HIGHLY CONFIDENTIAL

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Diagnostics Advisory Committee – Tuesday 16 April 2024

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

The following documents are made available to the Committee:

1. Overview

2. External Assessment Report produced by Southampton Health Technology Assessments Centre (SHTAC), University of Southampton, UK and Exeter Test Group, University of Exeter, UK - replacement version dated April 2024, used for decision making process.

3. Professional organisation submissions:
3a. Association of Respiratory Nurses
3b. British Thoracic Society
3c. Sleep Apnoea Trust

4. External Assessment Report (EAR) and economic model consultation comments
4a. Comments and responses, following 1st consultation (November 2023)
4b. Comments and responses, following 2nd consultation (March 2024)

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Diagnostics advisory committee:

16 April 2024

Lead team: John Cairns, Sophie West

External assessment group: Southampton Health Technology Assessments Centre (SHTAC)

NICE Technical team: Lirije Hyseni, Jessica Wilcock, Thomas Walker

NICE National Institute for Health and Care Excellence

ACIC

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome (OSAHS)

The following slides provide an overview of the external assessment group (EAG) report for this topic. Not all these slides will be presented at the committee meeting but the main information in this set of slides will be summarised. We have tried not to repeat information found in the other documents and references can be found in the slide notes.

Key documents in this assessment include:

- The final scope - contains the decision problem for the assessment
- The external assessment report (EAR)* - assessment of the included technologies by the EAG. The report has a more detailed executive summary which provides an overview of the EAG's work and links to the relevant sections of the report

The slides contain information that has been supplied in confidence. Academic in confidence information is underlined and highlighted in [redacted] and commercial in confidence information in [redacted]

* These documents are in the Committee pack and will be published at consultation

Background on OSAHS in adults

- OSAHS is a respiratory condition in which the upper airway becomes blocked repeatedly during sleep, reducing (hypopnoea) or intermittently stopping airflow completely (apnoea)
- Symptoms of OSAHS include stopping and starting breathing, making gasping, snorting or choking noises, waking up many times and loud snoring. Sleep interruptions can reduce quality of life, cognitive function and mental health
- About 9.6 million people (29.3%) in the UK aged between 30 and 69 years old have either mild (24.5%) or moderate to severe (4.8%) OSA[1]
    - In the UK, about 85% of people with OSA are undiagnosed.[2] If left untreated, OSA increases the risk of cardiovascular and cerebrovascular complications such as myocardial ischaemia, stroke and arrhythmias and can shorten life expectancy[3]
- COPD–OSAHS overlap syndrome occurs in people who have both chronic obstructive pulmonary disease (COPD) and OSAHS.

[1] Benjafield et al. (2019)
[2] British Lung Foundation (2014)
[3] Sharma and Culebras (2016)

OSAHS: obstructive sleep apnoea / hypopnoea syndrome

Background on OSAHS in children

- In children, the prevalence of OSA is between 1 to 3%. OSA is more common (up to 25%) in children with obesity, sickle cell disease or Down syndrome
- The most common cause is adenotonsillar hypertrophy (enlarged tonsils or adenoids) which can partially obstruct the airway during sleep[4]
- Daytime sleepiness causes a range of problems including sleep disruption, educational and neurocognitive impairment, and behavioural problems. OSAHS in children is also associated with failure to thrive, hypertension, cardiac dysfunction, and systemic inflammation
- For children with underlying conditions, OSA may also cause recurrent respiratory illness, hospital admissions and death[5]

OSAHS: obstructive sleep apnoea / hypopnoea syndrome

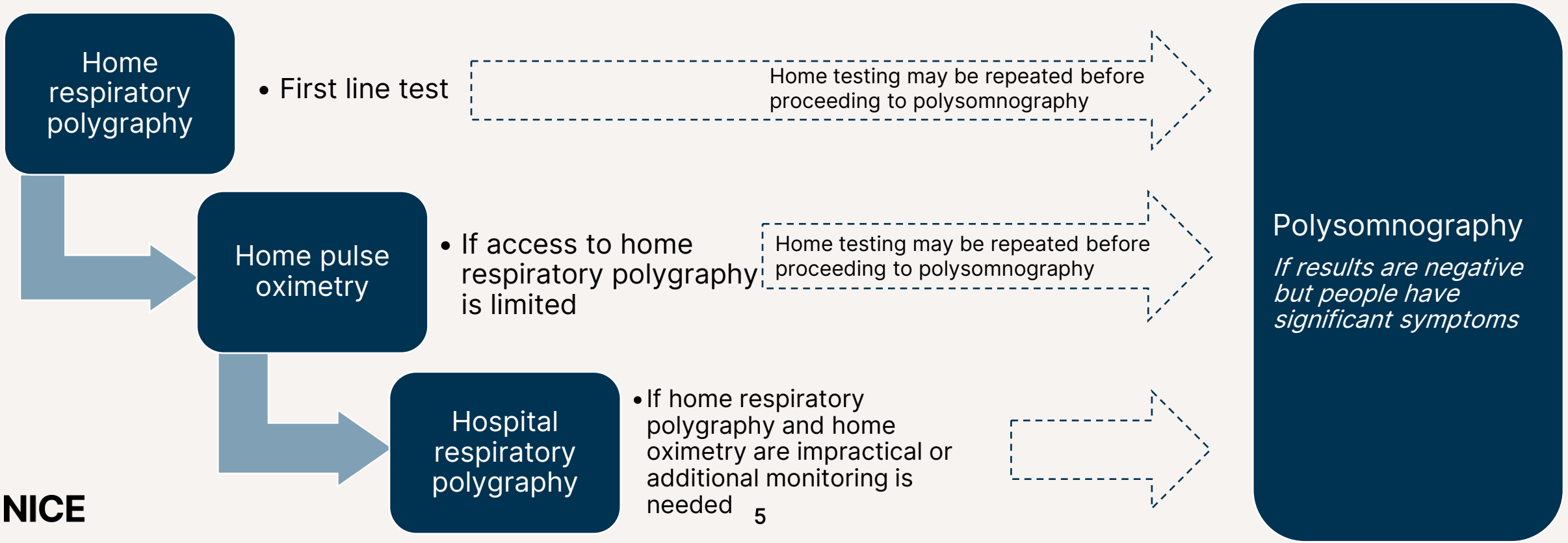
[4] GOSH NHS Foundation Trust, 2015
[5] BTS Guideline, 2022

Current practice (1)

A stakeholder noted that **oximetry is still used by a significant proportion of NHS sleep services** as it is cheaper and more accessible than other tests.

Diagnosing OSAHS in people over 16 years old

NICE's guideline on obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s (NG202) states that a person is referred to a sleep service if they have sleep history and symptoms indicating OSAHS to perform further diagnostic testing. The results of the sleep study can then be used to determine the severity of OSAHS



Current practice (2)

Children and young people aged 2 to 16 years old

- The British Thoracic Society guideline for diagnosing and monitoring paediatric sleep-disordered breathing (2023) recommend the use of home pulse oximetry and home respiratory polygraphy, where patients and carers are deemed appropriate for home sleep studies
- Preference may vary depending on patient features. Recommendations were made for the following groups:
    - Children without comorbidities
    - Children with comorbidities, for example neuromuscular disorders, Down syndrome or restrictive lung disease

Current practice (3)

Diagnosing and treating OSAHS

- NICE NG202 recommends using the results of the sleep study to diagnose OSAHS and determine the severity of OSAHS (mild, moderate or severe). This is determined by the number of events.
- The BTS guidance includes criteria for diagnosing OSAHS in children and people over 16 years old

	OAHI criteria	Severity criteria for both AHI and ODI:
Sleep apnoea type	People under 16 years old	People over 16 years old
Mild OSA	1 or more to less than 5	5 or more to less than 15 events per hour
Moderate OSA	5 or more to less than 10	15 or more to less than 30 events per hour
Severe OSA	10 or more	30 or more events per hour

- NICE NG202 recommends that personal treatment plans are tailored to the patient following an OSAHS diagnosis for people over 16 years of age
- For children over 2 years of age the ERS task force statements recommends a stepwise treatment approach

BTS British Thoracic Society; ESR European respiratory society; OAHI obstructive apnoea hypopnoea index

Decision problem

Decision question	Do novel home-testing devices for OSAHS represent a clinically and cost-effective use of NHS resources?
Populations	People with suspected OSAHS (who are considered suitable for a home sleep study). The population is separated by age groups: • People over 16 • Children and young people aged 16 and under
Interventions	• AcuPebble SA100 (Acurable) • Brizzy (Nomics) • NightOwl (ResMed) • Sunrise (Sunrise) • WatchPAT 300 (Zoll/Itamar) • WatchPAT ONE (Zoll/Itamar)
Comparators	Home respiratory polygraphy or home oximetry
Setting	Testing is to be done at home

For full decision problem see the final scope

Technologies under assessment

Six technologies are included

Companies note that devices have been given regulatory approval for different outputs and intended uses

Device name	Attachment details	Age*	Use	Severity Cut-off	Oximetry or PPG
AcuPebble SA100 (Acurable)	• Wireless sensor (throat)[a]	Adults	Reusable	AASM guidelines	Optional oximetry
Brizzy (Nomics)	• Waist belt hub • Wired sensors (chin and forehead)	3+	Reusable	Based on Martinot (2016)	Optional oximetry
NightOwl* (ResMed)	• Wireless sensor (finger)	13+	Disposable	AASM guidelines	PPG
Sunrise (Sunrise)	• Wireless sensor (chin)	3+	Disposable	Based on Pepin (2020)	Not applicable
WatchPAT 300* (Itamar/Zoll)	• Wrist strap • Oximeter (finger) • Wired sensor (chest)	12+	Reusable	AASM guidelines	PAT Oximetry
WatchPAT ONE* (Itamar/Zoll)			Disposable		

- The **NightOwl** technology available to the UK is the same as the CE marked NightOwl Mini and is awaiting a declaration of conformity based on the name change. A reusable version exists but will not be available in the UK.
- **WatchPAT300/ONE** replace an earlier version - WatchPAT 200U. The company state that the devices use the identical algorithm and produces identical signals. It also states that previous versions have considerably different functionality. WatchPAT ONE is a single-use version of WatchPAT 300

*Please see notes for more information; PAT: peripheral arterial tone; PPG: photoplethysmography

Internet connection Smartphone

Patient perspectives

Submission received from Sleep Apnoea Trust

- Sleep apnoea can have a devastating effect on relationships
- Diagnosing OSAHS can be onerous. Following a GP referral, it would include 2 to 3 visits to a sleep clinic. These novel home-testing devices have the capability to transform this process
- NG202 is an outstanding manual, but it is currently not backed up in reality
- More rapid and accurate diagnosis and treatment is vital. It is especially difficult for people with jobs that include driving or where vigilance is critical for safety
- Improving the diagnostic care pathway can improve people's lives and save the NHS millions of pounds by reducing the comorbidities that proliferate with undiagnosed OSAHS
- However, for new diagnostic technologies to be used, they must be proven to be of sufficient accuracy to help rather than hinder progress in identifying people with undiagnosed sleep apnoea

"Sleep Apnoea is unlike any other [condition]. There is no obvious sign of injury or disability, just an increasingly debilitating tiredness leading to an overwhelming and uncontrollable desire to sleep during normal waking hours."

"Primary care referral rates varies significantly across the UK and can be a barrier to successful progress in treating the millions as yet undiagnosed. Then the diagnostic pathway varies considerably, as does the use of advanced diagnostic equipment. Therefore, at present it is a postcode lottery."

"By improving the diagnostics technology pathway, it would elevate the whole treatment process to achieve the standards set by the new NICE NG202."

Clinical perspectives

Submissions received from

1. The British Thoracic Society
2. Association of Respiratory Nurses

- These novel home-testing devices may provide benefits; however, their accuracy, reliability, user friendliness, accessibility and efficiency needs to be demonstrated by the evidence
- It may benefit people who struggle to attend appointments in person
- It is important to understand the limitations of use. Need to consider the storage of devices, postage, and loss or no return of devices
- The IT, software or web-based infrastructure is an important consideration for clinicians and patients if they are expected to use their own mobile or computers for testing or to download the device data
- Sustainability should be considered

"Yes, [there is an unmet need]. There are a large number of referrals for suspected sleep disordered breathing to sleep services across the UK. Demand currently outstrips capacity."

"There is a pressing need to formally evaluate these newer devices, most importantly to understand validity and reliability and how they compare with existing devices as well as each other. They need to be user friendly, cost effective and have favourable environmental impact profile."

"To be able to have rapid and easy access to sleep study results would be beneficial to HCPs and patients."

Equality considerations (1)

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

- OSAHS is more common in people who are pregnant or overweight or obese
- Increased prevalence among older people and the condition is more common in men than women
- Many people with OSAHS may be protected under the disability provision of the Equality Act because their condition could have long-term adverse effects on their ability to do normal day-to-day activities
- People with comorbidities (e.g., Down syndrome) have a higher risk of developing OSAHS
- People who are frail or have cognitive impairment, or both, may struggle to use technologies that require more user-input. People with musculoskeletal issues (e.g., arthritis) may have trouble wearing devices that have multiple components. Technologies that are easier to use could offer additional benefit to these users
- Some people may be less familiar with, or confident, using technologies that require a smartphone and may need assistance with setting up and using smartphone applications

Equality considerations (2)

- Technologies that use light-based assessment, for example photoplethysmography (PPG) sensors and/or pulse oximetry sensors may overestimate levels of oxygen in the blood for people with darker pigmentation and skin tones
- Some technologies are contraindicated for people with pacemakers or other implantable devices, or with known or suspected arrhythmias, people with significant cardiopulmonary or neurological disorders, or people with a known allergy to acrylate
- Technologies that have adhesive sensors may not be suitable for people with physical features that impair adhesion (for example, skin growths or scars)
- The technologies differ in where sensors are attached, and some positions may be problematic if they require a beard to be shaved, and this has been grown for religious or cultural reasons
- Access to some technologies may be limited for those that do not have access to a smartphone or internet connection at home. Practices in rural or socioeconomically deprived areas could have more difficulty adopting these technologies

Clinical effectiveness

Objective

To conduct a systematic review of the clinical effectiveness (including diagnostic performance) of novel home-testing devices in people with suspected OSAHS

NICE National Institute for Health and Care Excellence

Included studies - people over 16 years

Included studies for people over 16 years

Novel device	Study name	No of pts.	Novel device Setting	Reference standard	
AcuPebble SA100	Devani (2021)	150	Home	Home	RP
	Sanchez Gomez (2024)	63	Hospital	Hospital	PSG
Brizzy	Martinot (2017)	92	Hospital	Hospital	PSG
NightOwl	Massie (2018)	101	Hospital	Hospital	PSG
	Massie (2022)	261	Hospital	Hospital	PSG
	Van Pee (2022)	167	Hospital	Hospital	PSG
	Lyne (2023)	100	Hospital	Hospital	PSG
Sunrise	Pepin (2020)	376	Hospital	Hospital	PSG
	Kelly (2022)	31	Home	Home	PSG
	Alsaif (2023)	[redacted]	Home	Home	RP
WatchPAT 300	Mueller (2022)	56	Home	Home	RP
WatchPAT ONE	Storey (2022)	600	Home	Home	RP
Supporting evidence					
WatchPAT 200-U	Tauman (2020)	101	Hospital	Hospital	PSG
	Pillar (2020)	84	Hospital	Hospital	PSG

RP respiratory polygraphy; PSG polysomnography

For detailed study descriptions and a general overview of study and patient characteristics see section 4.2 of the EAR

For further details see section 4.4

Study quality – people over 16 years

The EAG assessed study quality using the QUADAS 2 checklist

HIGH: Retrospectively set thresholds

Device	Study	Risk of bias				Applicability concerns		
		Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
AcuPebble SA100	Devani 2021	☺	☺	☺	☺	☺	☺	☺
	Sanchez Gomez 2024*	☺	☺	☺	☺	☺	?	☺
Brizzy	Martinot 2017*	☺	☺	☺	☺	☺	?	☺
NightOwl	Massie 2018 (reusable)	?	☺	☺	☺	?	?	☺
	Massie 2022 (reusable)	☺	☺	☺	☺	☺	?	☺
	Van Pee 2022 (reusable)	☺	☺	☺	☺	☺	?	☺
	Lyne 2023* (disposable)	?	☺	☺	☺	☺	?	☺
Sunrise	Pepin 2020*	☺	☹	☺	☺	☺	☹	☺
	Kelly 2022	☺	☹	☺	☺	☺	☹	☺
	Alsaif 2023	?	?	☺	☺	?	?	☺
WatchPAT 300/ONE	Mueller 2022	☺	?	?	☺	☺	☺	☺
	Storey 2022	☺	?	?	?	?	?	?
WatchPAT 200U	Pillar 2020	☹	☺	☺	☺	?	?	☺
	Tauman 2020*	?	☺	☺	☺	☺	?	☺

HIGH: Intentional selection for patients with heart failure

UNCLEAR: Hospital setting rather than home setting

* Studies used in base case

NICE

☺	☹	?
Low risk / concern	High risk / concern	Unclear risk / concern

Included studies - people under 16 years

Included studies for people under 16 years (N=3)

Novel device	Study name	No of Pts.	Novel device Setting	Reference standard	
AcuPebble SA100	NCT04031950 (2019)	[redacted]	Hospital	Hospital	[redacted]
Brizzy	Martinot (2015)	33	Hospital	Hospital	PSG
Sunrise	Martinot (2022)	140	Hospital	Hospital	PSG

Novel Device	Study	Risk of bias				Applicability concerns		
		Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
AcuPebble SA100	NCT04031950 (2019)*	☺	☺	☺	☺	☺	?	☺
Brizzy	Martinot (2015)	☺	?	☺	☺	☺	☹	?
Sunrise	Martinot (2022)	☺	☹	☺	☺	☺	☹	☺

*Unpublished interim results

HIGH: Retrospectively set thresholds

UNCLEAR: Hospital setting rather than home setting

HIGH: No reference standard

NICE

☺ Low risk / concern | ☹ High risk / concern | ? Unclear risk / concern

For further details on study quality see section 4.3-4

Patient demographics

Ethnicity and race

Only studies for AcuPebble and NightOwl reported the ethnicity and race of participants:

Devani (2021) - AcuPebble
White British (31.0%), Asian or Asian British (20.7%), White other (12.7%), White or Black Caribbean (3.3%), Black or Black British (2%), Indian (1.3%), Pakistani (1.3%), White or Black African (1.3%), Chinese (0.7%), Other (25.3%)

Massie (2022) – NightOwl
"Persons of diverse racial and ethnic backgrounds were included", no further information given.

Van Pee (2022) – NightOwl
Race: White (74%), Black (26%)
Ethnicity: Hispanic, Latino or Spanish (55%)

NCT04031950 (2019) - AcuPebble
[redacted]

For further details see section 4.2.2

Clinical effectiveness

People over 16 years of age

NICE National Institute for Health and Care Excellence

Diagnostic accuracy studies – people over 16 years

- All test accuracy studies assessed 2 tests that were used at the same time on the same people
- Most studies used hospital polysomnography (PSG) as a reference standard
    - 1 study (Kelly et al., 2022; Sunrise) used PSG at home
    - 1 study compared the novel device (Devani et al., 2021; AcuPebble) to home respiratory polygraphy (RP)
- No diagnostic accuracy studies were identified for WatchPAT 300/ONE. Supporting evidence from WatchPAT 200U device was identified.

EAG:
- Sleep study setting is known to influence the diagnostic performance of devices
- Home-based studies are more relevant to the decision problem than studies in which the novel device is used in a hospital-sleep laboratory

For further details see section 4.5.1 of the EAR

*Used in base case

Diagnostic accuracy (1)

Reference standard: Hospital PSG

[a] accuracy values for this cut-off, indicating test positivity for OSAHS (mild, moderate, severe), were estimated by the EAG based on data in the study publication

Author	Index test	No. pts	Cut-offs		Sensitivity % (95% CI)	Specificity % (95% CI)
			Index test	Reference test		
Sanchez-Gomez (2024)*	AcuPebble	63	AHI 15-30 or >30 (desat ≥3%)	AHI 15-30 or >30 (desat ≥3%)	93 (77 to 99)	97 (85 to 100)
			ODI 15-30 or > 30 (desat ≥3%)	ODI 15-30 or > 30 (desat ≥3%)	92 (74 to 99)	92 (79 to 98)
			AHI ≥5 (desat ≥3%)[a]	AHI ≥5 (desat ≥3%)[a]	[redacted]	[redacted]
Martinot (2017)*	Brizzy	92	MM-RDI > 5.9,	PSG RDI ≥5	93(87 to 97)	100 (51 to 100)
			MM-RDI > 13.5,	PSG RDI ≥15	89 (80 to 94)	100 (84 to 100)
			MM-RDI > 32.5,	PSG RDI ≥30	74 (58 to 86)	97(87 to 100)
Massie (2018)	NightOwl (Reusable)	101	NightOwl REI >5,	PSG-AHI >5	98 (92 to 99)	80 (44 to 97)
			NightOwl REI >15	PSG-AHI >15	97 (88 to 100)	83 (68 to 93)
			NightOwl REI >30	PSG-AHI >30	90 (76 to 97)	97 (89 to 100)
Van Pee (2022)	NightOwl (Reusable)	228	PAT AHI ≥5 (desat ≥3%)	PSG AHI ≥5 (desat ≥3%)	93 (89 to 97)	72 (54 to 91)
			PAT AHI ≥15 (desat ≥3%)	PSG AHI ≥15 (desat ≥3%)	91 (85 to 96)	76 (65 to 87)

AHI Apnea-hypopnea index; MSSS Michele Sleep Scoring System; MM mandibular movements; NOM NightOwl Mini (disposable); NOR NightOwl Reusable; ORDI Obstructive Respiratory Disturbance Index; Pts patients; , RDI Respiratory disturbance index

For details see section 4.5.1 of the EAR

*Used in base case

Diagnostic accuracy (2)

Reference standard: Hospital PSG

Author	Index test	No. pts	Cut-offs		Sensitivity % (95% CI)	Specificity % (95% CI)
			Index test	Reference std		
Lyne (2023)*	NightOwl (disposable)	94	NOM AHI < 5	PSG AHI < 5	93 (84 to 98)	77 (55 to 92)
	NightOwl (reusable)	96	NOR AHI < 5	PSG AHI < 5	89 (80 to 95)	91 (71 to 99)
Pepin (2020)*	Sunrise	376	Sr-RDI 7.63	PSG RDI ≥ 5	91 (89 to 92)	92 (90 to 94)
			Sr-RDI 12.65	PSGRDI ≥ 15	94 (91 to 97)	84 (81 to 87)
Pillar (2020)	WatchPAT 200U	84	WP AHI ≥15	PSG AHI ≥15	85 (NR)	70 (NR)
Tauman (2020)*	WatchPAT 200U	101	WP AHI ≥ 5	PSG AHI ≥ 5	96 (90 to 99)	25 (1 to 81)
			WP AHI ≥ 15	PSG AHI ≥ 15	88 (79 to 94)	63 (38 to 84)

AHI Apnea-hypopnea index; MSSS Michele Sleep Scoring System; MM mandibular movements; NOM NightOwl Mini (disposable); NOR NightOwl Reusable; ORDI Obstructive Respiratory Disturbance Index; Pts patients; , RDI Respiratory disturbance index

For details see section 4.5.1 of the EAR

Diagnostic accuracy (2)

Reference standard: Home PSG

Author	Index test	No. pts	Cut-offs		Sensitivity % (95% CI)	Specificity % (95% CI)
			Index test	Reference std		
Kelly (2022)	Sunrise	31	MM-ORDI 9.53	PSG ORDI >5	88 (69 to 97)	100 (54 to 100)
			MM-ORDI 12.65	PSG ORDI >15	100 (79 to 100)	75 (45 to 92)
			MM-ORDI 24.81	PSG ORDI >30	79 (NR)	96 (NR)

- Kelly et al. was the only study to use home PSG as a reference standard meaning that both novel devices and comparator were studied in the home

AHI Apnea-hypopnea index; MSSS Michele Sleep Scoring System; MM mandibular movements; NOM NightOwl Mini (disposable); NOR NightOwl Reusable; ORDI Obstructive Respiratory Disturbance Index; Pts patients; , RDI Respiratory disturbance index

For details see section 4.5.1 of the EAR

Diagnostic accuracy (3)

Reference standard: home RP

Author	Index test	No. pts	Cut-offs		Sensitivity % (95% CI)	Specificity % (95% CI)
			Index test	Reference test		
Devani (2021)	AcuPebble	150	15–30 AHI or >30 AHI (desat ≥3%)	15–30 AHI or >30 AHI (desat ≥3%)	93 (82 to 98)	97 (91 to 99)
			15–30 AHI or >30 AHI (desat ≥4%)	15–30 AHI or >30 AHI (desat ≥4%)	96 (86 to 100)	97 (92 to 99)
			15–30 ODI or >30 ODI (desat ≥3%)	15–30 ODI or >30 ODI (desat ≥3%)	91 (82 to 96)	93 (85 to 98)
			15–30 ODI or >30 ODI (desat ≥ 4%)	15–30 ODI or >30 ODI (desat ≥ 4%)	98 (89 to 100)	92 (85 to 97)
			≥5 AHI (desat ≥3%)*	≥5 AHI (desat ≥3%)*	92 (84 to 96)	96 (87 to 100)

EAG: Devani et al. used home RP as a reference standard which is not equivalent in accuracy to PSG and may overestimate AcuPebble's diagnostic accuracy. It is also different to the reference standard used for other novel devices, which makes it difficult to compare diagnostic accuracy with those for other novel devices that used PSG as the reference standard.

AHI Apnea-hypopnea index; MSSS Michele Sleep Scoring System; MM mandibular movements; NOM NightOwl Mini (disposable); NOR NightOwl Reusable; ORDI Obstructive Respiratory Disturbance Index; Pts patients; , RDI Respiratory disturbance index

For details see section 4.5.1 of the EAR

Intermediate outcomes (1)

Time to interpret device outputs

- Two studies reported estimates of the time it took for sleep study data to be scored and a diagnosis reached (Devani et al. 2021 **[AcuPebble],** Alsaif et al. 2023 **[Sunrise]**)
- Devani et al. compared AcuPebble with home RP (Embletta MPR Sleep System)
    - The study estimated that manual scoring of RP signals by experts to reach a diagnosis took 60 to 120 minutes, whereas for AcuPebble, zero time is required for the analysis to issue a diagnosis

EAG:

- Assumed that this range of estimates was based on all manual scoring of RP signals for all participants in the study (the source of the estimate is unclear in the publication)
- Sleep specialists would still need time to review the results of the automated sleep report for AcuPebble, but no estimate of this appears to have been included in the study publication

- Alsaif et al (2023) reported that [redacted]

For details see section 4.5.4 of the EAR

Intermediate outcomes (2)

Outcomes	No. of studies	Results
Agreement / concordance	12	In most studies the EAG noted satisfactory agreement between tests was reported.
Impact on clinical decision-making	[redacted]	[redacted]
Time to diagnosis or starting treatment	[redacted]	[redacted]
Number of repeat sleep studies	2	• In 1 study, no repeat tests were done (**AcuPebble SA100**). • In 1 study, 2 repeat tests were done because of operating errors (**WatchPAT300**). In the comparator group, 8 repeat tests were done because of insufficient recording time (n=2), failure to start the device (n=2), and loss of nasal pressure or inadequate examination time (n=3), and no reason stated (n=1)

For details see section 4.5 of the EAR

Intermediate outcomes (3)

Use of healthcare resources and costs

Devani (2021) for AcuPebble; Storey (2022) for WatchPAT ONE (conference abstract)

Use of healthcare resources	AcuPebble SA100	Home RP (Embletta MPR)
Cleaning	0.5 minutes	2 minutes
Device preparation	0.5 minutes	10 minutes
Time training patient on using the device	0	30 minutes
Analysis of signals by experts to issue a diagnosis	0	60 to 120 minutes
Cost	~£1	£250 to £500*

Use of healthcare resources	WatchPAT ONE	Home RP (NOX T3)
Number of appointments not attended by patients	13	42
Cost per appointment (including equipment, room, staff and postage)	£73.16	£39.91
Mean staff time taken per appointment (from check in to check out but excluding analysis)	12 minutes	21 minutes**
Mean patient time per appointment (including travel time)	12 minutes	223 minutes

For details see section 4.5.7 of the EAR

Clinical and patient reported outcomes

- Clinical outcomes: none of the studies reported mortality or morbidity outcomes
- Patient reported outcomes:
- None of the studies reported on health-related quality of life
- 3 studies reported that the novel devices were easy to use, and that sleep quality was good:
  - **Acupebble:** Most patients found it easy to use the mobile app and follow the app instructions. They found it easy to attach the AcuPebble SA100 sensor, finding it more comfortable and easier to attach than the sensors of the comparator (RP; Devani et al. 2021)
  - **WatchPAT 300:** Most patients said that falling asleep, and sleep quality was better with WatchPAT than with RP, however more patients experienced pain with WatchPAT than with RP (13% v 5%). 88% of patients expressed a preference for WatchPAT over RP if they were to undergo future testing (Mueller et al. 2022)
  - **Sunrise:** [redacted]
- 1 study assessed patient experience using the Sunrise device at home and reported that the ability to perform a sleep test at home [redacted] (Alsaif et al. 2023)

For details see section 4.7.2 of the EAR

Clinical effectiveness

Children and young people aged 2 to 16 years old

NICE National Institute for Health and Care Excellence

Diagnostic accuracy

For people under 16 years

- 2 of the 3 studies identified reported the accuracy of the novel devices in diagnosing OSAHS
  - The results for AcuPebble SA100 are unpublished preliminary results

Author	Index test	No. pts	Cut-offs	Cut-offs	Sensitivity % (95% CI)	Specificity % (95% CI)
			Index test	Reference std		
NCT04031950 (2019)	AcuPebble SA100	[redacted]	[redacted]		[redacted]	[redacted]
			[redacted]		[redacted]	[redacted]
Martinot (2022)	Sunrise (MMs)	140	Sr-RDI 5.75	PSG OAHI ≥1	83 (78 to 86)	53 (48 to 59)
			Sr-RDI 9.61	PSG OAHI ≥5	90 (87 to 93)	80 (76 to 84)
			Sr-RDI 13.07	PSG OAHI ≥10	100 (100 to 100)	88 (0.84 to 0.91)

AASM American Academy of Sleep Medicine; AHI Apnea-hypopnea index; MM mandibular movements; ORDI Obstructive; Pts patients; RDI Respiratory Disturbance Index; Sr Sunrise

For details see section 4.8.1 of the EAR

Clinical and patient reported outcomes

- **Clinical outcomes**: none of the studies reported mortality or morbidity outcomes
- **Patient reported outcomes:**
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]

For details see section 4.10 of the EAR

Ongoing studies

The EAG identified 3 references for 3 ongoing studies relevant to the review

Device	No. studies	Relevant details of ongoing studies (including PICO)	Estimated completion date
Sunrise	1	• **Population:** over 16s in France (n = 848) • **Comparator:** PSG (in lab or as outpatient) • **Outcomes:** includes time to diagnosis or starting treatment	March 2024
Sunrise	1	• **Population:** over 16s in Scotland (n = 100) • **Comparator:** 'detailed sleep test' (setting unclear) • **Outcomes:** not reported Limited details provided	Not reported
Sunrise	1	• **Population:** children in the UK (n = 100) Compares the diagnostic accuracy of home use of Sunrise to home cardio-respiratory polygraphy with Transcutaneous Carbon Dioxide monitoring, with the exception that children with significant co-morbidities or aged <9 years of age will undergo the same tests but in a sleep laboratory setting.	2023 (ongoing at the time of company submission)

For details see Appendix 2 Table 57 of the EAR

Issues for consideration (1)

Clinical effectiveness – in people over 16 years of age

Quality of evidence varied across devices

- There was no accuracy data available for **WatchPAT ONE** or **WatchPAT 300**. 2 studies with accuracy data for the predecessor version (WatchPAT 200U) were identified.
    - The company states that these devices use identical algorithms and produce identical signals
- For **Sunrise**, all studies had an unclear or high risk of bias related to the interpretation of the index test because they used post-hoc analyses to optimise diagnostic cut-off points
- For **NightOwl**, 2 of the 3 studies with accuracy data used a version (reusable) of the test not being marketed to the UK (Massie et al, Van Pee et al.). The company state that the only difference between the devices is the battery.
    - The other study, Lyne et al. (2023) had unclear risk of bias for patient selection

Issues for consideration (2)

Clinical effectiveness – in people over 16 years of age

Study design and setting

- Only 2 studies measured test accuracy in the home setting
    - Devani et al. (AcuPebble) and Kelly et al. (Sunrise)
- Only 1 study compared test accuracy against a comparator test for this assessment (home RP) in the same population (Devani et al. [AcuPebble])
- Different indices (AHI, ODI, RDI) and desaturation levels (3%, 4%) were used for accuracy studies across different devices. A stakeholder comment noted that these measure are not interchangeable.
- No studies compared the tests to pulse oximetry
- No studies compared different novel tests to each other

EAG: Cautioned against drawing conclusions about the relative superiority in diagnostic performance between the novel devices. The devices have not been formally compared in the same study with the same population and there is no formal statistical analysis to confirm any differences or equivalence between them

Issues for consideration (3)

Clinical effectiveness – in people under 16 years of age

Less evidence is available for this group

- Two novel devices for the monitoring of mandibular movements during sleep were evaluated (**Brizzy** with manual scoring; **Sunrise** with automated scoring)
    - However, accuracy data was only available for the **Sunrise** study. It used post-hoc analyses to optimise diagnostic cut-off points and was therefore scored as high risk of bias for the index test
- There is 1 ongoing study of **AcuPebble SA100** in the UK, with preliminary accuracy results available
- All studies were based in a sleep laboratory rather than at home
    - A stakeholder noted that carrying out trials with children at home poses ethical issues, and that ethics committees generally request them to be done in hospitals.

EAG: Overall, there is limited evidence from the available studies on the outcomes relevant to the decision problem. It is unlikely that the clinical results are directly transferrable from adults to children

Issues for consideration (4)

Equality considerations

No studies reported data for identified subgroups, including people from black, Asian and minority ethnic backgrounds

- Technologies that use light-based assessment, such as photoplethysmography (PPG) sensors or pulse oximetry sensors may overestimate levels of oxygen in the blood for people with low (more dangerous) levels of oxygen saturation, particularly for people with darker pigmentation and skin tones. This is an issue that has gained greater awareness since the COVID-19 pandemic and has been highlighted in the recently published Equity in Medical Devices: independent review (March 2024).
    - Pulse oximetry is a comparator for this assessment, and experts have indicated it is widely used in the NHS
    - The WatchPAT and NightOwl devices use a light based (PPG) technology
    - A stakeholder noted that one study on the performance of WatchPAT was missed, which showed a large drop in performance in a population with a high proportion of Black or African American participants (72%). This highlights an issue with using devices that rely heavily on PPG
- Sensor placement differs for different devices. Some positions may not be appropriate if they require a beard to be shaved which has been grown for religious or cultural reasons
- Some devices require a smartphone and internet connection, either to set up and download the sleep study or during the sleep study

Technologies under assessment

Equality considerations

Text in bold and underlined: Technology that uses light-based sensors which may overestimate levels of oxygen in the blood for people with darker pigmentation and skin tones.

Device name	Technology sensor mechanism (taken from manual)
AcuPebble SA100 (Acurable)	• Acoustic sensing of sounds generated during the process of respiration and cardiac operation. • **Optional oximetry**
Brizzy (Nomics)	• Mandibular movements • **Optional oximetry**
NightOwl (ResMed)	• Continuous recording of pulse waveform (also known as **photoplethysmography [PPG]**)
Sunrise (Sunrise)	• Mandibular movements
WatchPAT 300 (Itamar/Zoll)	• **Peripheral Arterial Tone (PAT) signal measures arterial pulse volume changes in the finger as a result of vasomotion (vasoconstriction and vasodilatation).** • **Oximetry**
WatchPAT ONE (Itamar/Zoll)	

Cost effectiveness

People over 16 years of age

Objectives

- To conduct systematic reviews of evidence to inform a health economic evaluation of novel home-testing devices in people with suspected OSAHS
- To conduct a health economic evaluation using decision-analytic modelling to assess the incremental cost-effectiveness of novel home-testing devices compared to home RP or home oximetry in people with suspected OSAHS

NICE National Institute for Health and Care Excellence

Company submitted models

One company submitted an economic model

- ResMed submitted a model that compared NightOwl with home respiratory polygraphy (RP)
- The diagnostic performance of NightOwl was from van Pee et al. (2022) and the diagnostic accuracy of home respiratory polygraphy was taken from the NG202 economic modelling
    - The EAG used different accuracy estimates in its base case
- Estimated that NightOwl saved £171 per diagnosis made compared to home RP

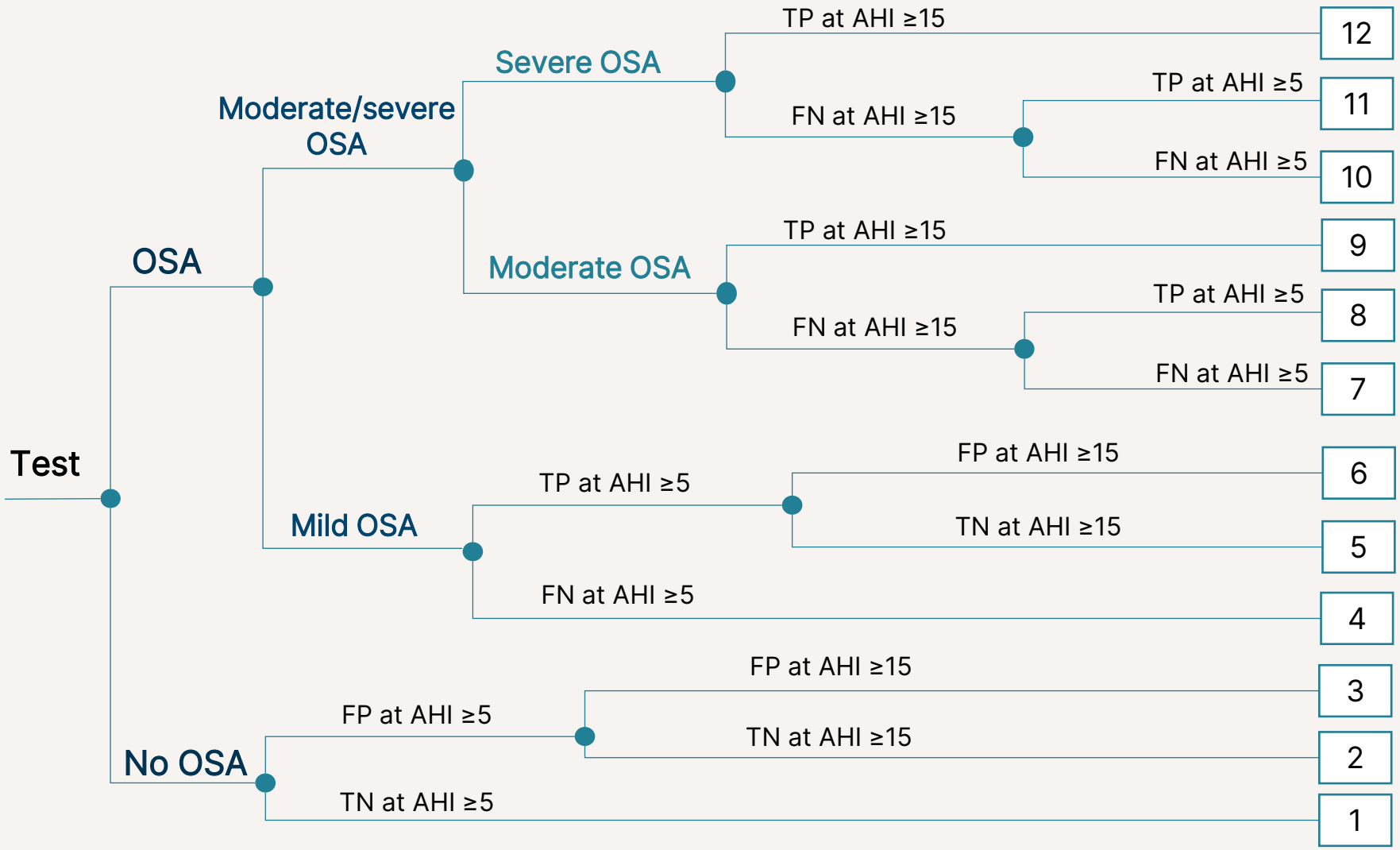
The EAG stated there were several limitations of the model, including:

- Omission of failure rates for both arms, and the subsequent impacts on costs
- Accuracy data is from van Pee (2022) which evaluated the reusable version of NightOwl, rather than the disposable version that will be commercialised in the UK
- A higher cost for home respiratory polygraphy was used than in the EAG model
- No consideration of treatment costs or longer-term impacts, or health outcomes beyond diagnostic accuracy

HRP: home respiratory polygraphy

For more detail see section 5.3 of the final EAR

EAG's model structure – Decision Tree



Test Result	True Condition
Moderate/Severe	Severe
Mild	Severe
No OSA	Severe
Moderate/Severe	Moderate
Mild	Moderate
No OSA	Moderate
Moderate/Severe	Mild
Mild	Mild
No OSA	Mild
Moderate/Severe	No OSA
Mild	No OSA
No OSA	No OSA

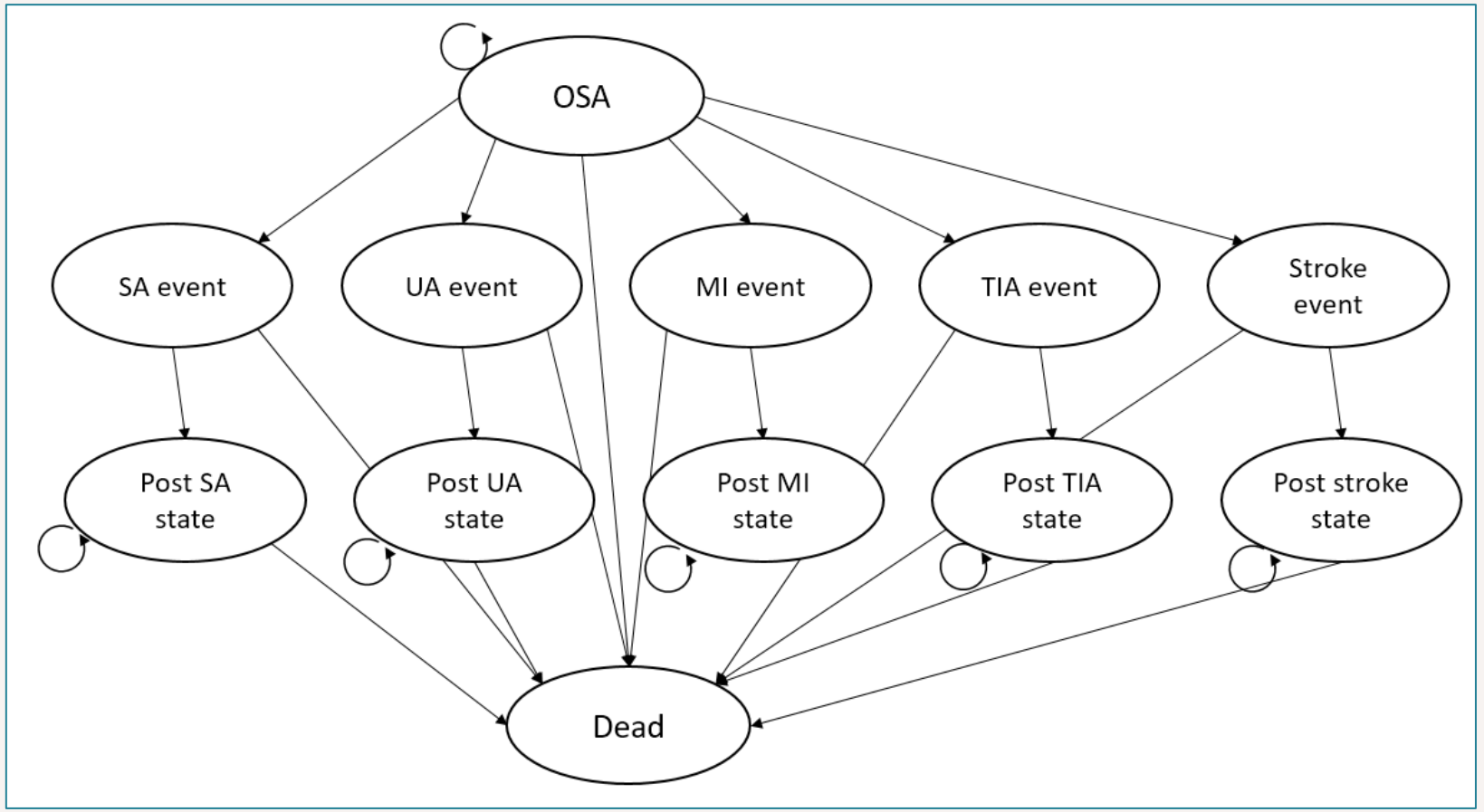
NICE

AHI: apnoea-hypopnea index, FN: False negative, FP: False positive, OSA: Obstructive sleep apnoea, TN: True negative, TP: True positive

For more detail see section 5.6.1 of the final EAR

EAG's model structure – Markov Model

The main purpose of the Markov model was to estimate the impacts of treatment decisions (informed by the diagnostic pathway) on the utility, risk of cardiovascular events and RTAs, alongside related costs, for people with OSAHS



SA: Stable Angina, UA: Unstable Angina, MI: Myocardial Infarction, TIA: Transient Ischemic Attack; RTAs: road traffic accidents

For more detail see section 5.6.2 of the final EAR

EAG's model inputs and assumptions (1)

EAG: Model largely based on the existing cost-effectiveness model from the NICE guideline for adults (NG202)

Input	Assumptions	Source
Time to diagnosis	3-months to diagnosis*	Clinical experts
Time to treatment	• 3-months to commence treatment post diagnosis • 1 month treatment delay for individuals needing a second sleep study due misdiagnosing moderate or severe as having no OSA	Clinical experts
Test failure rate	Differed between tests, from 0.6 to 11.5%* Failure rate for RP was 5.4%* • [redacted] used in scenario analysis	• Newcastle Regional Sleep Service • Diagnostic accuracy studies • Alsaif et al. (2023)

* Assumption investigated in scenario analysis

EAG: In one-way sensitivity analysis we assessed the impact of increasing time to diagnosis and time to treatment to 6 months and decreasing time to diagnosis and time to treatment to 1.5 months (not based on any data). We also conducted a scenario analysis investigating the impact of reduced time to diagnosis and treatment based on data from Alsaif et al. (2023). The scenario analysis made no difference to the overall cost-effectiveness results from the base case analysis

AHI: apnoea-hypopnea index, CM: conservative management, CPAP: continuous positive airway pressure, MAD: mandibular advancement device

For more detail see Section 5.7 & 5.8 and Appendix 9b Table 69 of the final EAR

*Used in scenario analysis (home based studies)

EAG's model inputs and assumptions (2)

Diagnostic accuracy estimates used in the base case model

	Interventions assessed in clinical setting with in-lab PSG as reference standard											
Device	AcuPebble		Brizzy		NightOwl		Sunrise		WatchPAT 300		WatchPAT ONE	
Cut-off	AHI ≥ 5	AHI ≥ 15	AHI ≥ 5	AHI ≥ 15	AHI ≥ 5	AHI ≥ 15	AHI ≥ 5	AHI ≥ 15	AHI ≥ 5	AHI ≥ 15	AHI ≥ 5	AHI ≥ 15
Sensitivity	[redacted]	0.93	0.93	0.89	0.93	0.89	0.91	0.92	0.96	0.88	0.96	0.88
Specificity	[redacted]	0.97	1.00	1.00	0.77	0.82	0.94	0.84	0.25	0.63	0.25	0.63
Source	Acurable (2023) Sanchez-Gomez (2024)		Martinot (2017)		Lyne (2023)		Pepin (2020)		Tauman (2020)		Tauman (2020)	

EAG selection and uncertainties of intervention accuracy data:

- **Sunrise**: Pepin (2020) data was chosen over Kelly (2022)* because in-lab PSG was the reference standard (vs home PSG), and it has a larger sample size (n=376 versus n=31)
- **NightOwl**: Lyne (2023) data was used rather than van Pee (2022) or Massie (2018); it used the device version that will be available in the UK
- **WatchPAT**: Assumed that the accuracy of earlier version (200U) is the same as that for WatchPAT 300 and WatchPAT ONE
- **AcuPebble**: the Macarena trial data was used because the reference standard was in-lab PSG compared to home RP for the Devani study*

AHI: apnoea-hypopnea index, PSG: polysomnography RP: respiratory polygraphy

For more detail including alternative diagnostic accuracy data see Section 5.7.3 (table 18) of the final EAR

EAG's model inputs and assumptions (3)

Diagnostic accuracy estimates for comparators used in the base case model

Device	Oximetry Base case		Respiratory polygraphy Base case		Respiratory polygraphy Scenario analysis		Respiratory polygraphy Scenario analysis	
Cut-off	ODI ≥ 5	ODI ≥ 15	AHI ≥ 5	AHI ≥ 15	AHI ≥ 5	AHI ≥ 15	AHI ≥ 5	AHI ≥ 15
Sensitivity	0.52	0.35	0.95	0.93	0.87	0.77	0.94	0.84
Specificity	0.96	0.99	0.69	0.85	0.67	0.95	0.58	0.89
Source	NG202		Xu (2017)		Pereira (2013)		NG202	

EAG selection of RP comparator accuracy data:

- For RP in the base case, Xu (2017) was chosen as it evaluates RP in the home-setting, is compared to in-lab PSG, uses a named device representative of what is currently used in England, and is 1 of the more recent studies
  - Xu (2017) was conducted in China and includes only 80 participants
- Pereira (2013) also evaluated RP in the home setting and compared to in-lab PSG, but was slightly older than Xu (2017)
  - Pereira (2013) included 128 participants with suspected OSAHS in Canada
- NG202 model used a pooled estimate from 8 studies (including both Xu and Pereira) for home RP

AHI: apnoea-hypopnea index, ODI oxygen desaturation index, RP: respiratory polygraphy

For more detail see Section 5.7.4 in final EAR

EAG's model inputs – Base case costs of novel devices

	First night	Repeat (test failure)	Source
Home oximetry	£18	£212	Inflated device based on NG202 costs
Home RP	£212	£212	National schedule of NHS costs 2021/22
AcuPebble	£74	£1 if automated*	Company submission
Brizzy	£70	£28	Company submission
NightOwl	£111	£13*	Company submission
Sunrise	£83	£84	Company submission
WatchPat 300	£82	£30	Company submission
WatchPat ONE	£103	£25	Company submission

*Assuming the device is still with the patient

Time to review device output:

- Companies suggested different lengths of time to review device outputs and noted that some devices have been given regulatory clearance for fully automated diagnosis
- The EAG assumed 20 minutes to review device outputs for all novel devices, and 10 mins for oximetry. A scenario analysis assumed that time to review was as reported by the companies

EAG: We understand that there is variation in how much reliance clinical services are prepared to place on automated diagnosis. Clinicians noted that they still access raw data and manually review sleep studies for some or all patients and that it varies between centres

FN: False negative, RP: respiratory polygraphy

For more detail see section 5.7.12-17 of the final EAR

Base case results (1)
Compared to oximetry (probabilistic)

EAG: When comparing the novel devices to oximetry, the results are generally consistent, in that novel devices are seen to be more cost-effective than oximetry

	AcuPebble	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Incremental cost (95% CI)	£535 (168, 992)	£481 (107, 903)	£630 (236, 1028)	£627 (262, 1041)	£755 (329, 1204)	£775 (328, 1197)
Incremental QALYs (95% CI)	0.085 (0.008, 0.223)	0.076 (0.003, 0.209)	0.083 (0.002, 0.204)	0.09 (0.012, 0.224)	0.089 (0.009, 0.224)	0.089 (0.008, 0.212)
ICER (£ per QALY gained)	£6,271	£6,360	£7,613	£6,989	£8,458	£8,750
INMB at £20,000 per QALY gained (95% CI)	£1171 (-196, 3549)	£1031 (-190, 3357)	£1025 (-306, 3249)	£1168 (-186, 3510)	£1031 (-321, 3419)	£996 (-411, 3276)
INMB at £30,000 per QALY gained (95% CI)	£2,023 (-64, 5774)	£1,787 (-129, 5462)	£1,852 (-218, 5294)	£2,066 (-59, 5736)	£1,923 (-204, 5707)	£1,881 (-270, 5408)
Probability cost-effective at £20,000/ QALY	94.3%	92.8%	90.3%	93.7%	88.9%	87.6%
Probability cost-effective at £30,000/ QALY	96.8%	95.9%	94.4%	96.7%	95.0%	93.8%

ICER: incremental cost effectiveness ratio, INMB: incremental net monetary benefit, QALY: quality adjusted life year

For more detail see section 5.10.1 of the final EAR

Base case results (2)

Compared to respiratory polygraphy (probabilistic)

* ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

EAG: High uncertainty over the relative diagnostic accuracy estimates for all devices and advised caution in interpreting these results. We have not reported a fully incremental analysis due to many differences between data sources used in the model for the devices.

	AcuPebble	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Incremental cost (95% CI)	-£264 (-475, -64)	-£318 (-524, -119)	-£169 (-395, 35)	-£171 (-319, -2)	-£44 (-298, 235)	-£24 (-286, 237)
Incremental QALYs (95% CI)	-0.006 (-0.047, 0.029)	-0.016 (-0.056, 0.017)	-0.009 (-0.051, 0.025)	-0.002 (-0.027, 0.026)	-0.002 (-0.04, 0.033)	-0.003 (-0.045, 0.036)
ICER (£ per QALY gained)	£ 43,505*	£20,199*	£19,640*	£108,795*	£21,216*	£ 8,570*
INMB at £20,000 per QALY gained (95% CI)	£143 (-519, 678)	£3 (-652, 482)	-£3 (-697, 487)	£140 (-238, 572)	£2 (-557, 533)	-£32 (-661, 508)
INMB at £30,000 per QALY gained (95% CI)	£82 (-993, 954)	-£154 (-1229, 640)	-£89 (-1204, 727)	£124 (-513, 821)	-£18 (-950, 840)	-£60 (-1131, 837)
Probability cost-effective at £20,000/ QALY	78.90%	58.40%	58.10%	81.60%	55.20%	48.10%
Probability cost-effective at £30,000/ QALY	65.30%	40.00%	46.40%	69.50%	51.50%	46.50%

ICER: incremental cost effectiveness ratio, INMB: incremental net monetary benefit, QALY: quality adjusted life year

For more detail see section 5.10.1 of the final EAR

Scenario analyses (1)

Alternative diagnostic accuracy estimates for home respiratory polygraphy

- Using data from Pereira (2013) or the NG202 economic model improves the estimated cost-effectiveness of all novel devices relative to RP: all now have a positive INMB
- Pereira and NG202 are less favourable for home RP than the estimates from Xu (2017) used in the base case

	AcuPebble	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Using home RP accuracy data from Pereira (2013)						
Incremental cost	£14	-£51	£108	£101	£228	£251
Incremental QALYs	0.037	0.026	0.035	0.041	0.040	0.040
ICER (£ per QALY gained)	£382	Dominant	£3,129	£2,471	£5,673	£6,247
INMB at £20,000 per QALY gained	£730	£547	£582	£716	£577	£553
INMB at £30,000 per QALY gained	£1,102	£835	£927	£1,125	£979	£956
Using home RP accuracy data from NG202						
Incremental cost	-£127	-£192	-£33	-£40	£87	£111
Incremental QALYs	0.018	0.007	0.016	0.022	0.021	0.021
ICER (£ per QALY gained)	Dominant	Dominant	Dominant	Dominant	£4,092	£5,174
INMB at £20,000 per QALY gained	£493	£337	£346	£480	£340	£317
INMB at £30,000 per QALY gained	£677	£410	£502	£700	£554	£531

ICER: incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years, RP: respiratory polygraphy

For more detail see section 5.10.2 of the final EAR

Scenario analyses (2)

Alternative diagnostic accuracy estimates for interventions

Scenario analyses used base case estimates for home RP (Xu et al.)

Parameterisation of decision tree for respiratory polygraphy, AcuPebble, NightOwl, WatchPAT 300/ONE based on the raw 4x4 contingency table data

- WatchPAT 300 showed an improvement in the INMB at £20,000 per QALY gained (£2 in the base case to £73) compared to RP
- There is a small improvement for NightOwl in the INMB at £20,000 per QALY gained (-£3 in the base case to £26), and a slight reduction for AcuPebble (£143 in the base case to £94) compared to RP

Alternative diagnostic accuracy estimates for NightOwl

- Using the estimates from Massie (2018) for NightOwl increased the INMB at £20,000 per QALY in the base case from -£3 to £281 compared to RP. When using estimates from Van Pee (2020) this would increase to £88

Alternative diagnostic accuracy from home-based studies

- Using estimates from Kelly (2022) for **Sunrise** increased the INMB at £20,000 per QALY in the base case from £140 to £451 compared to RP
- Using estimates from Devani (2021) for **AcuPebble** decreased the INMB at £20,000 per QALY in the base case from £143 to £138 compared to RP

INMB incremental net monetary benefit; QALYs quality-adjusted life-years, RP: respiratory polygraphy

For more detail see section 5.10.2 of the final EAR

Scenario analyses (3)

Alternative model assumptions

Scenario analyses used base case estimates for home RP (Xu et al.)

Values used in the EAG's base case are representative of those used in NG202

Failure rate data for RP

- When the EAG assumes a higher failure rate for RP of [redacted] as reported in the Alsaif (2023), WatchPAT ONE is the only device that is not estimated to be cost-effective compared to RP at the £20,000 threshold

Time for data review of novel devices

- Using company stated time to review outputs of novel devices only resulted in small changes to the INMB at £20,000 per QALY gained and only changed the cost-effectiveness results for NightOwl, which became cost-effective

Reduced time to diagnosis and treatment

- When the EAG assumes a reduced time to diagnosis of [redacted] rather than 3 months, and a time to treatment of [redacted] rather than 3 months after diagnosis (Alsaif, unpublished study), this made no difference to the overall results from the base case analysis.

ICER: incremental cost-effectiveness ratio; INMB incremental net monetary benefit; OSAHS: Obstructive sleep apnoea hypopnea syndrome, QALYs quality-adjusted life-years, RP: respiratory polygraphy

For more detail see section 5.10.2 of the final EAR

One-way sensitivity analyses

Across all novel devices, the parameters having the most impact in one-way sensitivity analyses were:

- Utilities for mild and moderate OSAHS
- Sensitivity and specificity estimates for the novel devices
- Sensitivity and specificity estimates for the comparators, and
- Prevalence parameters

- When comparing the novel devices to RP, changes in the sensitivity of RP or of the novel device at the high diagnostic cut-off would lead to changes in the results compared to the base case analysis
- When comparing the novel devices to oximetry, the results are generally consistent in that the novel devices are seen to be more cost-effective than oximetry, regardless of the parameter inputs

OSAHS: Obstructive sleep apnoea hypopnea syndrome, RP: respiratory polygraphy

For more detail see section 5.10.2 of the final EAR

Cost effectiveness

Children and young people aged 2 to 16 years old

Modelling for children and young people

There are several challenges for modelling in children and young people

EAG: Did not consider that a decision model is likely to resolve uncertainty over the cost effectiveness of home-based assessment with novel devices in children at this time

- Despite recent BTS guidelines (June 2023), there remains some uncertainty around the current clinical pathways for children with and without comorbidities
- The EAG considered the absence of relevant test accuracy data is a barrier to credible modelling of the cost-effectiveness of home-based devices for OSAHS in children at the present time
    - The EAG stated that there is consensus that evidence from adults is not automatically generalisable to children
    - A study in progress (NCT04031950) examining the diagnostic accuracy of AcuPebble in a paediatric population may have the potential to inform a future economic model
- There are questions regarding the clinical effectiveness of adenotonsillectomy as first line treatment in children without comorbidities, and there is uncertainty in the choice and effectiveness of treatment for mild OSAHS
- In children, the evidence on the extent and reversibility of longer-term impacts of untreated OSAHS is less clear
- No relevant utility estimates with a UK general population valuation identified

BTS: British Thoracic Society
OSAHS: Obstructive sleep apnoea hypopnea syndrome

For more detail see section 5.11.1 of the final EAR

Modelling for children and young people

EAG: Proposed potential model structures for children without comorbidities including parameter requirements

The EAG stated that the key evidence gaps for children, other than test accuracy, include:

- The impact of OSAHS on health-related quality of life for children stratified by OSAHS severity. Studies using preference-based utility instruments with a UK general population value set (e.g. using the CHU9D) would help to inform future economic evaluations

- The relationship between OSAHS in childhood and long-term effects on health outcomes and well-being, and the extent to which these effects can be assumed to be causative and reversible with appropriate treatment

OSAHS: Obstructive sleep apnoea hypopnea syndrome

For more detail see section 5.11.3 of the final EAR

Issues for consideration (1)

Data used for diagnostic accuracy estimates

- The same issues with test accuracy data for the devices raised in the clinical effectiveness section are relevant to their use as model inputs
- No data were available comparing novel tests done in same population. The EAG cautioned against comparing the cost effectiveness of the different novel tests based on available data
- Accuracy data used in the base case is based on hospital-based studies instead of at home. Cost effectiveness estimates using studies done at home were available for 2 devices (AcuPebble and Sunrise)

Alternative accuracy data for NightOwl

- Using alternative diagnostic accuracy inputs for NightOwl improved cost-effectiveness results when compared to home RP
  - Massie (2018) and Van Pee (2020) used a different version of NightOwl which will not be available in the UK

Source of diagnostic accuracy data for respiratory polygraphy

- Choice of source for sensitivity and specificity estimates for home respiratory polygraphy has a big impact on cost effectiveness estimates for the devices

Issues for consideration (2)

Novel devices compared with oximetry

- When comparing the novel devices to oximetry, the results are generally consistent, in that novel devices are more cost-effective than oximetry, regardless of the parameter inputs
- A stakeholder noted that oximetry is still used by a significant proportion of NHS sleep services

Sustainability

- Some of the technologies are reusable (AcuPebble, Brizzy, Sunrise, WatchPAT 300) and others are disposable (NightOwl, WatchPAT ONE)
  - Disposable devices will have an environmental cost and a larger number of devices may need to be stored
  - Any reduction in the need to travel to healthcare centres to collect, and return, equipment may have benefits in terms of reducing carbon dioxide emissions
  - If devices need to be returned, this may cause delays to the devices being available if not returned in a timely manner

Subgroups

- Due to a lack of data, the EAG were not able to assess cost-effectiveness separately for any of the subgroups highlighted in the scope. This included people with darker pigmentation and skin tones

Issues for consideration (3)

Potential impacts of novel devices that are not captured in the model raised by stakeholders

- Accessibility to home testing and number of patients lost to follow up or unwilling to undergo prescribed tests (influenced by device usability)
- Are there any potential benefits in terms of reducing inequalities? For example:
  - addressing health inequalities or in some cases enabling the diagnosis of vulnerable individuals
  - travel time to pick up RP

Uncertainty about cost effectiveness of technologies in children and young people

- There is limited evidence for this group
- Even if suitable accuracy data is available, the need to resolve further uncertainties may be required to obtain cost effectiveness estimates
  - These data may be more difficult to generate and require longer term studies to resolve uncertainties than the accuracy data

NICE National Institute for Health and Care Excellence

Thank you.

Appendix

NICE National Institute for Health and Care Excellence

Technologies under assessment

Company comments on differences between device versions

Device in scope	Devices with data included in assessment
NightOwl	• The company have noted that the only difference between the **NightOwl Mini** and **NightOwl Reusable** devices is whether the battery can be recharged. • The NightOwl device to be launched in the UK in 2024 is the **NightOwl Mini** under a new name ("NightOwl"). There is no change to the sensor or software. • There will be a Declaration of Conformity for a disposable product, named NightOwl (currently known as NightOwl Mini) to be worn on the finger. This is currently anticipated for April 2024.
WatchPAT 300/ WatchPAT ONE	• The company wish to advise that both **WatchPAT 300** and **WatchPAT One** are using identical algorithm to the one used in **WatchPAT 200U** and in the Tauman study. • Both devices produce identical signals to that of WatchPAT 200U. This similarity allowed FDA and CE for the new devices based on technological continuity. • The version before **WatchPAT200U**, the **WatchPAT 200** has considerably different functionality; PAT finger sensor did not include oximetry. Oximetry was measured from an adjacent finger by a separate sensor.

Diagnostic Assessment Report commissioned by the NIHR on behalf of the National Institute for Health and Care Excellence

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome - a systematic review and economic evaluation

ERRATUM

Produced by	Southampton Health Technology Assessments Centre (SHTAC), University of Southampton, SO17 1BJ, UK[1] and Exeter Test Group, University of Exeter, EX1 2LU, UK [2]
Authors	Jaime Peters, Senior Research Fellow[2] Jonathan Shepherd, Principal Research Fellow[1] Emma Maund, Research Fellow[1] Bogdan Grigore, Associate Research Fellow[2] Lois Woods, Senior Research Assistant t[1] Joanne Lord, Professorial Research Fellow in Health Economics[1] David Alexander Scott, Principal Research Fellow[1] Chris Hyde, Professor of Public Health and Clinical Epidemiology[2]
Correspondence to	Dr Jonathan Shepherd Principal Research Fellow Southampton Health Technology Assessments Centre (SHTAC) University of Southampton Alpha House Enterprise Road, Southampton Science Park Southampton, SO16 7NS, UK. Email: jps@southampton.ac.uk
Date completed	3rd April 2024

Declared competing interests of the authors

The authors have no competing interests to declare.

Acknowledgements

We are grateful to the following for providing expert methodological and clinical advice:

Adam Birdseye-Dastagir, Senior Medical Technologist, Guy's and St Thomas' NHS Foundation Trust

Richard Brown, Highly Specialist Medical Technologist, Guy's and St Thomas' NHS Foundation Trust

Dr Paddy Dennison, Consultant in Respiratory and General Internal Medicine, University Hospital Southampton

Dr Panagis Drakatos, Consultant in Sleep and Respiratory Medicine, Guy's and St Thomas' NHS Foundation Trust

Dr Hemant Kulkarni, Consultant in Paediatric Respiratory Medicine, Sheffield Children's Hospital

Sharon Stothard, Sleep Clinical Nurse Lead, Sunderland and South Tyneside NHS Foundation Trust

Dr Martin Samuels, Consultant Respiratory Paediatrician, Great Ormond Street Hospital, London and Staffordshire Children’s Hospital, Stoke-on-Trent

Bruno Santos, Lead Sleep Physiologist, University Hospital Southampton NHS Foundation Trust

Dr Chris Turnbull, Consultant in Respiratory Medicine, Oxford University Hospitals NHS Foundation Trust

We thank the NICE Specialist Committee Members (SCMs) on this assessment for their informative comments on a draft of this report and their expert clinical advice; The NICE Guideline Update Team who developed the NG202 economic model which informed development of the model for this assessment, and also to the NICE Centre for Guidelines for sharing the model.; The NICE Diagnostic Assessment Programme team for their assistance during the assessment.

We thank our colleagues Karen Pickett, Senior Research Fellow and Jo Picot, Senior Research Fellow from SHTAC for reading and commenting on a draft of this report.

Contributions of authors

Jaime Peters (Senior Research Fellow, Health Economics) carried out the review of economic evaluations, developed the independent economic model, and drafted the report;

Jonathan Shepherd (Principal Research Fellow, Evidence Synthesis) carried out the systematic review of clinical effectiveness, drafted the report, and was project co-ordinator;

Emma Maund (Research Fellow, Evidence Synthesis) carried out the systematic review of clinical effectiveness, and drafted the report;

Bogdan Grigore (Postdoctoral Research Fellow, Health Economics) carried out the review of economic evaluations, developed the independent economic model, and drafted the report;

Lois Woods (Senior Research Assistant, Evidence Synthesis) conducted literature searches, carried out the systematic review of clinical effectiveness, and drafted the report;

Joanne Lord (Professorial Fellow in Health Economics) reviewed the independent economic model, and drafted the report;

David Alexander Scott (Principal Research Fellow in Medical Statistics and Health Economics) carried out the systematic review of clinical effectiveness and drafted the report;

Chris Hyde (Professor of Public Health and Clinical Epidemiology) developed the independent economic model, drafted the report, and provided strategic guidance and insights on diagnostic test evaluation.

Data-sharing statement

Completed data extractions for all studies included in the systematic review of clinical effectiveness can be provided on request to the authors.

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR135895.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR. Any errors are the responsibility of the authors.

Commercial in confidence information is highlighted and underlined in [redacted]

Academic in confidence information is highlighted and underlined in [redacted]

ABSTRACT

Background

Obstructive sleep apnoea and hypopnoea syndrome (OSAHS) is a sleep-related breathing disorder in adults and children caused by intermittent narrowing or complete collapse of the upper airway. Diagnostic sleep studies for OSAHS are conducted overnight in a home or hospital setting using devices that monitor physiological parameters. Current methods are oximetry, respiratory polygraphy and polysomnography (PSG). Portable novel devices have been developed to facilitate home sleep studies. There is uncertainty over the clinical and cost-effectiveness of these novel devices. This report was commissioned to inform a Diagnostic Assessment of novel home-testing devices for OSAHS conducted by the National Institute for Health and Care Excellence (NICE).

Objectives

- To conduct a systematic review of clinical effectiveness (including diagnostic performance) of novel home-testing devices in people with suspected OSAHS.
- To conduct systematic reviews of evidence to inform a health economic evaluation of novel home-testing devices in people with suspected OSAHS.
- To conduct a health economic evaluation using decision-analytic modelling to assess the incremental cost-effectiveness of six novel home-testing devices compared to home respiratory polygraphy or home oximetry in people with suspected OSAHS.

Data sources and methods

A systematic review of clinical effectiveness was conducted in accordance with a pre-defined protocol. Searches were conducted between 22nd and 24th May 2023, and updated on 25th September. Screening, full text review, data extraction and critical appraisal was conducted by two reviewers. The extracted data was used to inform a structured descriptive synthesis of clinical effectiveness. Summary numerical and statistical data were tabulated with accompanying textual description. Clinical heterogeneity was substantial, so meta-analysis was not appropriate. For the economic evaluation, we reviewed published and submitted evidence for both adult and child populations and developed a model to assess the cost-effectiveness of the different novel devices compared to respiratory polygraphy and oximetry in an adult population.

Results

In people aged over 16 years, novel devices were compared to home RP in four studies, to hospital sleep-laboratory PSG in five studies, and home-based PSG in one study. No eligible studies were identified which compared novel devices to home-based pulse oximetry.

Sensitivity and specificity estimates vary between studies and within studies at different severity cut-offs. Sensitivity was generally high, in the range 80 to 100%, and fell below 80% in just two studies. In contrast, specificity was more variable with estimates ranging from 25% to 100%, with more estimates in the 70% to 80% range than was the case for sensitivity. Some estimates were uncertain, and we urge caution in making inferences about the relative differences between the novel devices.

Test failure rates were reported in most studies. The percentage of test failures varied across studies, from 0% to 18%. Test failures were often caused by technical issues relating to the functioning of the test equipment. Limited data were available on use of healthcare resources and costs; the number of repeat tests done; time to interpret device outputs and time to diagnosis and treatment initiation, and patient experience of using novel devices.

Evidence was sparse for children and young people. Two novel devices were evaluated but not in a home setting. A third study is ongoing at two centres in the UK. Preliminary results for one of the study centres are available.

In the base case economic analysis, all six novel devices are estimated to be less costly than respiratory polygraphy, but they are also associated with a small reduction in QALYs. For AcuPebble and Sunrise, the reduction in QALYs may be considered cost-effective compared to the reduction in costs (i.e. INMB > £0 at the £20,000 and £30,000 per QALY thresholds). We estimated that all six novel devices are more costly and provide more QALYs than oximetry, with incremental costs below £10,000 per QALY gained. Probabilistic, scenario and sensitivity analyses highlight the extent of uncertainty, especially in the comparison between the novel devices and respiratory polygraphy.

Limitations

There is high uncertainty over the cost-effectiveness results for the comparison of novel devices with respiratory polygraphy, and the results are sensitive to different sources of data on the accuracy of the novel devices and respiratory polygraphy. There is little uncertainty in the model results for novel devices compared to oximetry. However for both comparators, the analysis relies on data from a clinic, rather than home, setting. Some outcomes have not

been captured, such as effects on patient acceptability, potential delay in diagnosis of other conditions due to false positives, and the impact of contraindications and comorbidities to the use of some devices. There is insufficient evidence to assess the cost-effectiveness of novel home devices in children or in subgroups for whom there are potential equality considerations.

Conclusions

We estimate that the novel devices for home-based sleep studies in adults are a cost-effective alternative to oximetry. It is difficult to assess the cost-effectiveness of the novel devices compared with respiratory polygraphy, and the relative clinical and economic effects of the different novel devices are unclear.

Study registration: PROSPERO CRD42023443437

Funding: NIHR Evidence Synthesis Programme as project number NIHR135895.

Word count: 787

SCIENTIFIC SUMMARY

Background

Obstructive sleep apnoea and hypopnoea syndrome (OSAHS) is a sleep-related breathing disorder in adults and children caused by intermittent narrowing or complete collapse of the upper airway, leading to episodes of reduced (hypopnoea) or absent (apnoea) airflow.

Diagnostic tests for OSAHS are conducted overnight while the patient sleeps, with devices that monitor a range of physiological parameters. Sleep studies can be conducted in the patient's home, or in a hospital sleep clinic or laboratory. Current approaches are oximetry, respiratory polygraphy and polysomnography (PSG). A number of portable novel devices have been developed to facilitate home sleep studies. There is uncertainty over the clinical and cost-effectiveness of these novel devices.

This report was commissioned to inform a Diagnostic Assessment of novel home-testing devices for OSAHS conducted by the National Institute for Health and Care Excellence (NICE). It includes a systematic review of diagnostic accuracy and a health economic evaluation.

Objectives

The objectives of this diagnostic assessment are:

- To conduct a systematic review of the clinical effectiveness (including diagnostic performance) of novel home-testing devices in people with suspected OSAHS.
- To conduct systematic reviews of evidence to inform a health economic evaluation of novel home-testing devices in people with suspected OSAHS: including a systematic review of cost-effectiveness studies, resource use and costs; and a systematic review of health-related quality of life (utility) studies.
- To conduct a health economic evaluation using decision-analytic modelling to assess the incremental cost-effectiveness of novel home-testing devices compared to home respiratory polygraphy or home oximetry in people with suspected OSAHS.

Methods

Systematic review of clinical effectiveness

The proposed methods for the systematic review of clinical effectiveness were reported in advance in a published research protocol (PROSPERO registration number CRD42023443437). The final protocol was published on the NICE website in June 2023.

The search strategy was defined and piloted, and the final searches were conducted in relevant health and medical research databases and trial registers. The databases were initially searched between 22nd and 24th May 2023, and then again on 25th September 2023. No date limits were applied. Predefined inclusion and exclusion criteria were based on the decision problem defined in the NICE Scope. Screening of titles and abstracts was conducted by two reviewers independently, and disagreements were resolved by discussion or with a third reviewer. One reviewer screened the full texts of references judged potentially relevant, and a second reviewer checked the first reviewer's judgement. The reviewers discussed any discrepancies before agreeing whether to include the reference. Where study eligibility remained unclear, we contacted the authors of the study and requested the required information.

Relevant data was extracted from each included study: study design and methods, characteristics of the study population, intervention and comparator(s), and study outcomes. A single reviewer extracted data using a structured and piloted form, which was checked for accuracy and interpretation by a second reviewer and any discrepancies were resolved. Included studies were critically appraised for risk of bias and applicability using the QUADAS-2 instrument. Each study was appraised by one reviewer and checked by a second reviewer, with any disagreements resolved through discussion. The extracted data was used to inform a structured descriptive synthesis of clinical effectiveness. Summary numerical and statistical data were tabulated with accompanying textual description. We considered that clinical heterogeneity was substantial and that meta-analysis would not be appropriate. Likewise, it was not feasible to construct a network meta-analysis.

Systematic review of economic evaluations

We conducted a systematic review of the cost-effectiveness of the novel home-based devices compared to respiratory polygraphy and oximetry. The search strategy was the same as for the clinical effectiveness review, but the outcomes and study design differed. Included studies were full economic evaluations that assessed both costs and consequences for the different novel devices. Outcomes included measures of resource use and costs and health outcomes: life-years or quality-adjusted life-years (QALYs) gained. Economic evaluations not meeting the inclusion criteria and studies that reported on resource use and costs, and health-related quality of life (utilities) were assessed as potential sources of information for our economic model.

External Assessment Group (EAG) independent economic assessment

To evaluate the cost-effectiveness of novel devices in an adult population, we adapted an existing economic model used to inform recent NICE guidelines on the diagnosis and management of OSAHS in people ≥ 16 years of age. The model consists of a decision tree to capture the diagnostic outcomes associated with six novel devices and two comparators (home respiratory polygraphy and oximetry). It has a time horizon of 12 months to capture any delays to the start of treatment (should treatment be offered). A lifetime Markov model is used to estimate the long-term impacts associated with the performance of the devices. It models the risks of cardiovascular events and road traffic accidents for people with OSAHS and includes death from other causes for the total cohort. For all novel devices, in the base case analysis, accuracy data are taken from studies where the devices were evaluated in a laboratory setting.

Results

Systematic review of diagnostic test evaluation and clinical effectiveness

The combined May 2023 and September 2023 searches of literature and other sources identified a total of 290 references subjected to full text screening, 239 were excluded, the majority for reporting an intervention not relevant to the scope. A further 21 references did not report sufficient information to fully inform a screening decision to include or exclude. The remaining 30 publications reported a total 18 studies meeting the inclusion criteria for this systematic review. Of the 18 studies: 12 are relevant to the over 16 years age group and two provided supporting evidence. Three studies are relevant to the 2 -16 years of age group. One study met our inclusion criteria but did not report any results.

Over 16 years age group

The novel devices were compared to home RP in four studies, to hospital sleep-laboratory PSG in five studies, and home-based PSG in one study. There were no eligible studies identified which compared novel devices to home-based pulse oximetry. The Sunrise device was evaluated in three studies, NightOwl in four studies, AcuPebble SA100 in two studies and a single study each was included for, Brizzy, WatchPAT ONE and WatchPAT 300.

Most studies were prospective cross-sectional evaluations of patients referred to specialist sleep services with suspected OSA.. Risk of bias assessments of the studies indicated a low risk of bias for most domains, however there were instances of high or unclear risk of bias for some domains, including bias in the analysis of the index test. Mean age across the studies varied from 41 to 56 years, most commonly around 48 years. With a small number of

exceptions most studies gave limited detail on study population characteristics, notably comorbidities and ethnicity.

Diagnostic accuracy data were available for AcuPebble, Brizzy, NightOwl, and Sunrise but not for WatchPAT ONE or 300. We included two studies of the predecessor version WatchPAT 200U in lieu of evidence for the current version. The sensitivity and specificity estimates vary across the studies and also within studies at different severity cut-offs. Sensitivity was generally high, in the range 80 to 100%, and fell below 80% in just two studies at a high cut-off. In contrast, specificity was more variable with estimates ranging from 25% to 100%, with more estimates in the 70% to 80% range than was the case for sensitivity. In notable cases the available confidence intervals for sensitivity and specificity are wide, indicating greater uncertainty. We urge caution in making inferences about the relative superiority in diagnostic performance between the novel devices.

Test failure rates were reported in many studies. The total percentage of test failures per type of testing (i.e. novel device or comparator) varied across the studies, from 0% to 18%. In most studies the percentage of failed tests per type of testing was less than 10%. Test failures were often caused by technical issues relating to the functioning of the test equipment.

Limited data were available on use of healthcare resources and costs; the number of repeat tests done; time to interpret device outputs and time to diagnosis and treatment initiation, and patient experience of using novel devices.

Children and young people aged 2 -16 years

Two novel devices for the monitoring of mandibular movements during sleep were evaluated: Brizzy with manual scoring and Sunrise with automated scoring. A key limitation of the two studies is the fact that the novel device was not applied in the home setting, due to its concomitant administration with laboratory PSG. The third study is an ongoing study of the AcuPebble SA100 device at two centres in the UK. Preliminary results for one of the study centres are available.

Overall there is limited evidence from the available studies on the outcomes relevant to the decision problem.

Systematic review of economic evaluations

No economic evaluations met our inclusion criteria. We considered five studies with the potential to inform our model structure and parameters, including one conducted for the NICE clinical guideline on the diagnosis and management of OSAHS and obesity hypoventilation syndrome in people over 16. The guideline economic model is the most relevant to our decision problem, and it provided relevant data and assumptions that were used to inform the EAG model structure and parameters.

One company (ResMed) submitted a decision tree model in Excel comparing NightOwl with home respiratory polygraphy. Results suggest that NightOwl is cheaper than home respiratory polygraphy, saving £171 per person, and has better diagnostic performance than home respiratory polygraphy, with lower false positive and false negative rates.

External Assessment Group (EAG) independent economic assessment

In the base case analysis, all six novel devices are estimated to be less costly than respiratory polygraphy, but they are also associated with a small reduction in QALYs. For AcuPebble and Sunrise, the reduction in QALYs is considered cost-effective compared to the reduction in costs (i.e. INMB > £0 at the £20,000 and £30,000 per QALY thresholds). We estimated that all six novel devices are more costly and provide more QALYs than oximetry, with incremental costs below £10,000 per QALY gained.

However, there is a high level of uncertainty over the cost-effectiveness results, apparent from the probabilistic, scenario and one-way sensitivity analyses. In the probabilistic analyses, there are wide and overlapping confidence ranges for the incremental costs and QALYs for each novel device compared with oximetry, which are more pronounced for the comparisons with respiratory polygraphy. For example, the incremental costs for WatchPAT 300 compared with respiratory polygraphy range from -£298 to £235 and the incremental QALYs range from -0.040 to 0.033. This uncertainty is reflected in wide confidence ranges around the INMBs for the novel devices. Scenario analyses indicate that results are sensitive to many assumptions, including the data source used to estimate the performance and failure rates associated with respiratory polygraphy, the proportion of people diagnosed with mild OSAHS who are treated with CPAP, alternative parameterisation of the decision tree (using 4x4 contingency table data), and the impacts associated with false positives.

Limitations

The cost-effectiveness analysis is limited by the availability and quality of data for many of the model components, including limited accuracy data in the home environment and the effects of post-hoc optimisation of thresholds for sensitivity and specificity. Our analysis does not capture the potential for greater patient acceptability of home-based testing with novel devices over the comparators, neither does it include the impact of comorbidities. A key limitation is that our economic model is only relevant to people over the age of 16 years and that we have not been able to conduct subgroup analysis to investigate potential equality issues: for example, related to concerns over the sensitivity of oximetry and other light-based assessment methods for people with darker skin tones.

Conclusions

Implications for service provision

Based on the clinical reviews and economic evaluation, we suggest the following conclusions related to services for people aged over 16 years undergoing home-based testing for suspected OSAHS:

- The estimated cost of the diagnostic pathway is lowest for oximetry and highest for respiratory polygraphy, with the cost for the novel devices lying in between. This is also true for total costs, including costs of OSAHS treatment (if indicated) and costs for care and treatment after cardiovascular events and road traffic accidents, in addition to costs for diagnosis. Estimated total costs are similar for the six novel devices, and differences between these cost estimates are highly uncertain.
- Although oximetry has the lowest cost of the included devices, it has relatively poor sensitivity. In particular, oximetry is estimated to misclassify a high proportion of people with mild OSAHS as not having OSAHS, and a high proportion of people with moderate or severe OSAHS as having mild OSAHS. This implies that patients who would benefit from treatment are not treated or that their treatment may be delayed. All of the novel devices therefore appear to offer relatively good value for money when compared against oximetry.
- All novel devices are estimated to have lower total costs and to produce fewer QALYs than respiratory polygraphy, but these differences are very small and highly uncertain. We emphasise that there is high uncertainty over the relative diagnostic accuracy estimates for all devices and advise caution in interpreting these results.

Suggested research priorities

Research to address significant gaps in the availability and quality of evidence on the diagnostic accuracy of sleep studies for OSAHS, including:

- The diagnostic accuracy of home-based respiratory polygraphy compared against a laboratory PSG reference standard.
- The extent to which diagnostic accuracy evidence from older versions of devices for home-based testing for OSAHS is transferable to new versions.
- The diagnostic accuracy of sleep studies conducted in the home using conventional and novel devices, rather than in a clinic setting.
- Studies of diagnostic accuracy of home-based sleep studies in children and young people under the age of 16, including those with and without comorbidities
- Indirect comparisons between novel and conventional devices and reference standards, with appropriate adjustment for heterogeneity. This would facilitate more robust comparison of results and economic evaluation, but we acknowledge that this is challenging given the high degree of heterogeneity in the current evidence base.
- Alternative study designs for collecting comparative data should be considered, including trials and prospective observational studies. We note that there are also challenges in designing a trial, given heterogeneity of patient populations, variations in practice, and differing opinions on the appropriateness of oximetry.

Further research to provide data to evaluate the clinical and economic effects of home-based sleep studies in children. Key evidence gaps for children include:

- The impact of OSAHS on health related quality of life for children, stratified by OSAHS severity. Studies using preference-based utility instruments with a UK general population value set (e.g. using the CHU9D) would help to inform future economic evaluations.
- The relationship between OSAHS in childhood and long-term effects on health outcomes and well-being, and the extent to which these effects can be assumed to be causative and reversible with appropriate treatment.

Funding: This project was funded by the National Institute for Health Research (NIHR) Evidence Synthesis programme and will be published in full in *Health Technology Assessment*. See the NIHR Journals Library website for further project information. Project number NIHR135895

Study registration: PROSPERO CRD42021266443

Word count: 2532

Table of contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AASM	American Academy of Sleep Medicine
AE	Adverse event
AEs	Adverse events
AHI	Apnoea Hypopnoea index
BNF	British National Formulary
BTS	British Thoracic Society
CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CHD	Coronary heart disease
CI	Confidence interval
CIC	Commercial in confidence
CO_2e	Carbon dioxide equivalent
COPD	Chronic obstructive pulmonary disease
CPAP	Continuous positive airway pressure
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CUA	Cost-utility analysis
CVD	Cardiovascular Disease
EAG	External Assessment Group
ESS	Epworth Sleepiness Scale
EQ-5D-3L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 3 Levels
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
Incr	Incremental
INMB	Incremental net monetary benefit
ITT	Intent to treat
LY	Life-years
LYG	Life-years gain

MAD	Mandibular advancement device
MI	Myocardial infarction
MM	Mandibular movement
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
NMA	Network meta-analysis
NOM	Disposable NightOwl Mini
NOR	NightOwl reusable
NR	Not reported
ODI	Oxygen Desaturation Index
OSAHS	Obstructive Sleep Apnoea (OSA)/Hypopnoea Syndrome (HS)
PAT	Peripheral Arterial Tone
PSA	Probabilistic sensitivity analysis
PSG	Polysomnography
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life year
QALYs	Quality-adjusted life years
QoL	Quality of life
RCT	Randomised controlled trial
RE	Respiratory effort
REM	Rapid eye movement
RERA	Respiratory effort-related arousal
ROC	Receiver operating characteristics
RP	Respiratory polygraphy
RR	Relative risk/risk ratio
RTA	Road traffic accident
SAE	Serious adverse event
SCM	Specialist Committee Member
SD	Standard deviation
SDB	Sleep disordered breathing
SE	Standard error
SF-6D	Short-Form Six-Dimension
SMR	Standardised mortality ratio

TEAE	Treatment-emergent adverse event
TIA	Transient Ischaemic Attack
TSD	Technical Support Document
TST	Total sleep time
UK	United Kingdom
US	United States
VAS	Visual analogue scale
WTP	Willingness to pay

1 BACKGROUND

1.1 Description of the health problem

Obstructive sleep apnoea (OSA) is a sleep-related breathing disorder in adults and children caused by intermittent narrowing or complete collapse of the upper airway, leading to episodes of reduced (hypopnoea) or absent (apnoea) airflow. An apnoea is commonly defined as a complete pause in breathing lasting 10 seconds or more in adults or a minimum duration of two breaths during baseline breathing in children. A hypopnoea is defined as a reduction in breathing lasting for 10 seconds or more in adults or a minimum of 2 breaths in children.[1] The person affected may awaken or their sleep lighten during such episodes, but they may not necessarily be aware they have the condition. Many people with OSA experience episodes of both apnoea and hypopnoea, which is referred to as obstructive sleep apnoea and hypopnoea syndrome (OSAHS).

In adults, common symptoms include excessive daytime sleepiness, snoring, fatigue, morning headaches, impaired concentration and memory. If left untreated OSAHS increases the risk of developing cardiovascular and cerebrovascular complications (e.g. hypertension, atrial fibrillation, heart failure, coronary artery disease, type 2 diabetes, stroke) and death.[2]

In children common symptoms include snoring, restless sleep and hyperactivity. OSAHS in children is associated with cognitive and behavioural problems, poor school performance, cardiovascular morbidity, poor growth and weight gain, decreased health-related quality of life and increased health care utilisation.[3]

1.1.1 Epidemiology

Approximately 25% of the UK population aged 30–69 years have mild to severe OSAHS.[4] The prevalence OSAHS in children aged 2 to 18 years of age is approximately 2 to 4% and is increasing with the rise in childhood obesity.[3]

In adults a number of factors are associated with an increased risk of OSAHS, including being overweight or obese, treatment-resistant hypertension, type 2 diabetes, cardiac arrythmia (particularly atrial fibrillation), stroke or transient ischaemic attack, chronic heart failure, moderate or severe asthma, polycystic ovary syndrome, Down Syndrome, non-arteritic anterior ischaemic optic neuropathy (sudden loss of vision in 1 eye due to decreased blood flow to the optic nerve), hypothyroidism and acromegaly.[2]

One of the leading causes of OSAHS in children is enlarged adenoids and/or tonsils which partially block the airways. The prevalence of sleep disordered breathing is also increased in children with medical conditions including craniofacial anomalies (e.g. Down Syndrome), neurological conditions (e.g. cerebral palsy), and genetic disorders (e.g. sickle cell disease). Children with a history of premature birth are also at increased risk.[3]

1.2 Diagnostic tests for OSAHS

Diagnostic tests for OSAHS are conducted overnight during sleep, hence the term 'sleep study' which is sometimes used. The person being tested for OSAHS is required to sleep as they normally do whilst attached to specialist equipment which monitors a range of physiological parameters associated with sleep disordered breathing. Three main test approaches in use are pulse oximetry, respiratory polygraphy and polysomnography.

- **Pulse oximetry** involves a small monitor (a pulse oximeter) clipped to a finger to record oxygen saturation in the blood and heart rate continuously during sleep. This test can be done at home or in hospital. However, when used as the sole testing medium to diagnose OSAHS it is regarded as less sensitive than other tests.
- **Respiratory polygraphy (RP)** (sometimes referred to as cardiorespiratory polygraphy) records peripheral oxygen saturation, breathing movements, heart or pulse rate, pressure and/or temperature changes resulting from airflow in the respiratory tract, snoring and body position during sleep. Straps are fastened around the chest and abdomen, an oximetry probe placed on a finger and a nasal cannula inserted into the nose, all of which are attached to a recording monitor. The test can be used at home using a portable monitor or in hospital.
- **Polysomnography (PSG)** The PSG is widely considered to be the gold standard test for diagnosing OSAHS and includes the same parameters as RP combined with additional assessment of the quality and duration of sleep from brain activity, eye movement, and muscle tone. PSG is predominantly conducted in a specialist hospital sleep laboratory supervised by a qualified sleep technician. PSG may also be done in the patient's home, using ambulatory equipment, though this is not common practice.

Whilst pulse oximetry and/or RP can be sufficient to inform a diagnosis in patients with a high pre-test probability of having moderate to severe OSA, in some patients, such as those with significant comorbidities, confirmatory testing using PSG may be required. PSG would also be considered where pulse oximetry and/or RP test results are negative but symptoms continue (see section 1.4 for further details of the care pathway).

The signals recorded during a sleep study can be scored manually by a sleep physiologist through visual examination of the data traces recorded whilst the patient slept. Newer types of testing device offer automated scoring using specialist computer software. The software scores the data using a proprietary algorithm resulting in an autogenerated sleep report, with details that may include total sleep time, sleeping position(s), sleep stages, cortical arousals, heart rate, respiratory rate, and snoring patterns. These data are used to inform an assessment of the number of apnoea and hypopnoea episodes which are scored using standard criteria to determine the patient's diagnostic status (i.e. to rule in/rule out OSAHS). In cases where OSAHS is detected the sleep report classifies the level of severity using standard criteria. Many devices with automation provide access to the raw signal data recorded for manual scoring if required.

1.3 Description of the diagnostic technologies under assessment

In the last few years an increasing number of portable testing devices with novel features have become available for diagnosing sleep disordered breathing. These devices have been designed using advances in technology to improve the performance, convenience, and acceptability of home testing for OSAHS. Novel features of devices include the use of wireless electronic sensors in place of multiple wired connections used in PSG or RP testing. It is suggested that these devices may be more comfortable to use whilst sleeping than has previously been the case. This would allow a more natural sleep, with signals more representative of the true physiological state of the patients and hence the presence, absence and severity of the disease. Furthermore, there may be fewer sleep interruptions and an increase in the number of successfully completed tests, in turn reducing the need for repeat tests and in-hospital testing. Furthermore, it is suggested that with fewer wired attachments people will find it easier to correctly fit and operate the device equipment, reducing the need for patient training. The time saved may release staff capacity for other clinical priorities, potentially increasing the volume of patients a clinic can manage routinely.

The following CE marked devices (manufacturer's name in parenthesis) have been identified for inclusion following NICE stakeholder consultation:

- AcuPebble SA100 (Acurable)
- Brizzy (Nomics)
- Sunrise (Sunrise)
- WatchPAT 300 (Zoll/Itamar)
- WatchPAT ONE (Zoll/Itamar)

A further novel device, NightOwl (ResMed), is included; this is awaiting an updated European Union declaration of conformity change.
Although these devices have been identified as being novel they don't necessarily share the same novel features. The devices vary in terms of indication (e.g. some can be used in children and adults; others in adults only), contraindications for use, physiological parameters measured (e.g. respiratory sounds, body movement), lifespan (e.g. single or multi-use), and their connectivity (e.g. ability to send and receive data over the internet). Each device is described briefly and to the best of our understanding in the following subsections.

1.3.1 AcuPebble SA100 (Acurable)

The AcuPebble SA100 is a CE-marked class IIa multi-use device (up to 500 times) for adults only. It consists of a wireless sensor enclosed in a plastic case measuring 2.9cm x (diameter) 1.4cm (height), which is attached to the person's neck below the Adam's apple using double sided adhesive tape (if the sensor is intended to be attached on skin with hair, the hair must be shaved). There is also an option of adding a compatible third-party oximeter. The device records sound generated from physiological body processes, including but not limited to the respiratory and cardiac functions.

The AcuPebble SA100 system requires a compatible mobile device (eg: smartphone or tablet) to install the AcuPebble SA100 mobile app. A fully setup compatible smartphone can be provided by Acurable when the AcuPebble SA100 system is purchased. An internet connection is required during setup (to activate the sleep study on the app) and to finish the study (to upload the signals recorded overnight). Patients can conduct the sleep study without any internet connection, as the setup and upload can be done by the healthcare professional providing the AcuPebble system to the patient. The AcuPebble SA100 web application used by the healthcare professional requires an internet connection to create the sleep study and to review the sleep study results once it is completed. Healthcare professionals can receive an automatically generated report within minutes through the AcuPebble SA100 web application. This report includes the presence and severity of OSA (overall score (rated normal, mild, moderate or severe) based on the Apnoea Hypopnoea Index (AHI) or Oxygen Desaturation Index (ODI)). Outputs include: diagnosis information based on AHI 3% desaturation criteria (AASM guidelines), diagnosis based on AHI 4% desaturation criteria (AASM guidelines), diagnosis based on ODI 3% desaturation criteria, diagnosis based on ODI 4% desaturation criteria, classification of apnoea events, cardiac rate, respiratory rate, snoring evaluation, acoustic derived airflow, acoustic derived relative

desaturation and activity. The company have advised that the device can be used during pregnancy and that they plan to extend the intended use for AcuPebble to children. The paediatric version of AcuPebble has different algorithms to take into account different scoring rules for children, and an optional, slightly larger sensor in order to be in line with guidance about choking hazards. A clinical validation study of the paediatric version is currently ongoing in the UK and is included in the systematic review of this report (see sections 4.3, 4.8, 4.9, 4.10).

1.3.2 Brizzy (Nomics)

The Brizzy is a CE-marked class IIa device and is indicated for use in the screening and diagnostic evaluation of sleep breathing disorders in children and young people (over 3 years old) and adult patients. The intended use is as a portable sleep recorder for detecting sleep apnoea syndrome and for monitoring its treatment. The technology can be used at home and in sleep clinics.

The Brizzy consists of a recording device hub to which electromagnetic sensors are connected via cables. The Brizzy user manual advises that if the device is to be used on a child, a responsible adult should be clearly shown how to install the cables to avoid any risk of strangulation. Furthermore, it recommends securing the cables with either adhesive tape or by slipping an additional shirt over the device and the sensor cables. The sensors, which are fixed on the chin and forehead using adhesive tape, measure jaw activity signal (referred to as “Jawac” by the company): mandibular movement, mouth opening, and nervous gnathic twitch. A pulse oximeter or an electrocardiogram (ECG) with 3 electrodes are optional add-ons. The central device hub is attached to a fastening belt and is worn around the waist during sleep. The company advise having at least 4 hours of recording.

It is currently unclear how the device will be distributed between the user and sleep clinic. Once the device is returned, a physiologist uploads the study to the web portal (CERES software) using a wired USB connection to produce an automated report which can aid in the diagnosis of sleep breathing disorders or be used for further clinical investigation. Raw data from the recorded study can be accessed and manually scored by healthcare professionals if needed.

The Brizzy device measures an output called the ‘respiratory events index JAWAC’ (REI_JAWAC). Other outputs measured by the device are total sleep time (TST), sleep fragmentation, respiratory effort, number and frequency of apnoea events (broken down by

type: obstructive, central, or mixed), positional analysis (total sleep time in supine versus non-supine position, REI-JAWAC in supine versus non-supine position), and mandibular (lower jaw) activity. If using an add-on oximeter or an ECG, the device can also measure heart rate, oxygen saturation (SpO_2), ODI, and an ECG graph. The device provides an automated qualitative output of OSAHS severity based on the REI_JAWAC measure using the scoring criteria described earlier in section 1.2)

The device has a lithium polymer battery (rechargeable by USB), and the storage capacity and battery life allow for recording several nights if used without oximetry or ECG. The company states that there are no known contraindications, and the technology can be used during pregnancy. However, caution is advised when used by people with restless leg syndrome as the number of apnoea events can be overestimated in this group. Parkinson's disease and temporomandibular disorders could impact jaw movements and test results should therefore be interpreted accordingly. The central hub and JAWAC sensors are reusable, made of recyclable Acrylonitrile Butadiene Styrene plastic. The fastening belt may be machine-washed and reused.

1.3.3 NightOwl (ResMed)

NightOwl is suitable for adults and children aged 13 years or over. In terms of CE marking, the software is Class IIa but the sensor itself is Class I. The company have indicated that the NightOwl device to be commercialised in the UK has a built-in battery that allows for 10 nights of recording. The battery is usable for three years from manufacturing. Therefore, the device can be used intermittently and is not required to be used on consecutive days. The device is not rechargeable and, after use, it is to be discarded, ideally by any existing recycling programme for electronic waste.

NightOwl consists of a photoplethysmography (PPG) sensor and an accelerometer and is attached to the fingertip or forehead using an adhesive. The device measures peripheral arterial tone (PAT), oxygen saturation, actigraphy (body movement), and pulse rate. A probabilistic model determines a respiratory event from the co-occurrence of oxygen desaturation, vasoconstriction manifested as a PAT channel decrease, and a pulse rate increase.

A 3G or 4G smartphone with the NightOwl Companion app installed is needed for data to be automatically uploaded to the analytics platform once the test is concluded. An automated report provides diagnostic information and the severity of OSA (AHI and AHI severity

category). Other outputs include pulse rate, oxygen saturation (SpO_2), ODI, sleep/wake states (TST), presence or absence of a substantial changes in PAT that may be caused by the presence of irregular heart rhythms, information on the location of desaturations and signal artifacts. The raw data can be accessed on the analytics platform and manually scored, if needed. Performance of the device can be adversely impacted if a person has changes in their sympathetic response or has reduced blood flow to the fingers e.g., due to use of drugs that affect the autonomic system (for example, alpha-adrenergic antagonists) or because they have peripheral vascular disease (for example, secondary Raynaud's disease). The device should not be used on patients with known severe ventricular extrasystole (VES) as this is likely to lead to insufficient clean data segments and therefore a failed test. The company offers a service to send the device directly to the patient, though sleep centres can manage delivery if they prefer. In addition, healthcare professionals can specifically request a patient to pick up the device in person, although this is not the preferred method.

1.3.4 Sunrise (Sunrise)

The Sunrise is a CE-marked class IIa single-use device for adults and children aged over three years of age. It consists of wireless sensor, measuring 46.5mm x 20.0mm x 5.6mm, that is placed on the chin. Both patient and health care professional Sunrise user manuals warn against placing the device in the mouth or swallowing it as this could cause suffocation, and to keep the device away from children and pets to avoid accidental swallowing..[5 6] Furthermore the user manuals state that the device may not be suited for bearded users and advise close shaving or usage of provided adhesive bandages to ensure optimal adhesion. The device measures mandibular movements to estimate interruptions and breathing during sleep. An internet connection (wi-fi or 3G/4G) and a smartphone is needed for data to be transferred from the device to a secure cloud platform where it is automatically analysed. An automated report provides OSA severity scoring (non-OSA, mild, moderate, or severe) based on AHI and/or obstructive respiratory disturbance index (ORDI). Other outputs include sleep/wake states (TST), sleep stages, respiratory events (AHI, RDI, OAHI), central AHI (CAHI), obstructive respiratory disturbance index (ORDI), respiratory effort related arousal (RERA) index, respiratory effort, awakening and arousal index, SpO_2, heart rate, position changes index and sleep bruxism (extent of teeth grinding during sleep). Raw data can be accessed through an online web portal and manually scored, if needed.

The company offers a service to send the device directly to the patient. Alternatively, sleep services can manage delivery or have the patient pick up the device during their consultation

appointment. Sunrise provides a prepaid and self-addressed envelope to return the device for disposal in accordance with the European Union's Waste Electrical and Electronic Equipment (WEEE) directive. The device cannot be used for people with conditions affecting the rotation of the condyle (part of the jawbone) in temporo-mandibular joint.

1.3.5 WatchPAT 300 (Zoll/Itamar)

WatchPAT 300 (WP300) is a CE-marked class IIa multi-use device with some single use components. It is suitable for adults and children aged 12 and above and consists of a wrist worn device, finger probe and chest sensor. It measures a proprietary peripheral arterial tone signal (PAT), heart rate, oximetry, actigraphy (body movement), body position, snoring and chest motion. Snoring and body position safety and effectiveness are validated for an adult population only. After the sleep test, if the device belongs to the NHS trust, it needs to be returned to clinical setting where staff download the data via USB connection and analyse the results using the zzzPAT software. Otherwise, the WatchPAT Direct service provides delivery services directly from and return to the manufacturer, who sends the results to the sleep service.

An automated report provides diagnosis and severity of OSA (no apnoea, mild, moderate or severe apnoea) based on AHI and OHI. Additional outputs include: AHI, CAHI, RDI, ODI, sleep/wake states, sleep stages, body position, snoring, heart rate, chest movement, SpO_2 and actigraphy. Raw data is accessible through the WatchPAT software platform, where it can be reviewed and manually edited to adjust the test scoring. The manufacturer states that the WP300 is not indicated for use in people with injuries, deformities or abnormalities that may prevent proper application of the device and should not be used in people on medication including alpha blockers or short acting nitrates (taken less than 3 hours before the study), or people with a permanent pacemaker (atrial pacing or VVI without sinus rhythm), or people with sustained non-sinus cardiac arrythmias (in the setting of sustained arrhythmia the WatchPAT's automated algorithm might exclude some periods of time, resulting in a reduced valid sleep time. A minimum valid sleep time of 90 minutes is required for an automated report generation). The WatchPAT 300 is also not indicated for children who weigh less than 65 lbs. Additional precautions are stated for people aged 12 to 17 years of age, including patients with severe comorbidities such as Down Syndrome, neuromuscular disease, underlying lung disease or obesity hypoventilation to be considered for laboratory polysomnography (PSG) rather than a home sleep testing.t is recommended that the physician ensures the patient and his/her guardian are aware that the use of specific drugs and other substances used to treat ADHD, antidepressants, corticosteroids,

anticonvulsants, use of caffeine, nicotine, alcohol and other stimulants might interfere with sleep and affect the sleep study's conditions.

1.3.6 WatchPAT ONE (ZoII/Itamar)

WatchPAT ONE is a single-use version of WatchPAT 300. Unlike WatchPAT 300, an internet connection (wi-fi or 3G/4G) and a smartphone is needed to transfer sleep data from the device to a webserver for automated scoring using zzzPAT software. Raw data can be accessed through the WatchPAT software and manually scored if needed. The company state they are in the process of setting up a free-of charge recycling service for the device.

1.3.7 Measuring the severity of OSAHS

One of the key outcomes of a sleep study is the ability to attribute breathing disturbances as predominantly obstructive or non-obstructive in pattern. The former is indicative of sleep apnoea resulting from obstructions to breathing in the upper airways (i.e. OASHS) whilst the latter indicates sleep apnoea caused by brain signal disturbances, a condition known as central sleep apnoea. Central sleep apnoea is less common than OSAHS and has a different clinical management pathway. As will be explained in section 2, this diagnostic assessment report is restricted to obstructive rather than central sleep apnoea.

Commonly used diagnostic scoring criteria for OSAHS include the Apnoea Hypopnoea Index (AHI) and the Oxygen Desaturation Index (ODI). The AHI measures the number of apnoeas recorded per hour of sleep, averaged across the duration of the sleep study. The ODI measures the number of episodes of oxygen desaturation recorded per hour (using 3% or 4% desaturation criteria), averaged across the sleep study. The American Academy of Sleep Medicine (AASM) has defined severity criteria for adults and children.[7 8]

In people aged ≥16 years the AHI and ODI scoring thresholds for OSAHS severity are:

- Normal/no OSAHS: <5 events per hour
- Mild OSAHS: ≥5 to <15 events per hour
- Moderate OSAHS: ≥15 to <30 events per hour
- Severe OSAHS: ≥30 events per hour.

In people <16 years of age the corresponding AHI and ODI scoring criteria are lower to reflect the fact that children have a different physiology, i.e. different normal state, to adults and can experience morbidity at a lower AHIs.

- Normal/no OSAHS: < 1 event per hour
- Mild OSAHS: ≥1 and <5 events per hour
- Moderate OSAHS: ≥5 and <10 events per hour

Severe OSAHS: ≥10 events per hour

Other available scoring indices include the Respiratory Disturbance Index (RDI), which measures the average frequency of apnoea and hypopnoea per hour of recording using a portable monitor (e.g. at home) or per hour of sleep time using PSG (e.g. in hospital). It is important at this point to acknowledge the distinction between total sleep time and total recording time in a sleep study. Some portable monitors measure only the total recording time, which may include periods when the patient was awake. Because the total recording time often exceeds the actual sleep time of the patient, RDI from some portable monitors underrepresents the severity of OSAHS. PSG, in contrast, does have the ability to estimate total sleep time. The RDI also records the number respiratory effort-related arousals (RERAs), defined as a reduction in airflow with resultant arousal that doesn't meet the criteria for hypopnoea.

Table 1 shows a classification of sleep study types devised by the American Academy of Sleep Medicine (AASM), ranging from type 1 (sleep laboratory PSG) to type 4 (devices measuring 1 or 2 parameters). [1] The classification reflects the evolution of diagnostic testing for OSAHS; the ascending types reflecting the increasing use of limited channel portable device testing outside of hospitals (i.e. types 2 to 4). The classification (or adaptations thereof) is widely used in clinical practice and in the scientific literature.

Table 1 AASM classification of sleep study types

Sleep study	Description
Type I	Sleep laboratory PSG (gold standard). Multi-channel recordings including EEG, EOG, EMG, ECG, full set of respiratory measures, position and movement sensors, video and audio, and CO_2. Diagnostic outputs include sleep staging, sleep duration, arousals, AHI, OAHI, SpO_2 and CO_2 measures. Attended by qualified sleep technicians.
Type 2	Use the same monitoring sensors as sleep laboratory PSG (Type I) but are unattended and can be performed outside a sleep laboratory (e.g. at home).
Type 3	Use portable devices that measure limited cardiopulmonary parameters; two respiratory variables (e.g., effort to breathe, airflow) oxygen saturation, and a cardiac variable (e.g., heart rate or electrocardiogram). Can be performed in hospital or in the home (unattended).

Type 4	Use portable devices that measure only 1 or 2 parameters, typically oxygen saturation and heart rate, or in some cases, just air flow. Can be performed in hospital or in the home (unattended).
Other	Examples include devices that measure peripheral arterial tonometry (PAT) or mandibular jaw movements.

Source: Kapur et al, 2017[1]
AHI Apnoea Hypopnoea index; American Academy of Sleep Medicine (AASM); CO_2, carbon dioxide; ECG electrocardiogram; EEG electroencephalogram; EMG electromyogram; EOG electrooculogram; PSG polysomnography; SpO_2, oxygen saturation

The AASM classification doesn't necessarily include novel approaches to diagnostic testing. For example, some approaches use peripheral arterial tonometry (PAT) in place of measuring some of the parameters used in type 1 to 4 devices. The classification also doesn't explicitly account for the recent proliferation of wearable devices and non-contact systems. Some commentators have therefore appended an 'other' category to represent these newer technologies.

1.4 Care pathway

1.4.1 Care pathways in adults

People with suspected OSAHS may present to primary care with a range of self-reported symptoms including snoring, unexplained excessive sleepiness, tiredness, fatigue, choking during sleep, and insomnia. The patient's symptoms form the basis of a sleep history taken by the doctor. Additionally, instruments such as the self-administered Epworth Sleepiness Scale (ESS), the Epworth Sleepiness Scale (ESS) for Children and Adolescents (ESS-CHAD) and the STOP-Bang questionnaire may be also used. Other factors taken into account when making a referral include comorbidities and any patient occupational risk (e.g. vocational driving). If there is sufficient clinical suspicion of OSAHS based on these assessments the doctor will refer the patient to a specialist sleep service for further investigation.

Once referred specialist assessment of the patient's symptoms and needs will be undertaken to determine which approach to diagnostic testing is most appropriate. Factors taken into consideration include whether the patient has contraindications to any of the testing devices; the presence of multiple comorbidities; level of clinical suspicion of OSAHS; the likely severity of OSAHS and patient preferences, including practical considerations about the feasibility of home or inpatient testing. Expert clinical advice to the EAG is that this is commonly done by a consultant in respiratory medicine / sleep respiratory medicine, but in some centres, specialist nurses triage patients for diagnostic testing.

The NICE guideline on obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s (NG202, 2021)[2] recommends the following diagnostic approaches in people with suspected OSAHS:

- **Home RP** as the initial testing strategy, where practical.
- If access to home RP is limited, **home oximetry alone** may be used (with the caveat that this may be inaccurate for differentiating between OSAHS and other causes of hypoxaemia in people with heart failure or chronic lung diseases).
- **Hospital RP** can be used when home RP and home oximetry are impractical or where further RP monitoring is required.
- **PSG** can be used if RP results are negative but symptoms continue.

In addition to NICE NG202,[2] the American Academy of Sleep Medicine (AASM) Clinical Practice Guidelines on diagnostic testing for adult OSA is also widely used by sleep specialists in England.

Although both the AASM and NICE guidelines recommend home RP for diagnosing OSAHS (in uncomplicated cases), this approach is regarded has having limitations. The RP monitors include multiple wired components attached to the person as they sleep which can be uncomfortable, potentially interrupting sleep, affecting the patient's sleep position and therefore natural sleep patterns. If total sleep time is less than the minimum time required to complete the test (usually at least four hours) the results may be inaccurate or inconclusive. Consequently, one or more re-tests may be required, adding to costs and delaying a definitive diagnosis and commencement of treatment, if needed. If a successful home RP result cannot be achieved, then an in-hospital sleep study (if available) may be required. Expert clinical advice suggests, however, a reduction in hospital sleep testing capacity since the COVID-19 pandemic, creating greater reliance on home-testing as the primary approach to sleep testing. The criteria for assessing patient suitability for home testing has therefore widened and now includes some types of patients who previously would have received hospital PSG as the primary testing approach.

Sufficient training is needed to ensure correct use of testing devices in the home, with cost implications in terms staff time to provide instruction and support to patients. Furthermore, home testing equipment must be collected from, and later returned to, the hospital before diagnosis can be confirmed. This is often the patient's responsibility and requires means of transport and their availability during clinic operating hours. Patients will therefore incur transport costs, and may even need time off work, with consequent loss of earnings for

some. Transport costs will inevitably be excessive for patients whose nearest sleep clinic is some distance away. Some clinics send equipment to patients via the postal service or by courier, accumulating substantial costs to the NHS. Clinical experts consulted commented that home testing equipment is not always returned, leaving hospitals to cover the cost of replacements.

Some novel devices can be sent directly to the patient by the manufacturer. A direct delivery service could improve access to home testing thereby reducing excessive waiting times for home-testing currently seen in practice, particularly since the COVID-19 pandemic. Potentially this could reduce time to diagnosis, leading to more timely treatment initiation and symptom improvement. Furthermore, a delivery service managed by the manufacturer rather than the NHS would allow NHS staff to focus on other priorities.

1.4.2 Care pathways in children

In 2023 The British Thoracic Society (BTS) published a clinical guideline for diagnosing and monitoring paediatric sleep-disordered breathing.[9] The guideline, the first in the UK specifically covering OSAHS in children, proposes a number of care pathways covering initial assessments of symptoms and types of sleep study, as appropriate to the needs of the patient. A key factor governing choice of pathway is the presence or absence of comorbidities.

2 DEFINITION OF THE DECISION PROBLEM

The NICE scope for this diagnostic assessment includes the following decision question, developed and prioritised in consultation with relevant stakeholders.

Decision question

Do novel home-testing devices for OSAHS represent a clinically and cost-effective use of NHS resources?

The following sub-sections define the parameters relevant to the decision problem.

2.1 Population and relevant subgroups

The population of interest is people suspected to have OSAHS (who are considered suitable for a home sleep study) in both adults (over 16 years old) and children (2-16 years). This assessment reviews the evidence for these defined adult and child populations separately as not all devices included in this assessment are indicated for use in children or young people. The age ranges align with the NICE Guideline for the diagnosis and management of OSAHS and obesity hypoventilation syndrome in over 16s (NG202)[2] and the British Thoracic Society (BTS) guideline for diagnosing and monitoring paediatric sleep-disordered breathing.[9]

Subgroups of interest are:

- People with chronic obstructive pulmonary disease (COPD)
- People who have neuromuscular disorders
- People from black, Asian and minority ethnic backgrounds
- For children and young people aged 16 and under, with and without comorbidities (as defined in the BTS guidelines for the diagnosis of sleep disordered breathing in paediatrics)
- Pregnant women and pregnant people.

2.2 The intervention

As described in section **Error! Reference source not found.** of this report, six novel devices were identified for inclusion in the NICE scope. They were selected on the basis of having novel features that have the potential to demonstrate cost-effectiveness in relation existing standard practice, and because they are commercially available in England. The devices are: AcuPebble SA100 (Acurable), Brizzy (Nomics), NightOwl (ResMed), Sunrise (Sunrise), WatchPAT 300 (Zoll/Itamar) and WatchPAT ONE (Zoll/Itamar).

Four of these devices are indicated for use in children and/or adolescents as well as for use in adults: Brizzy (for patients over three years of age),[10] NightOwl (individuals aged 13 or older),[11] Sunrise (for patients over three years old),[5] and WatchPAT 300 (for 12 years and older, except for the Central Apnoea-Hypopnoea index parameter which is indicated for patients 17 years and older).[12] In the child population (2-16 years for this assessment) CO_2 monitoring may be used alongside the interventional device.

2.3 The comparator

The comparator technologies are RP or home pulse oximetry devices currently in use across the NHS in England for home testing. Home RP devices include at least four channels, for example to record oximetry, breathing rate, apnoeas and hypopnoeas, snoring and body position. Examples of RP devices used in home testing include Alice NightOne (Philips), ApneaLink Air (ResMed), Embletta MPR PG (Stowood), NoxT3s (ResMed); these examples are illustrative and other branded devices are also in use within the NHS.

Additionally, in the child population (2-16 years for this assessment) CO_2 monitoring may be used alongside the comparator technology of either home RP or home pulse oximetry.

2.4 Outcomes

The outcomes of relevance to the decision problem are grouped into the four categories below.

Intermediate outcomes can include measures of diagnostic accuracy (e.g. sensitivity and specificity) and the ability to assess severity of OSAHS; the time taken to interpret device outputs, reach a diagnosis, and/or start treatment; test failure rate and the number of repeat sleep studies done; use of healthcare resources, e.g. hospital admissions and use of pharmacological and non-pharmacological interventions for the management of OSAHS.

Clinical outcomes evaluate morbidity and mortality.

Patient reported outcomes evaluate aspects that have an impact on patients on a personal and/or functional level. They reflect ease of use for the patients and their carers, with the extent to which assistance from a healthcare professional is needed to set up and operate the device. Acceptability of the device is relevant such as being more comfortable to wear, and the potential to reduce anxiety and stress and provide a more representative night of sleep data.

Costs outcomes are considered from an NHS and Personal Social Services perspective.

2.5 Overall aims and objectives of the assessment

The aim of this diagnostic assessment is to estimate the clinical effectiveness and cost-effectiveness of novel home-testing devices in people with suspected OSAHS. The results will inform NICE guidance to the NHS on use of this diagnostic technology.

The objectives of this diagnostic assessment are:

1. To conduct a systematic review of the clinical effectiveness (including diagnostic performance) of novel home-testing devices in people with suspected OSAHS.
2. To conduct systematic reviews of evidence to inform a health economic evaluation of novel home-testing devices in people with suspected OSAHS. We will conduct a systematic review of cost-effectiveness studies and of health-related quality of life (utility) studies. We will take a systematic approach to identifying relevant resource use and cost data relating the diagnosis and treatment of OSAHS.
3. To conduct a health economic evaluation using decision-analytic modelling to assess the incremental cost-effectiveness of novel home-testing devices compared to home RP or home oximetry in people with suspected OSAHS.

3 METHODS OF CLINICAL AND DIAGNOSTIC ASSESSMENTS

The proposed methods to produce the systematic review of clinical effectiveness were reported *a priori* in a published research protocol (PROSPERO registration number CRD42023443437). The final protocol was published on the NICE website shortly after the final scope of this assessment was published in June 2023. The following sub-sections report further detail on the methods used.

3.1 Identification of studies

Comprehensive, systematic literature search strategies were designed and tested by an experienced information specialist from the project team to inform the systematic review of diagnostic test evaluation and clinical effectiveness (see section 4), and systematic reviews of cost effectiveness evidence and economic model input parameters (see section 5).

The draft search strategy for diagnostic test evaluation and clinical effectiveness was piloted in MEDLINE, following which revisions were made and a final version produced. The final search strategy was implemented in the following health and medical research databases and trials registers:

- Ovid MEDLINE(R) ALL
- Ovid Embase Classic + Embase
- Cochrane Library (Wiley) for the Cochrane Register of Controlled Trials (CENTRAL) and the Cochrane Database of Systematic Reviews (CDSR)
- Web of Science for Science Citation Index Expanded (SCI-EXPANDED) and Conference Proceedings Citation Index – Science (CPCI-S)
- International HTA Database (database.inahta.org)
- CRD Database for the Database of Abstracts of Reviews of Effects (DARE) and the NHS Economic Evaluations Database (NHS EED) (crd.york.ac.uk/CRDWeb)
- Epistemonikos (epistemonikos.org)
- ClinicalTrials.gov
- BePartOfResearch (formerly the UK Clinical Trials Gateway)
- NIHR Clinical Research Network
- OpenGrey
- PROSPERO register of systematic reviews

The databases were initially searched between 22nd and 24th May 2023, and were then searched again on 25th September to identify any relevant references added to the databases between May and September. No date limits were applied.

We searched for conference abstracts for the last three years only, on the assumption that after three years studies presented atconferences would likely have been published in full and identified by our database searches. We identified several specific sleep medicine conferences from which the abstracts are indexed in the Embase, MEDLINE and/or SCI-Expanded databases listed above. For this reason we did not hand search them separately. The conferences were: Clinical Update Sleep™ (Guy's and St Thomas' NHS Foundation Trust), Sleep Europe (European Sleep Research Society), ESRS Congress (European Sleep Research Society), The Sleep and Breathing Conference (European Respiratory Society and the European Sleep Research Society), ERS International Congress (European Respiratory Society), World Sleep (World Sleep Society), and Sleep (American Academy of Sleep Medicine and the Sleep Research Society).

To identify any further relevant primary studies we also searched:

- the reference lists of the included studies,
- the reference lists of relevant systematic reviews identified in the database searches,
- the manufacturer and distributor evidence submissions to NICE,
- any references brought to our attention by our clinical experts and NICE specialist committee members.

Further details and the full search strategies applied to each database are reported in *Appendix 1*

3.2 Inclusion and exclusion criteria

The predefined inclusion and exclusion criteria are based on the decision problem as outlined earlier in section 2, and are described below. An extended PICO tabulation of these criteria is included in *Appendix 2* . This table is the basis of the worksheet we used to systematically apply the criteria to each study screened.

3.2.1 Population

The relevant population is people with suspected OSAHS.

The population is stratified by the following age groups:

- People over 16 years old
- Children and young people between 2 and 16 years of age (NB. some technologies included for assessment are not indicated for use in children or young people. None of the technologies are indicated for children aged under 2 years).

Where data permits, the following subgroups were considered:

- People with COPD
- People who have neuromuscular disorders
- People from black, Asian and minority ethnic backgrounds
- For children and young people aged 2 to 16 years, with and without comorbidities (as defined in the BTS's guidelines for the diagnosis of sleep disordered breathing in paediatrics)
- Pregnant women and pregnant people

3.2.2 Interventions

The following technologies were eligible for inclusion:

- AcuPebble SA100 (Acurable)
- Brizzy (Nomics)
- NightOwl (ResMed)
- Sunrise (Sunrise)
- WatchPAT 300 (Zoll/Itamar)
- WatchPAT ONE (Zoll/Itamar)

For children and young people (2-16 years), use of the interventions may be alongside CO_2 monitoring.

3.2.3 Comparators

For people over 16: home RP or home oximetry (can include home test devices currently used in clinical practice but cannot include any of the named novel devices in 5.2 above). For people with COPD, home oximetry alone is not recommended and will therefore not be considered a suitable comparator for this subgroup.

For children and young people aged 2 to 16 years: home RP or home oximetry. CO_2 monitoring may be used alongside these technologies.

Home RP or home oximetry can include home test devices currently used in clinical practice but cannot include any of the named novel devices included as interventions (above).

The reference standard can include in-hospital PSG, PSG done outside hospital or RP done in a healthcare setting (rather than at home).

3.2.4 Study design

We did not limit inclusion by type of study design, because a range of study designs could potentially be used to assess the clinical effectiveness of novel home testing devices.

3.3 Inclusion screening process

At the first stage of screening, two reviewers independently applied the above criteria to the titles and abstracts using an inclusion/exclusion worksheet (see *Appendix* 2). Any disagreements between reviewers in judgements about study eligibility were resolved through discussion or with the opinion of a third reviewer where necessary.

At the second stage of screening one reviewer screened the full texts of references judged potentially relevant on title and abstract screening. A second reviewer checked the first reviewer's judgement on eligibility based on the full text. The reviewers discussed any discrepancies in judgement and before agreeing a final decision to include or exclude the reference. Where study eligibility remained unclear due to missing information to inform reviewers' judgement, we contacted the authors of the study and requested the required information.

To ensure consistency between reviewers in the application of the inclusion/exclusion criteria, the EAG developed decision rules to be followed when screening studies with complex characteristics or ambiguously reported procedures.

3.4 Data extraction strategy

Relevant data was extracted from each included study, covering details of the study design and methods, the socio-demographic and health characteristics of the study population, the intervention, and comparator(s), and the study outcomes. Each study underwent data extraction by a single reviewer using a structured and piloted data extraction form (see *Appendix 3* for the data extraction template). Data was extracted from each publication available for a given study, including journal article supplemental information where available. The extracted data was checked for accuracy and interpretation by a second reviewer and any discrepancies between them were resolved through discussion. The

finalised agreed data extraction form for each study informed the synthesis of clinical effectiveness (see section 3.6) (NB. The finalised data extraction forms are available on request).

3.5 Critical appraisal of study methodology

Included studies were critically appraised for risk of bias and applicability using the QUADAS-2 instrument.[13] Each study was appraised by one reviewer and their judgements checked by a second reviewer, with any disagreements between reviewers resolved through discussion. The results of our critical appraisal are summarised in section 4.4 with further detail given in Appendix 5.

3.6 Method of data synthesis

The data extracted from the included studies was used to inform a structured descriptive synthesis of clinical effectiveness. Summary numerical and statistical data from the included studies are tabulated with accompanying textual description, for each outcome measure in turn (see sections 4.5 to 4.7).

In the protocol for systematic review of clinical effectiveness we stated our intention to assess the appropriateness and feasibility of meta-analysis, based on considerations such as whether sufficient data were available and whether heterogeneity across the included studies could be considered acceptably low. Having made these assessments our position is that clinical heterogeneity is substantial enough that meta-analysis would not be appropriate. The respective novel devices are distinct in terms of their design, technology and clinical application and for this reason it would be inappropriate to statistically pool outcomes across different devices as if they were interchangeable. We did, however, consider that a set of pairwise meta-analyses for each respective novel device would be appropriate (e.g. novel device 1 vs comparator, novel device 2 vs comparator). This approach has been followed in previous systematic reviews in this topic area.[14] However, as will be reported in Section 4, pairwise meta-analysis was not feasible due to the low number of relevant studies included.

For the same reasons mentioned above, it was not feasible to construct a network meta-analysis to indirectly compare the novel devices to inform an incremental assessment of cost effectiveness.

4 RESULTS OF CLINICAL AND DIAGNOSTIC ASSESSMENTS

4.1 Quantity of research available

Our initial literature searches, done in May 2023 (see section 3.1 and Appendix 1) identified a total of 3541 potentially relevant references after duplicate references were removed. Independent screening of titles and (where provided) abstracts by two reviewers determined that 3313 of these references did not meet the inclusion criteria, whilst the full text of the remaining 228 references were obtained for further screening. Of the 228 full texts examined, 201 did not meet our inclusion criteria and were excluded. Updated literature searches conducted in late September 2023 (see section 3.1 and Appendix 1) identified a further 144 unique titles and abstracts, of which 130 did not meet our inclusion criteria. Of the 14 full texts examined, 12 did not meet our inclusion criteria and were excluded. In total 15 full texts identified through our initial and updated database searches meet our inclusion criteria.

Identification of studies via other methods included:

- Searching company submissions for relevant documents and references. This yielded 36 full texts, of which 16 did not meet our inclusion criteria.
- Citation checking of relevant systematic reviews (one full text identified, which did not meet our inclusion criteria)
- Citation checking of included studies (three full texts identified, one of which meets our inclusion criteria)
- Contacting study authors. Overall, authors of 34 full texts were contacted for clarification as to whether their paper met our inclusion criteria. Authors of 15 responded, some of whom provided references they believed relevant to our review. This resulted in an additional nine full texts being identified, of which seven did not meet our inclusion criteria.

Of the 49 full texts identified via other methods, 15 meet our inclusion criteria.

In summary, the combined May 2023 and September 2023 searches of literature and other sources identified a total of 290 references subjected to full text screening, 239 were excluded, the majority for reporting an intervention not relevant to the scope (reasons for exclusion are given in Appendix 2). A further 21 references did not report sufficient information to fully inform a screening decision to include or exclude. The remaining 30 publications reported a total 18 studies meeting the inclusion criteria for this systematic review. These are listed in Table 59 in Appendix 4. The PRISMA 2020 flowchart in Figure 1 illustrates the flow of studies during the stages of screening.

Of the 18 included studies:

- 12[15] studies are relevant to the over 16 years age group in this diagnostic assessment and an additional two studies provide supporting evidence (Table 2).
- Three studies are relevant to the 2-16 years of age group (Table 4).
- The remaining study, NCT0476473 2021,[16] compared NightOwl to PSG, with an age-related inclusion criterion of “13 years and older”. However, results are not yet available even though the study has completed. Attempts by the EAG to obtain clarification or data from the study investigators have so far proved unsuccessful. For the sake of clarity, we will not refer to this study in the remainder of this chapter.

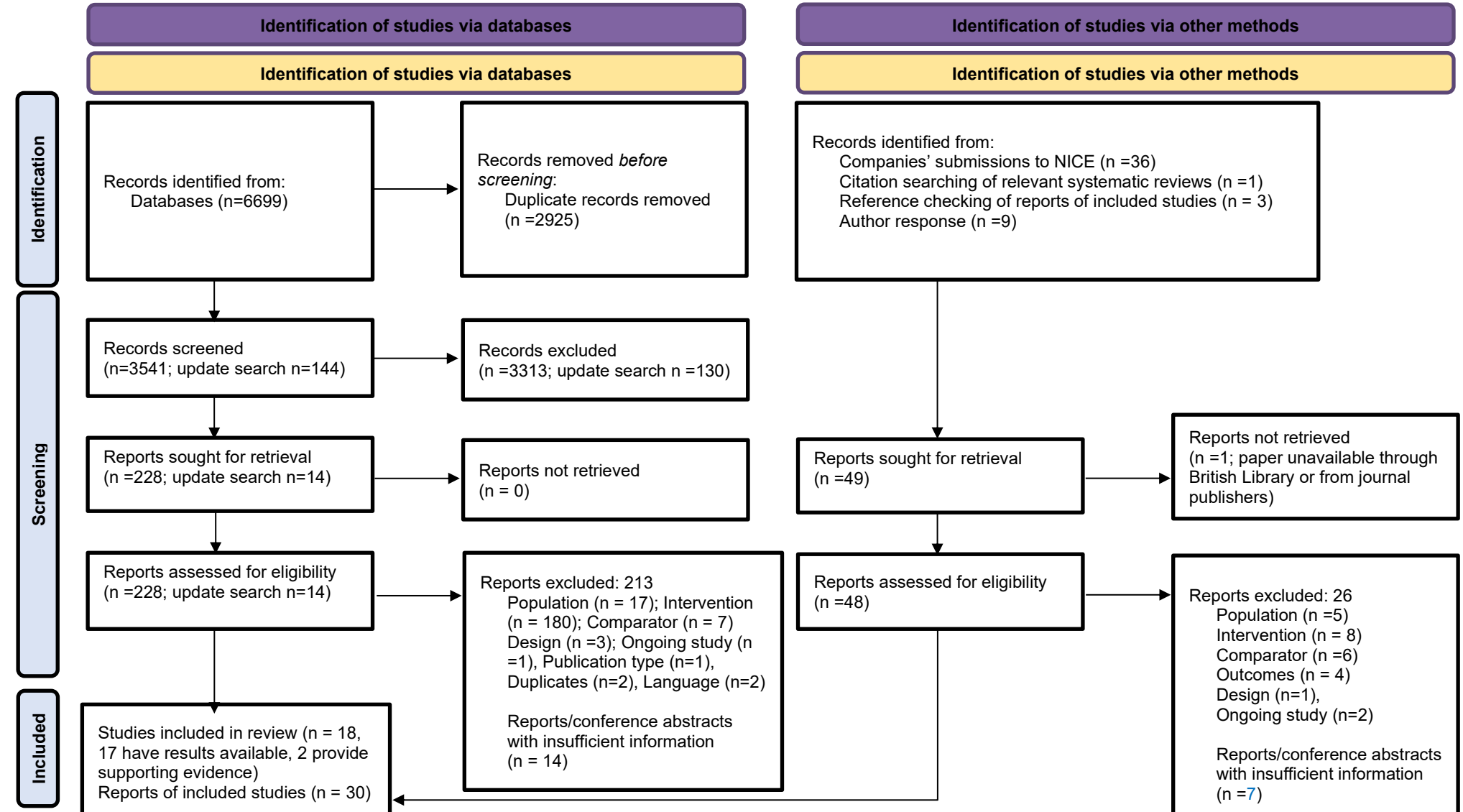


Figure 1 PRISMA 2020 flowchart

4.1.1 Studies of people over 16 years of age

Table 2 lists the 12 studies relevant to the over 16 years age group in this diagnostic assessment, plus two studies of an earlier version of one of the novel devices, included as supporting evidence. (NB. for simplicity, from here onwards we state the number of included studies in this review to be 14). All studies were published in peer-reviewed scientific journals, with the exception of:

- Alsaif et al (2023)[17] manuscript submitted for publication (available to the EAG academic in confidence by the study authors). The manuscript provides final results of the study, superseding interim results presented in a conference abstract. [18]
- Storey et al (2022)[25] conference poster presentation supplied by the company to the EAG (not confidential). No further reports or publications for this study are available currently.

In each study the novel device under evaluation was compared to one of the NICE guideline (202) recommended diagnostic testing approaches for OSA (e.g. RP or PSG). In all studies both the novel test and the comparator test were the performed with the same participants. No eligible studies were identified in which a novel device was compared against more than one other test, and there were no studies in which novel and comparator test were performed in separate groups of participants.

Table 2 Overview of included studies by novel device and comparator (people over 16 years)

Intervention novel device	Comparator/reference standard	
	Respiratory polygraphy (RP)	**Polysomnography[a] (PSG)**
AcuPebble SA100	Devani et al (2021)[19] [15]	Sanchez Gomez (2024)[20]
Brizzy	-	Martinot et al (2017)[21]
NightOwl	-	Massie et al (2018)[22] Massie et al (2022)[23] [b] Van Pee et al (2022)[24] [c] Lyne et al (2023) [25] [d]
Sunrise	Alsaif et al (2023)	Pepin et al (2020)[26] Kelly et al (2022)[27] [e]
WatchPAT ONE	Storey et al (2022)[28]	-
WatchPAT 300	Mueller et al (2022)[29] [f]	-
Supporting evidence		
WatchPAT 200-Unified[g]	-	Tauman et al 2020[30] Pillar et al 2020[31]

The '-' symbol means that no studies met the inclusion criteria

[a] Hospital sleep laboratory-based PSG in all studies except Kelly et al (2022) where PSG was home-based.

[b] Not stated whether the disposable or reusable version of NightOwl is used; we assume the reusable version.

[c] described as the "NightOwl reusable version"

[d] includes versions described as "disposable NightOwl Mini and the NightOwl reusable"

[e] home-based PSG is the comparator

[f] we note there is some uncertainty over which version of WatchPAT was used in this study. The publication describes the novel device as "WatchPAT 200/300" without further explanation on what this means. In the absence of a response to our request for clarification from the authors we made a pragmatic assumption that WatchPAT 300 was the novel device evaluated by the study. From this point onwards when referring to this study we state WatchPAT 300 rather than "WatchPAT 200/300".

[g] In accordance with the systematic review protocol, we permitted inclusion of studies evaluating an earlier version of a novel device if evidence for the current version of the device was unavailable. WatchPAT 200 Unified is the predecessor to WatchPAT 300 / ONE; two studies of WatchPAT 200 Unified were identified by our systematic searches.

As Table 2 shows, novel devices were compared to home RP in four studies, to hospital sleep-laboratory PSG in seven studies, and home-based PSG in one study. There were no eligible studies identified which compared novel devices to home-based pulse oximetry.

The EAG would like to highlight that use of the term comparator reflects the diagnostic tests that studies used to evaluate the novel device against, and that these do not necessarily correspond to the comparators as defined by the NICE scope. In many studies the comparator test was the reference standard for assessing diagnostic performance of the novel device. Some study publications explicitly refer to the comparator test as being a reference standard. The NICE scope uses the terms comparator and reference standard to refer to specific types of diagnostic testing. To illustrate this Table 3 shows the eligible comparators and reference standards as presented in the NICE scope and which of these feature in each included study. Table 3 is, therefore, an alternative presentation of the information in Table 2 from the perspective of the NICE scope.

Table 3 Overview of included studies by novel device and comparator as presented in the NICE scope (people over 16 years)

Study	Novel device	NICE scope comparator		NICE scope reference standard		
		Home RP	Home oximetry	In-hospital PSG	Other setting PSG	Healthcare setting RP
Novel devices compared to home RP						
Devani et al (2021)[10]	AcuPebble SA100	✓ [a]	X	X	X	X
Alsaif et al (2023)[17]	Sunrise	✓ [a]	X	X	X	X
Storey et al (2022)[28]	WatchPAT ONE	✓	X	X	X	X
Mueller et al (2022)[29]	WatchPAT 300	✓	X	N/A	N/A	N/A

Novel devices compared to PSG						
Sanchez Gomez (2024) [20]	AcuPebble SA100	X	X	✓	X	X
Martinot et al (2017)[11]	Brizzy	X	X	✓	X	X
Massie et al (2018)[22]	NightOwl	X	X	✓	X	X
Massie et al (2022)[12]	NightOwl	X	X	✓	X	X
Van Pee et al (2022)[13]	NightOwl	X	X	✓	X	X
Lyne et al (2023)[14]	NightOwl	X	X	✓	X	X
Pepin et al (2020) [26]	Sunrise	X	X	✓	X	X
Kelly et al (2022)[9]	Sunrise	X	X	X	✓	X
Supporting evidence						
Pillar et al (2020)[31]	WatchPAT 200-Unified	X	X	✓	X	X
Tauman et al (2020)[30]	WatchPAT 200-Unified	X	X	✓	X	X
[a] home RP is a reference standard for diagnostic accuracy measures in this study N/A Not Applicable (study does not measure diagnostic accuracy)						

In the novel device vs RP studies, both tests were done at home by the patient. In contrast, the novel device vs PSG studies were done in the hospital sleep laboratory setting, typically with simultaneous administration of novel device testing and PSG testing. (NB. Again, Kelly et al 2022 [27] is an exception as both novel device and PSG testing were done in the patient's home). As will be discussed later, home-based novel device testing studies are more relevant to the decision problem than studies in which the novel device is used in a hospital-sleep laboratory. Furthermore, the two settings are not necessarily comparable in terms of device efficacy. The fact that disproportionately fewer home-based studies are represented in this review is a limitation to bear in mind when interpreting the synthesis of clinical effectiveness.

Table 2 illustrates further disproportionality in the evidence for novel devices. The NightOwl device was evaluated in four studies, the Sunrise device in three studies, AcuPebble SA100 in two studies, and only a single study each was included for Brizzy, WatchPAT ONE and WatchPAT 300.

Predecessor versions of novel devices

Of the two WatchPAT studies included in the review,[28 29] neither reports diagnostic accuracy results (e.g. sensitivity and specificity). Without these data it would not be possible to assess the cost effectiveness of WatchPAT 300/ONE using our economic model (see section 5). The protocol for this review states that evidence for earlier, comparable versions of the devices would be considered if necessary. We therefore considered the feasibility of using evidence from earlier versions of WatchPAT in lieu of the current versions. The NICE scope notes that software or algorithms used by the devices may have been periodically updated,

which may impact performance. Therefore, evidence based on earlier versions of the software (e.g. WatchPAT 100, 200 and 200 Unified) may not accurately reflect the effectiveness of the current versions.

The manufacturer confirmed with NICE that both WatchPAT 300 and WatchPAT ONE use an identical algorithm to WatchPAT 200U. In addition, both devices produce identical signals to that of WatchPAT 200U. This similarity enabled the company to obtain FDA and CE approval for the current devices based on technological continuity. We therefore took a pragmatic decision to include any eligible studies of WatchPAT 200 Unified to inform economic modelling. Our systematic literature searches were designed to identify all versions of WatchPAT and we were able to find two relevant WatchPAT 200U studies; these are included as supporting evidence in the systematic review. [30 31]

4.1.2 Studies of children and young people aged 2 to 16 years

Table 4 lists the three included studies relevant to the 2 to 16 years age group in this diagnostic assessment. In two studies the novel devices studied (Brizzy and Sunrise) were performed overnight in hospital sleep laboratory simultaneously with PSG, the comparator/reference standard.

[redacted]

Unlike in the over 16 years group, we did not identify any relevant studies of predecessor versions of WatchPAT (e.g. WatchPAT 200U) which could have been included as supportive evidence.

Table 4 Overview of included studies by novel device evaluated (children and young people aged 2 to 16 years)

Novel device	Comparator	
	Respiratory polygraphy (RP)	Polysomnography[a] (PSG)
AcuPebble SA100	[redacted]	[redacted]
Brizzy	-	Martinot et al (2015)[33]
NightOwl	-	-
Sunrise	-	Martinot et al (2022)[34]
WatchPAT 300	-	-
WatchPAT ONE	-	-
The '-' symbol means that no studies were included. [redacted]		

Table 5 shows the eligible comparators and reference standards as presented in the NICE scope and which of these features in each included study in the 2-16 years age group. This is an alternative presentation of the information in Table 4, from the perspective of the NICE scope.

Table 5 Overview of included studies by novel device evaluated as presented in the NICE scope (children and young people aged 2 to 16 years)

Study	Novel device	NICE scope comparator			NICE scope reference standard	
		Home RP	Home oximetry	In-hospital PSG	Other setting PSG	Healthcare setting RP
Novel devices compared to PSG						
NCT04031950 (2019)[32] [a]	AcuPebble SA100	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Martinot et al (2015)[33]	Brizzy	X	X	✓	X	X
Martinot et al (2022)[34]	Sunrise	X	X	✓	X	X
[a] [redacted]						

4.2 Characteristics of included studies (people over 16 years)

4.2.1 Overview of general study characteristics (people over 16 years)

Table 6 gives an overview of the general characteristics of studies included in the review for people over 16 years. (NB. Participant characteristics are reported in the next sub-section, 4.2.2). In terms of geographical location, most studies were conducted in northern Europe, with a minority from further afield, including the USA and Australia. Within Europe a cluster of studies originate from Belgium, reflecting the concentration of specialist scientific and clinical expertise in that area. Three studies were UK-based, two from specialist sleep centres in London (Alsaif et al., 2023; Devani et al., 2021), additionally in the Scottish Highlands (Alsaif et al., 2023), and the third study [redacted] (Storey et al., 2022). In a fourth study the main study centre was in France with a reference centre for expert PSG scoring located in London (Kelly et al., 2022).

Study sample sizes

The number of participants enrolled in the studies varied from 40 (Kelly et al., 2022 and Alsaif et al.,2023) to 600 (Storey et al., 2022). There was variability in how study sample sizes were decided:

- Seven studies did not report statistical power calculations for the number of participants necessary to recruit for hypothesis testing (Mueller et al., 2022; Storey et

al., 2022, Massie et al., 2018; Massie et al., 2022; Kelly et al., 2022; Pillar et al., 2020 and Tauman et al., 2020).

- Six studies (Devani et al., 2021; Martinot et al., 2017; Van Pee et al., 2022; Lyne et 23; Sanchez Gomez et al., 2024 and Pepin et al., 2020) reported sample size calculations and all were subsequently adequately powered, except for Pepin et al., (2020) which recruited just under the minimum target number of patients in one of the study groups (46 patients instead of 50 patients for the non-OSA group).
- One study (Alsaif et al., 2023) stated that [redacted]

In most studies the number of participants analysed was lower than the number enrolled due to patient exclusions. Various reasons for exclusion were reported including patient withdrawal from the study and study administrative errors. Participants were also excluded in cases of test failure, where it was deemed that the novel home test and/or the comparator test wasn't performed according to standard protocol or the recorded data did not meet criteria for validity (see section 4.5.6 for failure rates).

Comparator / reference standard

The comparator used in most studies (n=9) was sleep laboratory-based polysomnography (PSG). This was the standard of care in many study centres and widely regarded by investigators as the "gold standard" testing approach for diagnosing and assessing the severity of OSA. In studies assessing both diagnostic accuracy and other types of outcomes (e.g. clinical measures, patient reported outcomes) PSG can be considered a reference standard for the former and comparator test for the latter. The novel device was used concomitantly with laboratory-based PSG in all 9 studies rather than in the home setting, thus limiting their applicability to the decision problem. Sleep study setting is considered to influence the diagnostic test performance and estimates from laboratory-based studies are not necessarily representative of home-based studies where, for example, the patient is responsible for correct administration of their own tests.[14] In an additional study the PSG was installed in the patient's home and used concomitantly with the novel device, Sunrise (Kelly et al., 2022).

There were just four studies which compared novel devices against respiratory polygraphy in the home setting. In Devani et al (2021), the study was designed to represent the conditions in which AcuPebble is typically used in practice, i.e. the home environment. The company stated that this was a requirement necessary to obtain regulatory approval. Respiratory

polygraphy, which is a commonly used home test, was therefore an appropriate comparator. In Mueller et al (2022), the choice of level 3 cardiorespiratory polygraphy as a comparator to the WatchPAT device was because polygraphy is well validated and commonly used in practice. The choice of home polygraphy as a comparator in the third study, the Sunrise OSA Trial (SOSAT) (Alsaif et al 2023), was prompted by the results of previous studies demonstrating comparability in performance between the novel device (Sunrise) and in-lab / home-based PSG. In the fourth study (Storey et al.,2022), the need to manage infection control and waiting lists for home sleep study tests due to the COVID-19 pandemic was the impetus behind the authors comparing WatchPAT ONE to their hospital's standard home sleep study test (NOX T3).

Study designs

There was general uniformity in study designs used. Most were prospective cross-sectional evaluations of patients referred to specialist sleep services with suspected OSA. Patients received overnight standard of care testing with concomitant administration of novel device testing. Comparisons between tests are therefore performed within a single cohort of patients in each study. A limitation of many of the studies with this design is that the novel device and comparator / reference standard test are evaluated in the sleep laboratory rather than in its intended setting (i.e. the patient's home). Sleep study setting is known to influence the diagnostic performance of devices and estimates from laboratory-based studies are not representative of home-based studies.[14]

Two studies introduced minor variations to the single cohort design. In the Sunrise OSA Trial (SOSAT) (Alsaif et al 2023) patients received novel and standard testing simultaneously during a single night, and additionally they were randomised to receive their treatment decision based on either the novel device (Sunrise) or the comparator (home respiratory polygraphy). Patients in the study by Mueller et al (2022) underwent testing with the novel device (WatchPAT 300) and the comparator, respiratory polygraphy, in a randomized order over two consecutive nights (RP then PAT or PAT then RP).

In all studies the novel device test and comparator test were done on a single night. In practice however, some testing protocols permit multi-night testing if required to obtain a successful sleep test. As sleep patterns often vary from one night to another this is a notable limitation to be considered when interpreting the results of the clinical effectiveness systematic review and cost effectiveness analysis.

Whilst all studies evaluated the efficacy of novel devices as a diagnostic approach, the precise focus of inquiry varied. We observed a number of themes of inquiry, and present some examples for illustration:

- **Validation studies**. Some studies were done to validate a novel diagnostic technology, or certain novel features of a technology. For instance, Martinot et al (2017) hypothesised that analysis of mandibular movement (MM) during sleep (in this study recorded by the Brizzy magnetic sensor) would compare favourably to PSG on measures of test agreement and diagnostic accuracy. The authors make a scientific case showing how analysis of MM during sleep can accurately identify the occurrence of cortical arousal, increased respiratory effort (RE) and RE-related arousal (RERA), all of which are associated with sleep apnoea events.
- **Real world studies.** Some studies were designed to assess the efficacy of a novel device in settings typical of those in which it is intended to be used (Devani et al 2021; Alsaif et al, 2023, Mueller et al., 2022). These studies assessed patient usability of the test, comfort levels during testing, and overall acceptability to the patient, amongst other outcomes.
- **Specialised investigations.** Some studies investigated particular aspects of sleep testing to advance knowledge in poorly understood areas. For example, Massie et al (2022) aimed to evaluate the rapid eye movement (REM) phenotyping performance of the PAT signal from the NightOwl device.

Table 6 Overview of included studies (people over 16 years)

Study ID	Country. No. centres	Study design	Intervention. Setting	Comparator/reference standard. Setting	Study population. No. patients	Outcome measures
Novel devices compared to home RP						
Devani et al (2021)[19]	UK (London). Single centre	Prospective, single cohort	AcuPebble SA100 automated diagnosis. Home-based, (Unattended) (Patients were not trained on the use of the device under evaluation)	Cardiovascular respiratory polygraphy (CR-PG) (Embletta MPR Sleep System). with manual scoring Home-based (Unattended)	People with suspected OSA referred for examination. N=182 enrolled N=150 analysed	Diagnostic test accuracy; Accuracy in event classification, including central versus obstructive apnoeas Diagnostic test agreement; Diagnostic test failure rates; Patient acceptability and usability; Healthcare resources used and costs.
Alsaif et al (2023)[17] SOSAT trial	UK (Scottish highlands and inner-London). Two centres	Prospective, randomised, single cohort blinded pilot study.	Sunrise (MMs) with autoscoring). Home-based. (Unattended)	RP (ApneaLink Air device) with manual scoring. Home-based. (Unattended)	People with suspected OSA undergoing investigation. [redacted]	Time to treatment decision (days); Diagnostic test accuracy; Diagnostic test agreement (AHI; treatment decisions); Diagnostic test failure rates. [redacted]
Storey et al (2022)[28]	UK [redacted] Single Centre	Prospective randomised study	WatchPAT ONE Home-based (Unattended)	RP (NOX T3) Home-based (Unattended)	Patients referred by Sleep, ENT, Insomnia, Dental or Respiratory consultants N=600 enrolled (300 randomised to WatchPAT ONE and 300 to NOX T3	Mean patient time (including travel time) to receive and return equipment; Number of appointments not attended by patients for intervention versus comparator; cost per appointment (equipment, room staff, postage); mean staff time taken per appointment (excluding analysis)
Mueller et al (2022)[29]	Germany. Single centre	Prospective randomised study	WatchPAT 300 with manual scoring (based on manual editing with software) Home-based. (Unattended)	RP (Miniscreen plus device) with manual scoring. Home-based. (Unattended)	People with suspected OSA needing home sleep testing. N=61 enrolled N=56 analysed	OSA diagnosis rates; OSA severity classification; Diagnostic test failure rates; Time spent in supine sleep position; Number of repeat sleep studies; Perceived quality of sleep and test related discomfort.

Study ID	Country. No. centres	Study design	Intervention. Setting	Comparator/reference standard. Setting	Study population. No. patients	Outcome measures
Novel devices compared to PSG						
Sanchez Gomez et al (2024) [20]	Spain. Single centre	Prospective, single cohort	AcuPebble SA100 automated diagnosis. Sleep laboratory-based.	PSG (Philips Sleepware G3 version 2.8.78) with manual scoring. Sleep laboratory-based (Attended)	Patients referred for assessment of potential OSA N=80 enrolled N=63 analysed	Diagnostic test accuracy; Diagnostic test failure rates; OSA severity classification.
Martinot et al (2017)[21]	Belgium. Single centre	Prospective, single cohort	Brizzy (MMs) with manual scoring. Sleep laboratory-based.	Routine PSG (SomnoscreenPlus) with manual scoring. Sleep laboratory-based. (Unattended)	People with suspected OSA referred for laboratory sleep test (with moderate to high pre-test probability) N=100 enrolled N=92 analysed (inc. 13 healthy volunteers)	Diagnostic test accuracy; Diagnostic test agreement; Diagnostic test failure rates OSA severity classification.
Massie et al (2018)[22]	Belgium. Single centre	Prospective, single cohort	NightOwl (reusable version) with autoscoring Sleep laboratory-based.	PSG (device not stated) with a combination of manual and automated scoring Sleep laboratory-based.	Patients who underwent a diagnostic in-hospital PSG in the sleep laboratory N=101 enrolled N= 101 analysed	Diagnostic test accuracy; Diagnostic test agreement; OSA severity classification.
Massie et al (2022)[23]	USA and Belgium. Four centres (3 in USA, 1 in Belgium)	Prospective, single cohort.	NightOwl with autoscoring. Sleep laboratory based. (Attended)	Routine PSG (Alice 6 PSG (European centres) or Cadwell Easy PSG (USA centres)). Sleep-laboratory based (Attended) Each PSG manually scored independently by local centre & by a	People with suspected OSA scheduled for in-lab PSG. N=261 enrolled N=261 analysed	Diagnostic test accuracy;[a] Diagnostic test agreement;[a] Minimum required REM sleep time; OSA diagnosis rates.

Study ID	Country. No. centres	Study design	Intervention. Setting	Comparator/reference standard. Setting	Study population. No. patients	Outcome measures
				separate expert centre (reference standard)		
Van Pee et al (2022)[24]	USA and Belgium. Four centres (3 in USA, 1 in Belgium).	Prospective, single cohort.	NightOwl (reusable version) with autoscoring Sleep laboratory-based. (Attended)	As per Massie et al (2022) above	People with suspected OSA scheduled for in-lab PSG. N=228 enrolled N=167 analysed	Diagnostic test accuracy; OSA severity classification; Diagnostic test agreement. Diagnostic test failure rates
Lyne et al (2023)[25]	Australia (Melbourne). Single centre	Prospective, single cohort	NightOwl Mini disposable (NOM) with autoscoring NightOwl reusable (NOR) with autoscoring. Sleep laboratory-based	Reference standard: PSG (Compumedics Grael Profusion PSG system) with manual scoring. Sleep laboratory-based	People with suspected OSA scheduled for in-lab PSG. N= 115 enrolled, N= 100 analysed	Diagnostic test agreement; Diagnostic test accuracy; OSA severity classification; Diagnostic test failure rates
Pepin et al (2020) [26]	Belgium. Single centre	Prospective, single cohort	Sunrise system (MMs) with autoscoring. Sleep laboratory-based.	PSG (Somnoscreen Plus system) with manual scoring. Sleep laboratory-based	People with suspected OSA scheduled for in-lab PSG. N=376 enrolled N=376 analysed	Diagnostic test accuracy [b] Diagnostic test agreement; OSA severity classification; Diagnostic test failure rates
Kelly et al (2022)[27]	France and UK (London)	Prospective, single cohort	Sunrise system (MMs) with autoscoring. Home-based. (Unattended)	PSG (Nox A1 portable acquisition system) with manual scoring. Home-based. (Unattended)	People with suspected OSA undertaking a diagnostic home sleep study. N=40 enrolled N=31 analysed	Diagnostic test agreement; Diagnostic test accuracy [b;] OSA severity classification; Diagnostic test failure rates
Supporting evidence[c]						
Pillar et al (2020)[31]	USA, Israel, Germany and Canada. 11 centres (5 in USA, 4 in Israel, 1 in	Prospective, single cohort	WatchPAT 200-Unified with autoscoring Sleep laboratory based	PSG ("full in-lab PSG", "FDA approved in-lab PSG from multiple manufacturers" used in the 11 centres) with manual scoring.	Selective recruitment of heart failure patients in order to have a substantial representation of	Diagnostic test agreement; Diagnostic test accuracy; OSA severity classification; Measures of concordance or agreement

Study ID	Country. No. centres	Study design	Intervention. Setting	Comparator/reference standard. Setting	Study population. No. patients	Outcome measures
	Germany and 1 in Canada			Sleep laboratory based	patients with central sleep apnoea. N=84	
Tauman et al (2020)[30]	USA, Israel, Germany and Canada. 11 centres	Prospective, single cohort	WatchPAT 200-Unified with automated scoring. Sleep-laboratory based.	PSG ("FDA approved in-lab PSG from multiple manufacturers" used in the 11 centres) with manual scoring. Sleep-laboratory based	Patients previously diagnosed with atrial fibrillation (permanent, persistent or paroxysmal) and suspected to have sleep apnoea, N=101	Diagnostic test agreement; Diagnostic test accuracy; OSA severity classification; Test failure rate: Measures of concordance or agreement

[a] test accuracy and agreement in detecting Rapid Eye Movement (REM)-related OSA (REM OSA a distinct OSA phenotype).
[b] Post-hoc analysis was performed to optimize the diagnostic cut-offs
[c] see section 4.1 for an explanation of supporting evidence.
MM mandibular movements; PAT peripheral arterial tonometry; PSG polysomnography; RP respiratory polygraphy.

4.2.2 Participant characteristics (people over 16 years)

Table 60 in Appendix 4 gives the demographic and general health profile of participants across the studies. A limited range of characteristics were reported by the studies, with a focus on age, sex, and weight. Other factors, such as race and ethnicity were rarely mentioned, likewise socio-economic status and health-related lifestyle.

Mean age across the studies varied from 41 to 56 years, most commonly around 48 years. Likewise, mean age ranges varied, with the youngest age 18 and the highest 84 years. There were disproportionately more males than females across studies, the biggest differential being around 70%/30% male/female (Alsaif et al., 2023; Devani et al., 2021; Mueller et al., 2022). Mean study BMI ranged from 28 to around 37 kg/m^2 with many studies reporting a mean of around 30 kg/m^2. This indicates a generally overweight/obese patient population and is to be expected given that excess weight is a key risk factor for OSA.

Details of study participant ethnicity and race were only reported by two studies: Devani et al. (AcuPebble) and Van Pee et al. 2022 (NightOwl). The ethnicity profile in the Devani et al study, set in London, was mixed, comprising White British/Other, Asian/Asian British, and Black/Black British participants (see Table 60 Characteristics of study participants in the systematic review of clinical effectiveness (people over 16 years). In the Van Pee et al study, co-located in Belgium and the US, the study sample featured participants defining themselves as: White; Black; or Hispanic, Latino, or Spanish. Of these groups the largest representation in the study was White. A third study, Massie et al. 2022 (NightOwl) stated that “Persons of diverse racial and ethnic backgrounds were included” but did not report the which ethnic groups were represented and in what proportions.

Four studies reported the presence of comorbidities. In the study by Devani et al. (2022), the most reported comorbidities were high blood pressure and diabetes, affecting approximately 25% and 11% of patients respectively. The study by Sanchez Gomez (2024) reported a wide range of comorbidities, the most common of which were hypertension (27%), cardiac diseases (11%) and neurological diseases (14%).In the two supporting evidence studies by Pillar et al., 2020 and Tauman et al., 2020, where patients with cardiac disorders were specifically recruited for study inclusion, 23% and 42% of patients respectively had diabetes.

Some studies excluded people with major health conditions (e.g. musculoskeletal diseases) from taking part, whilst other studies were less restrictive. Given that patients were recruited consecutively into studies at the time of their referral they can, to some extent, be regarded

typical of patients seen in clinical practice. It would be reasonable to assume that they would therefore have a typical comorbidity profile even though these are not described by the studies.

Baseline Epworth Sleepiness Scale (ESS) scores were reported in approximately a quarter of studies. The ESS questionnaire assesses the likelihood of falling asleep while doing different daily activities and is scored from 0 to 24 (higher scores indicating increased severity of symptoms). The range in mean ESS scores was 7.7 (normal daytime sleepiness) to 15.5 (moderate excessive daytime symptoms).

The limited information given on patient characteristics presents a significant challenge for an assessment of the clinical effectiveness and cost effectiveness of novel devices in the subgroups included in the NICE scope (e.g. People with COPD / neuromuscular disorders / People from black, Asian and minority ethnic backgrounds and pregnant women and pregnant people).

4.3 Characteristics of included studies (children and young people aged 2 to 16 years)

4.3.1 Overview of general study characteristics (children and young people aged 2 to 16 years)

Details of the three studies which met the inclusion criteria in the 2 to 16 years age group are presented in Table 7. Two studies were conducted by the same team of investigators, based in Belgium and the USA, and both evaluated the monitoring of mandibular jaw movements during sleep to diagnose OSAHS. Two novel devices for the monitoring of mandibular movements during sleep were evaluated: Brizzy with manual scoring (Martinot et al., 2015)[33] and Sunrise with automated scoring (Martinot et al., 2022)[34] . None of the three studies reported a sample size calculation.

Martinot et al (2022)[34] is the larger of the two studies, comprising 140 children aged 3-17 years referred to a hospital sleep laboratory in Belgium with clinically suspected OSA. The children underwent overnight laboratory-based PSG testing concurrently with analysis of mandibular movements by the Sunrise device. Outcome measures include agreement between the two testing approaches, and diagnostic accuracy of Sunrise based on ROC curve analysis using the Respiratory Disturbance Index (RDI).

Martinot et al (2015)[33] is an older investigation with a much smaller sample size of 33 children with suspected OSA scheduled to undergo adenotonsillectomy. The children received overnight laboratory-based PSG testing concurrently with monitoring of mandibular movements by the Brizzy device. Outcome measures include median obstructive sleep apnoea-hypopnoea index (OAHI) values; correlations between mandibular movement rates and respiratory effort associated with OAHI; total sleep time, number of obstructive and central apnoea events and other measures relating to mandibular movements and indices of sleep disordered breathing. The study does not, however, report the diagnostic accuracy of the novel device, and in general the outcome measures are of limited relevance to the decision problem in this assessment.

With regard to the decision problem for this diagnostic assessment a key limitation of the two Martinot et al studies is the novel devices were tested in the sleep laboratory rather than in the home setting.

The third study, NCT04031950 2019, is an ongoing study of the AcuPebble SA100 device at two centres in the UK.

[redacted] The company confirmed to NICE in November 2023 that this is an in-hospital based study and that the reference standard is [redacted]

[redacted] Outcome measures include diagnostic accuracy and ease of use and acceptability for parents/carers. It should be noted that AcuPebble SA100 is currently only indicated for use in adult patients. On completion of the study the company intend to extend the indication to include the paediatric population (i.e. under 18 years) with the regulation application expected to complete in 2024.

4.3.2 Participant characteristics (children and young people aged 2 to 16 years

Table 61 in Appendix 4 gives the demographic and general health profile of participants across the three studies.

[redacted] Demographic data were reported for the two studies by Martinot (2015 and 2022). As is the case with the over 16 years population (discussed above), only a limited range of characteristics were reported by these two studies. Median age ranged from 5 to 6.9 years and median BMI was 15.8 to 16.6 kg/m^2. The presence of any comorbidities was not stated in two of the studies, but specific comorbidities were ruled out by study exclusion criteria. For example, Martinot et al (2022) excluded children with significant, chronic medical conditions, such as genetic syndromes, diabetes mellitus, craniofacial anomalies, or neurologic disease. Furthermore, Martinot et al (2015) appear to have excluded children with craniofacial or neuromuscular disorders. In study NCT04031950 [redacted]

Table 7 Overview of included studies (children and young people aged 2 to 16 years)

Study ID	Country. No. centres	Study design	Intervention. Setting	Comparator. Setting	Study population. No. patients	Outcome measures
Novel devices compared to PSG						
NCT04031950 (2019)[32]	UK Two centres (only one has results currently)	Prospective cohort	AcuPebble, automated scoring. Sleep laboratory-based.	[redacted]	Inclusion criteria 1-18 yrs referred to sleep clinic with suspected OSA. [redacted]	Diagnostic test accuracy; Ease of use and acceptability for patients and parents/carers
Martinot et al (2015)[33]	Belgium. Single centre	Prospective single cohort	Brizzy (MMs), manual scoring. Sleep laboratory-based.	Routine PSG with Dream Medatec device with manual scoring. Sleep laboratory-based.	Aged 2-16 yrs with adenotonsillar hypertrophy and suspected OSA N=33 enrolled N=33 analysed	Correlations between MM rates and respiratory effort associated with OAHI
Martinot et al (2022)[34]	Belgium. Single centre	Prospective single cohort	Sunrise system with MMs and automated scoring. Sleep laboratory-based.	Routine PSG using XDream Medatec device with manual scoring. Sleep laboratory-based.	Aged 3-17 yrs referred to sleep laboratory with suspected OSA N=155 enrolled N=140 analysed	Diagnostic test agreement;[a]

[a] Post-hoc analysis was performed to optimize the cut-offs
OAH obstructive sleep apnoea-hypopnoea; OAHI obstructive sleep apnoea-hypopnoea index, MM mandibular movements; PSG polysomnography

4.4 Results of critical appraisal of study methodology

In this section we summarise the results of our critical appraisal of all the studies included in this systematic review (i.e. for the 2-16 years age group and for the 16 years and older group). Further detail on our critical appraisal judgements are presented in Appendix 5.

We applied the QUADAS-2 tool[13] to each of the included studies to assess the risk of bias and the applicability of the study to the decision problem. The QUADAS-2 tool appraises the likelihood of bias arising from: the selection of participants; the conduct and interpretation of the index test and the reference standard; the flow of participants through a study and the timing of the index test and reference standard. It also assesses the applicability of the participants selected and the index test and reference standard to the review's research question. Table 8 shows the results of our critical appraisal and Figure 2 presents the results graphically, for all studies included in this systematic review (i.e. both the 'people over 16 years' and children and young people aged 2 to 16 years' sub-groups).

The majority of studies were judged to be at low risk of bias overall, but in four studies a high risk of bias judgement was made in one of the four bias domains. Patient selection was judged to be at high risk of bias in the study by Pillar et al., 2020. The study selectively recruited heart-failure patients but it is not clear if this resulted in inappropriate exclusions. The intentional bias towards selecting patients with congestive heart failure may, therefore, have introduced other unintentional biases.[31] A high risk of bias was judged in the conduct or interpretation of the index test in three studies(Kelly et al., 2022, Martinot et al., 2022 and Pepin et al., 2020)[26 27 35] all of which used post-hoc analyses to optimise diagnostic cut-off points, potentially over-estimating novel device diagnostic accuracy.

Regarding applicability to the decision problem, most studies were judged as low concern for the patient selection and the reference standard domains. However in many studies it was unclear whether the conduct, or interpretation of the index test was relevant to the decision problem. This judgement was made for all studies where the novel testing device was used in a sleep laboratory (concomitant to PSG testing), rather than its intended setting (i.e. the patient's home). Two studies were also rated unclear for this domain although they were conducted in a home setting. Alsaif et al (2023) did not report on the thresholds used in their study and Storey et al., 2022 did not report details of the conduct and interpretation of the index test. For four studies, the judgements were of high concern – in Kelly et al., 2022, Martinot et al., 2022 and Pepin et al., 2020 all used post-hoc analyses to optimise diagnostic cut-off points, while in Martinot et al., 2015 diagnostic accuracy results for against PSG or

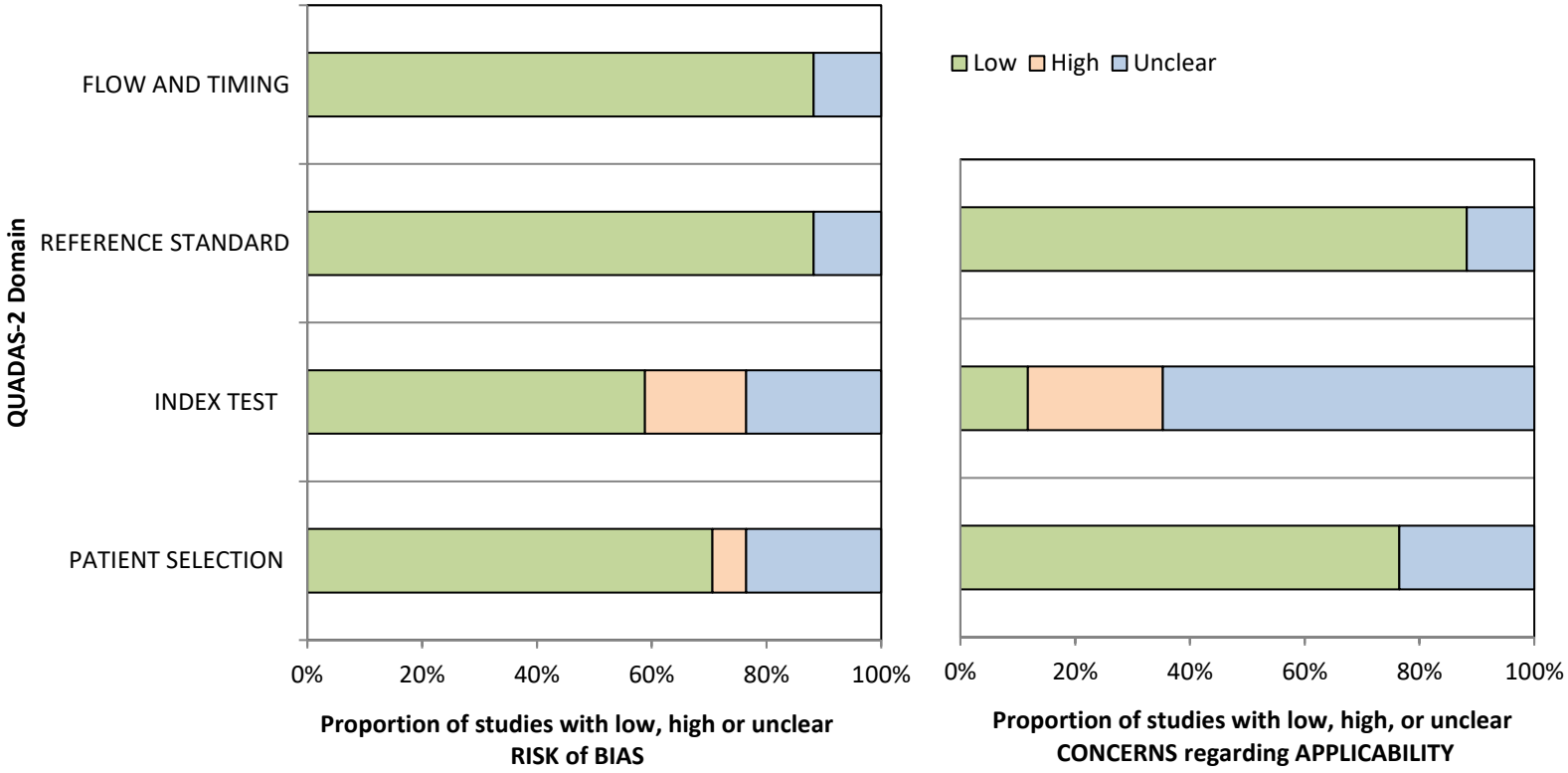
any other reference standard were not reported and the study was conducted in a sleep laboratory rather than the home setting.

Table 8 Overview of QUADAS-2 assessments for all studies

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
AcuPebble SA100							
Devani 2021	☺	☺	☺	☺	☺	☺	☺
Sanchez Gomez 2024	☺	☺	☺	☺	☺	?	☺
NCT04031950 (child)	☺	☺	☺	☺	☺	?	☺
Brizzy							
Martinot 2017	☺	☺	☺	☺	☺	?	☺
Martinot 2015 (child)	☺	?	☺	☺	☺	☹	?
NightOwl							
Massie 2018	?	☺	☺	☺	?	?	☺
Massie 2022	☺	☺	☺	☺	☺	?	☺
Van Pee 2022	☺	☺	☺	☺	☺	?	☺
Lyne 2023	?	☺	☺	☺	☺	?	☺
Sunrise							
Pepin 2020	☺	☹	☺	☺	☺	☹	☺
Kelly 2022	☺	☹	☺	☺	☺	☹	☺
Alsaif 2023	?	?	☺	?	?	?	☺
Martinot 2022 (child)	☺	☹	☺	☺	☺	☹	☺
WatchPAT 300/ONE							
Mueller 2022	☺	?	?	☺	☺	☺	☺
Storey 2022	☺	?	?	?	?	?	?
Supporting evidence (WatchPAT 200U)							
Pillar 2020	☹	☺	☺	☺	?	?	☺
Tauman 2020	?	☺	☺	☺	☺	?	☺

☺Low Risk ☹High Risk ? Unclear Risk

Figure 2 Proportion of studies with low, unclear or high risk of bias and proportion of studies with low, unclear or high concerns regarding applicability (all studies)



The following sections, 4.5, 4.6, and 4.7, present a synthesis of study intermediate outcomes, clinical outcomes, and patient-reported outcomes, respectively, for the over 16 years population. Subsequent sections, 4.8, 4.9 and 4.10 present, respectively, a synthesis of the outcomes for the children and young people aged 2 to 16 years population.

4.5 Intermediate outcomes (people over 16 years of age)

4.5.1 Diagnostic accuracy (people over 16 years of age)

Table 9 reports diagnostic accuracy estimates for the novel testing devices, as reported in 10 studies. For each study we present summary estimates of sensitivity, specificity, positive and negative predictive values and overall accuracy. Estimates are given for respective cut-off values of the diagnostic indices as used by each study (e.g. the AHI). The cut-offs are expressed as the average number of events occurring per hour of total sleep time. Hence, an AHI score of 15 means an average of 15 apnoeas or hypopneas recorded per hour. The PSG cut-offs of 5, 15 and 30 correspond to established severity categories for mild, moderate and severe OSA, respectively. Some studies used alternative diagnostic indices such as the ODI and RDI.

Not all studies reported the full set of estimates we required hence, where possible, we calculated missing estimates constructing diagnostic contingency tables to record the number of true/false positive/negative tests. The diagnostic metrics were estimated using a online diagnostic test evaluation calculator (MedCalc®). Where possible we used digitisation software (Engauge Digitizer) to extract the data-points from scatter plots in study publications, providing us with the diagnostic index value (e.g. AHI) for each data-point (participant) estimated by the novel device and the reference standard. The index value for each participant was then counted against the relevant disease severity category (mild/moderate/severe), using a 4x4 diagnostic contingency table.

Novel devices referenced to RP

Of the four studies comparing novel devices to RP (Alsaif et al 2023; Devani et al 2021; Mueller et al, 2022; Storey et al (2022)) only Devani et al reported diagnostic test accuracy (AcuPebble SA100). They present accuracy values according to the current recommended American Academy of Sleep Medicine (AASM) AHI-based diagnostic criteria (using a ≥3% threshold for oxygen desaturation). A second set of accuracy values are reported using the same AHI criteria but the threshold for oxygen desaturation increased to ≥4%. A further two sets of estimates are provided using the ODI-based criteria, again varying oxygen desaturation between ≥3% and ≥4 respectively. Devani et al's selection of diagnostic criteria

was intended to represent the criteria used in clinical practice. As Table 9 shows, the sensitivity and specificity estimates for the four sets of criteria are all in the 90-100% range, indicating good diagnostic performance for AcuPebble referenced to home RP.

Also of note, the objective of the Devani et al study was to assess the efficacy of AcuPebble SA100 for automated diagnosis of *moderate-severe* OSA, as opposed to *mild, moderate, and severe* OSA. Consequently, their accuracy calculations combine AHI cut offs for 'no OSA' and 'mild OSA' into a single cut-off representing test negativity. For comparability to other studies we recalculated Devani's accuracy estimates using the conventional AHI cut off for test positivity (≥5 AHI) and oxygen desaturation ≥3%. Our estimates are similar to Devani's with the exception that the NPV decreases into the 80-90% range (corresponding to the estimate 15–30 AHI or >30 AHI (desat ≥3%) in Table 9 for Devani et al).

Table 9 Accuracy of novel devices in diagnosing OSAHS (people over 16 years of age)

Author, Novel device	**No. pts**	**Cut-offs** Novel device, Reference standard	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**PPV % (95% CI)**	**NPV % (95% CI)**	**Accuracy % (95% CI)**
Novel devices compared to home RP							
Devani et al (2021), AcuPebble SA100	150	15–30 AHI or >30 AHI (desat ≥3%)	93 (82 to 98)	97 (91 to 99)	94 (85 to 98)	96 (90 to 98)	95 (91 to 98)
		15–30 AHI or >30 AHI (desat ≥4%)	96 (86 to 100)	97 (92 to 99)	94 (84 to 98)	98 (93 to 99)	97 (92 to 99)
		15–30 ODI or >30 ODI (desat ≥3%)	91 (82 to 96)	93 (85 to 98)	93 (85 to 97)	91 (82 to 95)	92 (86 to 96)
		15–30 ODI or >30 ODI (desat ≥ 4%)	98 (89 to 100)	92 (85 to 97)	86 (76 to 92)	99 (93 to 100)	94 (89 to 97)
		≥5 AHI (desat ≥3%)[a]	92 (84 to 96)	96 (87 to 100)	98 (92 to 100)	87 (77 to 93)	93 (88 to 97)
Novel devices compared to PSG							
Sanchez Gomez (2024) [20]AcuPebble SA100	63	15-30 AHI or >30 (desat ≥ 3%)	92.86 (76.50 to 99.12)	97.14 (85.08 to 99.93)	96.30 (78.98 to 99.45)	94.44 (81.71 to 98.48)	95.24 (86.71 to 99.01)
		15-30 ODI or > 30 (desat ≥ 3%)	92.00 (73.97 to 99.02)	92.11 (78.62 to 98.34)	88.46 (72.01 to 95.81)	94.59 (82.19 to 98.51)	92.06 (82.44 to 97.37)
		≥5 AHI (desat ≥ 3%)[a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Martinot et al (2017), [b] Brizzy	92	MM-RDI > 5.9 PSG RDI ≥ 5	93 (86 to 97)	100 (51 to 100)	NR	NR	93 (86 to 97)
		MM-RDI > 13.5 PSG-RDI ≥ 15	89 (80 to 94)	100 (83 to 100.0)	NR	NR	91 (84 to 95)
		MM-RDI > 32.5 PSG-RDI ≥ 30	74 (58 to 86)	NR	NR	NR	NR
Massie et al (2018), NightOwl	101	NightOwl REI >5 MSSS PSG-AHI >5	98 (92 to 99)	80 (44 to 97)	98 (93 to 99)	80 (50 to 94)	96 (90 to 99)
		NightOwl REI >15 MSSS PSG-AHI >15	97 (88 to 100)	83 (68 to 93)	89 (81 to 94)	94 (81 to 99)	91 (84 to 96)
		NightOwl REI >30 MSSS PSG-AHI >30	90 (76 to 97)	97 (89 to 100)	95 (82 to 99)	94 (85 to 97)	94 (88 to 98)
Van Pee et al (2022), NightOwl	228	PAT AHI ≥5 (desat ≥3%) PSG AHI ≥5 (desat ≥3%)	93 (89 to 97)	72 (54 to 91)	96 (92 to 99)	62 (42 to 80)	90 (86 to 95)
		PAT AHI ≥15 (desat ≥3%)	91 (85 to 96)	76 (65 to 87)	82 (72 to 92)	88 (81 to 94)	86 (80 to 91)

Author, Novel device	**No. pts**	**Cut-offs** Novel device, Reference standard	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**PPV % (95% CI)**	**NPV % (95% CI)**	**Accuracy % (95% CI)**
		PSG AHI ≥15 (desat ≥3%)					
Lyne et al (2023), NightOwl (NOM)	94	NOM AHI < 5 PSG AHI < 5	93 (84 to 98)	77 (55 to 92)	93 (86 to 97)	77 (55 to 92)	89 (81 to 95)
NightOwl (NOR)	96	NOR AHI < 5 PSG AHI < 5	89 (80 to 95)	91 (71 to 99)	97 (90 to 99)	71 (56 to 83)	90 (82 to 96)
Pepin et al (2020), [b] Sunrise	376	Sunrise-RDI 7.63 PSG-RDI ≥5	91 (89 to 92)	94 (91 to 97)	99 (99 to 99)	59 (55 to 63)	92 (90 to 94)[c]
		Sunrise-RDI 12.65 PSG-RDI ≥15	92 (90 to 94)	84 (81 to 87)	89 (88 to 91)	88 (85 to 91)	88 (86 to 90)[c]
Kelly et al (2022),[b] Sunrise	31	MM-ORDI 9.53 PSG-ORDI >5	88 (69 to 97)	100 (54 to 100)	100 (85 to 100)	89 (NR)	94 (NR)
		MM-ORDI 12.65 PSG-ORDI >15	100 (79 to 100)	75 (45 to 92)	80 (NR)	100 (NR)	88 (NR)
		MM-ORDI 24.81 PSG-ORDI >30	79 (NR)	96 (NR)	95 (NR)	82 (NR)	87 (NR)
Supporting evidence (novel device compared to PSG)							
Pillar et al (2020), WatchPAT 200U	84	WP AHI ≥15, PSG AHI ≥15	85 (NR)	70 (NR)	78 (NR)	79 (NR)	0.86 (NR)
Tauman et al (2020), WatchPAT 200U	101	WP AHI ≥ 5 PSG AHI ≥ 5	96 (90 to 99)	25 (1 to 81)	NR	NR	NR
		WP AHI ≥ 15 PSG AHI ≥ 15	88 (79 to 94)	63 (38 to 84)	NR	NR	NR

[a] accuracy values for this cut-off, indicating test positivity for OSAHS (mild, moderate, severe), were estimated by the EAG based on data in the study publication
[b] post-hoc optimisation of the diagnostic cut-off points for the novel device against reference standard cut-offs
[c] described as balanced accuracy

AHI Apnoea-hypopnoea index; MM mandibular movements; MSSS Michele Sleep Scoring System; NOM NightOwl Mini (disposable); NOR NightOwl Reusable; NPV Negative predictive value, NR Not reported; ODI Oxygen Desaturation Index; ORDI Obstructive Respiratory Disturbance Index; PAT Peripheral Arterial Tone; PPV positive predictive value; Pts patients; PSG Polysomnography, RDI Respiratory disturbance index; WP WatchPAT

Novel devices referenced to PSG

Table 9 also shows the available diagnostic accuracy data for AcuPebble SA100, Brizzy, NightOwl, Sunrise and WatchPAT, referenced to sleep-laboratory PSG in all studies with the exception of Kelly et al 2022 (Sunrise) which referenced to home based PSG.

Of note, one of the NightOwl studies (Lyne et al 2023) evaluated both the reusable model (NOR NightOwl Reusable) and the disposable model (NOM NightOwl Mini) against PSG. Of the two models, only the disposable NightOwl is intended for use in the UK, but for completeness we report estimates for both models. Likewise, the NightOwl model assessed by Van Pee et al and Massie et al (2018) is the reusable version. (NB. The company confirmed to NICE that the only difference between the two NightOwl devices is whether the battery can be re-charged. The sensors and software are identical). Thus, of the four NightOwl studies, only one appears to include the version to be launched in the UK (i.e. Lyne et al 2023, NightOwl Mini disposable). The other studies evaluate a device which is near identical but is not intended for regulatory approved use in the UK.

Diagnostic accuracy estimates are presented for all studies, except:

- Massie et al (2022) which reported the accuracy of the novel device (NightOwl) in determining the REM sleep stage, as opposed to accuracy to detect OSAHS.
- The studies evaluating WatchPAT 300 (Mueller et al 2022) and WatchPAT ONE (Storey et al 2022) did not report diagnostic accuracy, hence we report accuracy estimates for the predecessor version, WatchPAT 200U (as explained earlier in section 4.1). These estimates were taken from the two WatchPAT 200U studies included in this review as supporting evidence to be considered in situations such as this (Pillar et al (2020) and Tauman et al (2020)).

Sensitivity and specificity estimates from all but one of the studies in Table 9 (Pillar et al 2020) are used as input parameters for the assessment of diagnostic accuracy in base case or scenario analyses for our economic model (see section 5.7.3).

As can be seen, the sensitivity and specificity estimates vary across the studies and also within studies at different severity cut-offs. Sensitivity was generally high, in the range 80 to 100%, and fell below 80% in just two studies (at one cut-off each from (Martinot (2017) and Kelly et al (2022)). In contrast, specificity was more variable with estimates ranging from 25% to 100%, with more estimates in the 70% to 80% range than was the case for

sensitivity. In notable cases (e.g. Kelly et al (2022); Tauman et al (2020)) available confidence intervals for sensitivity and specificity are wide, indicating greater uncertainty in the estimates. This is due to relatively small sample sizes, particularly the case for Kelly et al (2022). Caution is required when interpreting the accuracy estimates for Tauman at the AHI 5 cut-off, when digitising the scatterplot to obtain the summary estimates we were able to extract 100 of the 101 datapoints. These estimates are at increased uncertainty.

AcuPebble SA100 and Sunrise were the only two novel devices in this diagnostic assessment in which evidence was available comparing the novel device to RP and to PSG. In the two AcuPebble studies (Devani et al, 2021; Sanchez Gomez et al, 2024) the diagnostic performance metrics (sensitivity and specificity) were similar, demonstrating consistency in device performance irrespective of the reference test approach and setting. For Sunrise, diagnostic performance was not reported by Alsaif et al (2023), thus it is currently unclear how similar, or otherwise, the respective comparisons to RP and PSG would be for this outcome.

We suggest caution in making inferences about the relative superiority in diagnostic performance between the respective novel devices. The devices have not been formally compared in the same study with the same population and there is no formal statistical analysis to confirm any differences or equivalence between them.

It is also important to note that at least three of the studies performed post-hoc optimisation of the diagnostic cut-off points for the novel device against reference standard cut-offs (Kelly et al 2022; Martinot et al, 2017; Pepin et al, 2020). In this approach, optimal cut-offs on the ROC curve (defined as maximum sensitivity and specificity values simultaneously) are assessed and diagnostic performance metrics for this cut-off are estimated. However, this approach can be open to selective reporting of results from the cutoffs that perform well, thus over-estimating diagnostic accuracy.

Pepin et al (2020), for example, sought to optimise the clinical performance of the RDI derived from the Sunrise system analysis in ruling in a diagnosis of OSA at the two reference thresholds of PSG of at least 5 events/h or at least 15 events/h (leading to the classification of participants as not having OSA, or having OSA with comorbidities or having OSA irrespective of comorbidities, respectively). They used ROC curves and defined the trade-off between true-positive rates and false positive rates at PSG-RDI of at least 5 events/h and at least 15 events/h. The optimal diagnostic cutoff was adjusted, and the diagnostic PSG-RDI cutoffs of at least 5 events/h and at least 15 events/h were extrapolated to Sr-RDI cutoffs of

at least 7.63 events/h and at least 12.65 events/h. Whilst the results of the Pepin et al study can be at increased risk of bias subsequent studies using the cut-offs established by Pepin would not be at such risk because the thresholds will be pre-specified. (NB. The Sunrise manufacturer confirmed to NICE that the device uses pre-specified thresholds established in Pepin et al., 2020).

Notably, Kelly et al 2022, a more recent evaluation of Sunrise, did not report using pre-specified thresholds from Pepin et al., 2020. Instead, a post hoc analysis was done to optimise the cut-off points of MM-ORDI for diagnostic decisions, compared with reference standard cut-off values of obstructive PSG-ORDI. It is unclear why Pepin's thresholds were not used, but it might be because the diagnostic indices are not the same (Pepin used RDI, Kelly used ORDI).

4.5.2 Agreement / concordance (people over 16 years of age)

Twelve studies reported measures of agreement between the novel device and reference standard/comparator test.[19 21 23-27 29 17 15 30 31 22]

Agreement was assessed using standard statistical approaches such as Bland-Altman plots and by comparing mean values for the novel device and comparator on the AHI and estimating the mean difference between them and limits of agreement. In most studies satisfactory agreement between tests was reported.

4.5.3 Impact on clinical decision-making (people over 16 years of age)

One study, (Alsaif et al (2023), reported this outcome, for Sunrise (MM).

[redacted]

4.5.4 Time to interpret device outputs (people over 16 years of age)

Only two studies reported estimates of the time it took for sleep study data to be scored and a diagnosis reached (Alsaif et al (2023); Devani et al (2021). Devani et al (2021) estimated that manual scoring of respiratory polygraphy signals (the comparator technology) by experts in order to issue a diagnosis took 60–120 minutes to complete. The EAG assumes this range of estimates were based on all manual scoring of all patients in the study (the source of the estimate is unclear in the study publication). The novel device evaluated in this study, AcuPebble SA100, uses fully automated signal scoring which, we understand, is an

instantaneous process at the end of the sleep study. Devani et al (2021) estimate that zero time is required for the analysis of signals to issue a diagnosis. However, the EAG notes that sleep specialists would still need time to review the results of the automated sleep report, but no estimate of this appears to have been included in the study publication. We use expert clinical opinion to inform our assumptions about this parameter in the cost-effectiveness analysis (section 5.7.6)

Alsaif et al (2023) reported that

[redacted]

4.5.5 Time to diagnosis or starting treatment (people over 16 years of age)

Only one study assessed time to making a treatment decision, the Sunrise OSA Trial (SOSAT) (Alsaif et al 2023). Patients received novel and standard testing simultaneously during a single night and were randomised to receive their treatment decision based on either the novel device (Sunrise, MM monitoring, autoscored) or the comparator (home RP, manually scored).

[redacted]

[redacted]

4.5.6 Test failure rate (people over 16 years of age)

Eleven of the 14 included studies reported the number of sleep study tests (one night of testing for one person with one or more devices simultaneously) which failed to produce a valid diagnostic outcome (this could apply to the novel device and the comparator device) (Table 62 in Appendix 4). These were described as test failures in most studies, but other terms were used, such as ‘technically unacceptable’ or ‘inadequate tests’, or simply ‘exclusions’. Broadly speaking these terms have a similar meaning, that is, lack of a valid diagnostic outcome, but the reasons given for this varied. Some studies reported the criteria they used to determine the validity of a sleep study, for example recommendations on minimum sufficient sleep time by device manufacturers. Recommendations from clinical

guidelines were also used. For example, the AASM[1] defines home sleep apnoea tests as technically adequate if at least four hours of analysable signal can be obtained; for PSG all of the channels should be interpretable by sleep technicians, which in practice means that all attachments, such as nasal cannula, pulse oximeters, remain in place.

The novel device test failure (or equivalent definition) rates varied across the studies, ranging from 0% in Devani et al, (2021),[19] and Pepin et al (2020), [26] to around 18% (Van Pee et al (2022).[14] Most studies reported rates around the 10% level. Table 62 in Appendix 4 summarises the reasons given for failed/inadequate tests. We use the study author's own descriptions rather than attempt to classify the failures ourselves, as the descriptions given in study publications are often open to interpretation. Having said that we observe that failures tend to be technical or non-technical, with the former including occurrences such as signal acquisition errors and the latter covering a variety of factors including mistakes made in administration of the study protocol, or device operator errors, as well as insufficient sleep time for valid results. These factors are not necessarily independent of each other, but are often related, for example, an interrupted signal can be caused by the sensor not staying in position.

Given the relatively low number of available studies it's difficult to identify meaningful patterns or trends in test failures between novel test devices and their comparators. The evidence is mixed, in some studies rates were similar for novel and comparator tests, in others the novel devices had fewer failures, and in other studies the opposite was found.

Test failure rates are one of the outcome parameters included in the cost effectiveness analysis in this report (see 5.7.5). As we explain later, the economic model only includes test failures which potentially incurs a cost to the NHS, e.g. to organise a repeat test.

4.5.7 Use of healthcare resources and costs (people over 16 years of age)

Two studies reported an assessment of reported an assessment of health care resources and costs (Devani et al (2021); Storey et al (2022)).

Devani et al reported brief data on use of healthcare resources in their evaluation of AcuPebble SA100, including time taken for cleaning, device preparation and training. AcuPebble was shown to be more efficient in terms of resource use (Table 63 in Appendix 4 summarises the estimates given).

Storey et al (2022) compared the WatchPAT ONE to home RP (NOX T3) in terms of four measures of resource efficiency: number of missed appointments; mean patient travel time (minutes); cost per appointment, and mean staff time per appointment. Table 64 in Appendix 4 reports the results, showing that WatchPAT ONE was more efficient than NOX T3 on three of the four measures. However, the cost per appointment was higher. The authors speculated that WatchPAT ONE has the potential to improve clinic efficiency, for example by reducing waiting times.

4.5.8 Number of repeat sleep studies done (people over 16 years of age)

When the results of a test are invalid, eg. due to test failure, it will be necessary to repeat the test during another night. Whether or not repeat tests were done was rarely mentioned, with just two studies reporting information.

Devani et al (2021) reported that no repeat tests using the AcuPebble SA100 were conducted, with patients required to return the device to the hospital the day after the overnight sleep study.

Mueller et al (2022)[17] reported that two study participants repeated the WatchPAT 300 novel device sleep study due to operating errors (participants forgot to switch on the device). The comparator, home RP was repeated by 8 participants following test failures arising from insufficient recording time (n=2), failure to start the device (n=2) and loss of the nasal pressure sensor or inadequate examination time (n=3); the reason for the repeat test was not stated for the remaining participant.

4.6 Clinical outcomes (people over 16 years of age)

None of the studies reported clinical outcomes in terms of mortality or morbidity (though see section 4.7 for patient views on discomfort during sleep studies).

4.7 Patient reported outcomes (people over 16 years of age)

4.7.1 Health-related quality of life (people over 16 years of age)

None of the studies reported assessments of health-related quality of life.

4.7.2 Ease of use and acceptability for patients and carers (people over 16 years of age)

Three studies provided patient-reported outcomes related to ease of use and acceptability. Devani et al., (2021) assessed patient acceptability and usability for the AcuPebble SA100 device on completion of the home-based study via a voluntary questionnaire. One hundred and twenty-three patients out of the 150 cohort completed the usability questionnaire. Results are shown in Table 65 in Appendix 4. In summary, the vast majority of patients found it easy to use the mobile app and follow the app instructions for using AcuPebble SA100. Furthermore, they found it easy to attach the AcuPebble SA100 sensor, finding it more comfortable and easier to attach than the sensors of the comparator (RP).

Mueller et al., (2022) reported perceived quality of sleep and test-related discomfort with the WatchPAT 300 device and with RP used on separate nights (see Table 66 and Table 67). Of 56 patients, 54 provided questionnaire responses regarding testing with WatchPAT and 55 regarding testing with RP (see Table 66 in Appendix 4). Approximately three quarters of patients reported that falling asleep was not disturbed when using WatchPAT and they slept well during the night. In contrast, nearly two thirds of patients reported that falling asleep was disturbed when using RP and they did not sleep well during the night. Nearly four times more patients lost sensors during testing with RP than with WatchPAT (20% versus 6%), however nearly three times more patients experienced pain with WatchPAT than with RP (13% versus 5%). The reasons for pain differed between the devices - with WatchPAT the pain was finger and finger-probe related, while with RP the pain was due to the nasal cannula dynamic pressure measurement and to the device itself. Most patients (70%) reported that during testing they were not woken up by WatchPAT, whereas 50% said they were not woken by RP. The perceived number of awakenings was statistically significantly lower during testing with WatchPAT (mean 0.62, range 0 to 6) compared to RP (mean 1.8, range 0 to 10) ($p=0.004$). After experiencing testing with each type of device, 80% of patients said they slept better with WatchPAT than with RP. Furthermore, 88% of patients expressed a preference for WatchPAT over RP if they were to undergo future testing. Patient self-reported overall sleeping comfort for all 56 patients was also reported and was better with WatchPAT compared to RP (see Table 67).

Alsaif et al., 2023 found that

[redacted]

[redacted]

It is unclear whether a fourth study, Sanchez Gomez et al (2024), which compared AcuPebble SA100 to PSG in a hospital setting, assessed ease of use and acceptability. A preliminary study report[15] stated the aim was *"to identify whether there was something specific to the Spanish population in the user journey of the app (related to the wording since the previous trial was done with an app in English) that might be difficult to understand for patients".* However, no usability/acceptability data was included in the study report or the subsequent journal publication.[20]

4.7.3 Patient and carer experience (people over 16 years of age)

One study (Alsaif et al., 2023), which assessed the home use of Sunrise, reported on [redacted]. This study found that

[redacted]

The following sections 4.8, 4.9 and 4.10 provide a summary of the outcomes reported by the studies of novel devices in **the children and young people aged 2 to 16 years** population group.

4.8 Intermediate outcomes (children and young people aged 2 to 16 years)

Below we present outcome data for diagnostic accuracy, agreement between diagnostic tests and test failure rates. There were no available data for the following intermediate outcomes from the NICE scope: time taken to interpret device outputs, reach a diagnosis, and/or start treatment; the number of repeat sleep studies done; use of healthcare resources.

4.8.1 Diagnostic accuracy (children and young people aged 2 to 16 years)

Two of the three studies reported the accuracy of the novel devices in diagnosing OSAHS (NCT 04031950 (2019); Martinot et al 2022).

[redacted]

[redacted]

Table 10

[redacted]

Martinot et al (2022) reported the performance of Sunrise (mandibular movements monitoring) in diagnosing and classifying the severity of OSA compared to laboratory PSG. A total of 140 children were consecutively referred for clinical suspicion of OSA. The hypothesis was that Sunrise-derived ORDI (obstructive respiratory disturbance index) would provide satisfactory clinical accuracy to rule in a diagnosis of OSA, using diagnostic criteria from the International Classification of Sleep Disorders, Third Edition (ICSD‐3) and AASM severity thresholds.

Table 10 Accuracy of novel devices in diagnosing OSAHS (children and young people aged 2 to 16 years)

Author	Index test	No. pts	Cut-offs		Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Accuracy % (95% CI)
			Index test	Reference std					
NCT04031950 (2019)[a]	AcuPebble	[redacted]	[redacted]		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
			[redacted]		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Martinot et al (2022)[d]	Sunrise (MMs)	140	Sr-RDI 5.75	PSG OAHI ≥1	83 (78 to 86)	53 (48 to 59)	64 (59 to 68)	75 (70 to 80)	68 (65 to 71)
			Sr-RDI 9.61	PSG OAHI ≥5	90 (87 to 93)	80 (76 to 84)	82 (78 to 86)	89 (85 to 92)	85 (82 to 88)
			Sr-RDI 13.07	PSG OAHI ≥10	100 (100 to 100)	88 (0.84 to 0.91)	89 (0.86 to 0.92)	100 (100 to 100)	94 (92 to 96)

[a] Preliminary results of a study in progress; results are from one of the two study centres.
[b] All paediatric AASM events (including the post-sigh central apnoea) lasting for at least two breaths
[c] All scored paediatric AASM events excluding post-sigh apnoea and events associated with movement
[d] Diagnostic performance estimates in this study are medians

AASM American Academy of Sleep Medicine; AHI Apnoea-hypopnoea index; MM mandibular movements; ORDI Obstructive; Pts patients; RDI Respiratory Disturbance Index; Sr Sunrise

In a post-hoc analysis the study optimized the diagnostic performances of Sunrise RDI in ruling-in a diagnosis of paediatric OSA at three cut-off thresholds of PSG_OAHI ≥ 1 events/h, PSG_OAHI ≥ 5, events/h, or PSG_ OAHI ≥ 10 events/h. (The PSG_OAHI comprises obstructive and mixed apnoeas/hypopnoeas but excludes respiratory effort related arousals (RERA).

The area under the ROC curves (AUC) targeting PSG OAHI ≥ 1, PSG OAHI ≥ 5, or PSG OAHI ≥ 10 reached 0.75 (95%CI: 0.72–0.78), 0.90 (0.86–0.92), and 0.95 (0.90–0.99), respectively. Optimized best thresholds for Sunrise RDI were 5.75, 9.60 and 13.07 for the three PSG OAHI severity thresholds respectively. The corresponding accuracy estimates were 66%, 85% and 94% respectively.

Table 10 reports the summary diagnostic performance measures expressed as medians and 95% CIs. Sensitivity increased at each severity threshold reaching 100% at PSG OAHI ≥ 10. Specificity was relatively low at 53% for PSG OAHI ≥ 1 but increased to just below 90% at the highest severity threshold. Caution is advised in the interpretation of these findings given the post-hoc optimisation of Sunrise RDI thresholds.

Caution is also advised when comparing the estimates from the two studies in Table 10 particularly because they use different reference standard tests (home RP in one study, laboratory PSG in the other) and the use of mean performance estimates in one study and medians in the other.

4.8.2 Agreement / concordance (children and young people aged 2 to 16 years)

Two of the three studies in this population group reported measures of the agreement between novel device and reference standard (Martinot et al 2015; Martinot et al 2022)

In the 2015 study, 33 children (median age 5 years old) with suspected OSA were concurrently enrolled and received laboratory-based PSG testing concurrently with the monitoring of mandibular movements using the Brizzy device. The aim of the study was to explore the relationship between the mandibular movements observed during sleep in children with adenotonsillar hypertrophy and the presence of respiratory effort assessed with the pulse transit time measurement (a more established indicator of respiratory effort). They also examined the temporal relationship between mandibular movements and pulse transit times during OAH and central sleep apnoea. Several patterns of mandibular movement were compared to concomitant changes in pulse transit time suggestive of OSAS. The publication

reports a variety of data values and pattern analyses for subtypes of mandibular movement and associations/correlations with respiratory effort and mixed apnoeas or hypopnoeas from the obstructive sleep apnoea-hypopnoea index (OAHI).

Most sleep events scored were obstructive apnoea/hypopnoea (as opposed to central apnoea) with a median (95% CI) of 5.6 (6.4–11.3) events per hour. Overall, 94% of obstructive apnoea/hypopnoea events were associated with mandibular movements. The rate of mandibular movements per hour (MML [mandibular movement large] or MMO [mandibular movement mouth opening] or MMS [mandibular movement sharp and sudden] correlated significantly with the OAHI (Spearman rho = 0.511; p = 0.003).

The authors concluded that mandibular movements analysis is helpful to detect respiratory effort during sleep in children with upper airway obstruction, and can be regarded as a sensitive tool to identify and characterise sleep disordered breathing. Due to the specialised topic of investigation this study is less relevant to the decision problem than others in this systematic review, but can be seen as earlier evidence in support a novel mechanism to diagnose OSAHS (i.e. mandibular movements), which has been evaluated for clinical diagnostic performance in later studies (albeit most of the studies have been in the adult population).

The Martinot et al (2022) study reported acceptable agreement between the two methods (PSG vs. Sunrise) in estimating RDI, as suggested by an intraclass correlation coefficient (ICC) of 0.79 (95% CI: 0.72 to 0.85; p< .001). A Bland–Altman analysis between Sunrise RDI and PSG RDI yielded a median difference between the two tests of 1.57 events per hour with a CI including the zero value and no systematic bias between the two measures.

4.8.3 Test failure rates

Two of the three studies in this population group reported the number of tests classed as having failed (Martinot et al (2022); NCT 040319-50 (2019)). Martinot et al (2022) reported that fifteen of the 155 children (9.6%) enrolled were excluded for reasons of:

- Incomplete data (7 children),
- Total sleep time less than 4 hours (3 children),
- Technical failures (5 children)

It is not clear whether these apply to the Sunrise device or the comparator (lab PSG) or both. However, it is stated that there were three technical failures related to the connected Sunrise system due to the loss of the wireless connection.

[redacted]

4.9 Clinical outcomes (children and young people aged 2 to 16 years)

None of the studies reported clinical outcomes in terms of mortality or morbidity.

4.10 Patient reported outcomes (children and young people aged 2 to 16 years)

[redacted]Table 68[redacted]

4.11 Ongoing studies

The EAG identified three ongoing studies relevant to this review, all of which are evaluating the Sunrise novel device.[39-41] These studies are summarised in Table 58 in Appendix 2. Of the three studies, two include people over 16 and one includes children. One of two studies in people over 16 is being conducted in Scotland.[40] This study plans on recruiting 100 adults with suspected OSAHS to compare Sunrise to an unspecified "detailed sleep test". However, details of the planned conduct and setting of the tests, the outcome measures and the estimated completion date are not reported. The second study is a considerable larger study in France, with a planned recruitment of 848 patients and estimated completion date of March 2024.[39] In this study, patients are randomised to home based use of Sunrise or to lab or outpatient PSG. One of the planned outcomes for this study is the time to diagnosis or starting treatment, which could be informative given the extremely limited evidence we found in this review (see section 4.5.5)

The ongoing Sunrise study in children is being conducted in the UK and plans on recruiting 100 patients.[41] In our review, all three included studies of children,[32 35 42] including one study of Sunrise,[35] compared the simultaneous use of the novel devices to PSG in a sleep laboratory.[32 35 42] In contrast, this ongoing study is comparing the diagnostic accuracy of home use of Sunrise to home cardio-respiratory polygraphy with Transcutaneous Carbon Dioxide monitoring, with the notable exception that children with significant co-morbidities or aged <9 years of age will undergo the same tests but in a sleep laboratory setting.

5 ECONOMIC ANALYSIS

The aim of this chapter is to assess the cost-effectiveness of novel home-testing devices for diagnosing OSAHS. It comprises:

- A systematic review of cost-effectiveness studies (section 5.1)
- A systematic review of health-related quality of life (utility) (section 5.2)
- An overview of economic evidence from company submissions to NICE (section 5.3)
- A model developed by the EAG to evaluate novel home-testing devices for diagnosing OSAHS in people aged over 16 years (sections 5.4 to 5.10)
- A discussion of the challenges and potential for developing an economic model for people aged 16 years and under (section 5.11)

5.1 Systematic review of cost-effectiveness evidence

The aim of this systematic review was to identify studies reporting on the cost-effectiveness of the novel home-testing devices compared to home respiratory polygraphy or pulse oximetry not using these devices. The results of the systematic review were intended to inform our modelling of the research question; and to provide alternative analyses, preferably in the UK context, with which we could cross-validate the findings of our model.

5.1.1 Methods for review of economic studies

The database searches were carried out on 24th May 2023 and updated on 25th September 2023. The full search strategies are shown in Appendix 1b (Table 53). The database searches were based on the search strategy used for the systematic review of clinical effectiveness, with the addition of published filters to identify economic evaluations, estimates of resource use and costs, and health-related quality of life. The relevant population, interventions and comparators are the same as for the systematic review of test performance and clinical effectiveness (section 3.1), but the inclusion criteria differed in terms of the relevant study design and outcomes. We only included full economic evaluations (cost-effectiveness analysis (CEA), cost-utility analysis (CUA) or cost-benefit analysis (CBA)) comparing home testing with a named novel device (Acupebble SA100, Brizzy, NightOwl, Sunrise, WatchPAT 300 or WatchPAT ONE) to home testing with an oximetry or respiratory polygraphy device (not using one of the six named devices) in people with suspected OSAHS. Any study meeting our inclusion criteria, whether it was a trial-based economic evaluation, a decision analytic model or other type of evaluation was considered. Studies that only reported resource use or cost were excluded, but these studies were considered separately as possible sources of data for our model. Two reviewers independently screened all titles and abstracts identified from the literature searches. Both

reviewers then independently screened the full texts of any studies included at title and abstract screening, using pre-defined inclusion and exclusion criteria (see Appendix 6 Table 69) . All disagreements were discussed and resolved by the two reviewers.

The EAG planned to extract data related to the study design, methods, parameter sources, relevant model inputs and results of the included cost-effectiveness studies. The credibility of the included cost-effectiveness studies and their relevance to current UK practice were assessed using a pre-defined checklist, shown in Appendix 6 Table 71. This checklist was based on the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [45] and Philips and colleagues' [46] checklists.

5.1.2 Results of the review of economic studies

We identified 768 references from the literature searches. After title and abstract screening, 23 references were selected for full-text screening (Appendix 6 Figure 10). Of these, none were found to meet our inclusion criteria. The main reasons for exclusions were that the intervention was not one of the named novel devices, the comparator was not home oximetry or home respiratory polygraphy, or it was not a full economic evaluation (see Appendix 6 Table 70).

5.1.3 Overview of other published economic studies of interest

Although not meeting our inclusion criteria, five studies with the potential to inform our model structure and parameters were identified from the systematic review. They all related to an adult population and can be grouped into evaluations of tests for OSAHS (n=3) [47-49] and evaluations of CPAP treatment for OSAHS (n=2).[50 51] The characteristics of these studies are summarised in Table 11 below, with brief descriptions of each study in Appendix 6 Table 71. A sixth study, conducted for the NICE clinical guideline on the diagnosis and management of OSAHS and obesity hypoventilation syndrome in people over 16s, was identified from other sources as being of particular importance (the NG202 economic model)[52]

The NG202 economic model is most relevant to our decision problem, and it provided relevant data and assumptions that we used to inform the EAG model structure and parameters. The NG202 model compared eight strategies, defined by the diagnostic sleep study and extent of CPAP treatment for those diagnosed with OSAHS:

1. Home oximetry (and CPAP treatment for all diagnosed)
2. Home respiratory polygraphy (and CPAP treatment for all diagnosed)

3. Hospital respiratory polygraphy (and CPAP treatment for all diagnosed)
4. Home oximetry screening followed by home respiratory polygraphy for those who were negative for OSAHS after oximetry screening (and CPAP treatment for all diagnosed)
5. Home oximetry (and CPAP treatment for moderate and severe only)
6. Home respiratory polygraphy (and CPAP treatment for moderate and severe only)
7. Hospital respiratory polygraphy (and CPAP treatment for moderate and severe only)
8. Home oximetry screening followed by home respiratory polygraphy for those who were negative for OSAHS after oximetry screening (and CPAP treatment for moderate and severe only)

The only home studies included in the NG202 economic model were oximetry and respiratory polygraphy, not novel home-based devices. Hence the NG202 economic model does not meet our inclusion criteria for the systematic review of economic evaluations.

The NG202 economic model takes a linked evidence approach, consisting of a decision tree covering the diagnostic pathway, and Markov models representing the long-term impacts of treatment. Within the decision tree, a distinction is made between individuals who truly have OSAHS (as defined by an AHI score ≥ 5, or ODI ≥ 5 for oximetry) and those who do not have OSAHS (AHI score <5, or ODI < 5 for oximetry). The decision tree further differentiates OSAHS by severity, based on the American Academy of Sleep Medicine Task Force 1999[53] thresholds of event frequency: mild (5 ≥ AHI or ODI < 15), moderate (15 ≥ AHI or ODI < 30) and severe (AHI or ODI ≥ 30).

Depending on estimates of sensitivity and specificity for the diagnostic strategies evaluated, individuals are classified as true positives (correctly identified by the diagnostic test as having OSAHS), true negatives (correctly identified by diagnostic test as not having OSAHS), false positives (incorrectly identified by the diagnostic test as having OSAHS), and false negatives (incorrectly identified by the diagnostic test as not having OSAHS). These classifications are further distinguished by the severity of the true OSAHS condition, as well as the severity result produced by the diagnostic test. For instance, an individual with mild OSAHS may be misdiagnosed as having moderate or severe OSAHS; while an individual with moderate or severe OSAHS may be misdiagnosed as having mild OSAHS.

The decision tree uses sensitivity and specificity estimates at two thresholds (AHI or ODI ≥5 and ≥15) to model the accuracy with which oximetry and respiratory polygraphy would categorise patients with no, mild, moderate and severe OSAHS (see NG202 Economic

analysis report Figure 1 and Table 4[52]). It is assumed that false negatives who truly have moderate or severe OSAHS will go on to have additional sleep study testing, as they are likely to continue to be symptomatic. No further testing is assumed for people with mild OSAHS who are falsely diagnosed as having no OSAHS. The costs of any additional testing are accounted for in the NG202 model.

The diagnosed severity of OSAHS informs the type of treatment, and subgroups defined by true underlying severity (or absence) of OSAHS and treatment transition into Markov models, which estimate long term costs and outcomes. Treatment of OSAHS is assumed to impact health outcomes and costs through three mechanisms:

- Improved HRQoL due to improved symptoms of sleepiness as measured by the ESS
- Reduced risk of longer-term cardiovascular events via a reduction in systolic BP
- Reduced risk of road traffic accidents that are associated with untreated OSAHS.

The Markov models consist of 12 health states: OSAHS, five acute cardiovascular event states (for stable angina, unstable angina, myocardial infarction, transient ischaemic attack (TIA) and stroke), five post cardiovascular event states, and death (Figure 2 of NG202 economic analysis report [52]). Individuals are assumed to have at most one cardiovascular event over the lifetime time horizon. The risks of a slight, serious or fatal road traffic accident are modelled as events from any of the alive health states. The cycle length is 12 months, with a half-cycle correction. The economic model was developed in consultation with the guideline development committee, and it has a similar form as models used in previous evaluations, including TA139.[54]

As the NG202 model is recent and directly relevant to the evaluation of novel home-testing devices, we took it as the starting point for development of the EAG adult economic model. Further details on the methods used in the NG202 model are presented in the description of the EAG adult model below (section 5.5).

Table 11 Details of economic studies of interest

Study, Country	Decision problem	Type of study	Time horizon	Potentially relevant evidence	Limitations on applicability to EAG model
Phua et al 2021 [49] Singapore	Evaluate differences in costs between WATCHPAT 200 and PSG	Retrospective review of sleep studies	From initial sleep study to treatment	Time to treatment for WatchPAT200 vs PSG	Earlier version of WatchPAT, based in Singapore, different patient mix between WatchPAT and PSG
Di Pumpo et al 2022 [47] Italy	Cost-minimisation analysis of WatchPAT 200 for hospital attendance vs telemedicine	Cost model	From initial appointment to diagnosis	Limited	Lack of detailed reporting and Italian setting limits use of data from this study
Geessinck et al 2018 [48] Netherlands	Cost-utility analysis of screening tool (DiagnOSAS) vs no use of DiagnOSAS in primary care to aid diagnosis of OSA in men	Markov model	10 years (5 years in scenario analysis)	Utilities, transition probabilities Their justification of a 10 year time horizon based on lack of evidence on CPAP adherence after this time informed a scenario analysis in the EAG model	Many sources taken from previous analyses.
NG202 [52] UK	Cost-utility analysis of different diagnostic pathways for OSA and treatment for OSA	Decision tree and Markov model	Lifetime	Model structure and parameter estimates directly relevant to the EAG model, especially given the NG202 model was recently developed for NICE	Some parameter estimates are from old data sources
McMillan 2015 et al [51] UK	Cost-utility analysis of CPAP and best supportive care vs best supportive care only in adults ≥ 65 years	Economic evaluation alongside RCT Markov model	12 months Lifetime	Model structure, treatment effectiveness, utilities, 12-month treatment adherence	Potentially superseded by more recent data used in NG202 [52]
Guest et al 2008 [50] UK	Cost-utility analysis of CPAP vs no treatment for OSAHS	Markov model	Lifetime	Limited	The age of the study, and its data sources limit its applicability
Abbreviations: CPAP continuous positive airway pressure; EAG external assessment group; PSG polysomnography					

5.2 Systematic review of health-related quality of life studies

The aim of this systematic review was to identify studies reporting on the heath-related quality of life (HRQoL) associated with OSAHS to inform utility values for use in the EAG economic model.

5.2.1 Methods for review of health-related quality of life

We undertook searches to identify data on HRQoL for adults or children with suspected or diagnosed OSAHS. The search strategy is shown in Appendix 1c Table 54. The population, interventions and comparators are the same as for the systematic review of test performance and clinical effectiveness, with the addition of a published filter for HRQoL. Only primary research studies were included. We planned to extract data related to the study design, country and sample size, HRQoL instruments used, and health states assessed.

5.2.2 Results of the review of health-related quality of life studies

The database searches identified 2,095 potentially relevant references, of which 261 met our inclusion criteria at title and abstract screening (see Appendix 7 Figure 11*)*. Due to the high number of studies in adult populations, we prioritised articles for full-text screening that reported EQ-5D, SF-6D, HUI, QWB and 15D outcomes and/or studies based in the UK. All studies conducted in children that met our criteria at title and abstract screening were included in the full-text screening. This resulted in a short-list of 59 references for full-text screening, of which 44 were excluded from further consideration. The excluded references and reasons for exclusion are shown in Appendix 7 Table 73.

HRQoL studies in an adult population

From the short-list of 59 EQ-5D, SF-6D, HUI, QWB, 15D and UK-based studies, we identified 14 studies in an adult population with the potential to provide utility estimates for the EAG model (Table 12). This evidence is mixed in terms of the severity of OSAHS. Only one study (Skirko 2020) [55] reported utility by severity for newly diagnosed and untreated individuals. However, the results lack face validity: with estimates for mild and severe OSA associated with a utility of 0.60, and moderate OSA with a utility of 0.61). This does not reflect the literature on QoL associated with severity of OSA, which does suggest that impacts do differ by severity.[56] We therefore decided to take the same approach as in NG202 and use the mapping algorithm developed by McDaid et al 2009 for the NICE appraisal of CPAP.[54] This provides internal consistency between utilities for different health states, but also consistency with the approach taken in the NICE NG202 guideline[52] and

CPAP appraisal.[54] See section 5.7.10 below for an explanation of the McDaid mapping algorithm.

HRQoL studies in children

Only one study in children met our inclusion criteria, Sakki 2021.[57] This was a Finnish study evaluating the impact on HRQoL and resource use for children undergoing tonsillotomy due to sleep-disordered breathing. Seventy-five children aged 8-11 years old were asked to complete the 17D before surgery, and 6 and 12 months after surgery. The 17D is a preference-based instrument designed to assess utility in children.[58] It was developed from the 15D instrument for adults and includes 17 health attributes (dimensions), with values provided by parents of 8-11 year olds (in Finland). In the Sakki 2021 study, parents were asked to help their children complete the questionnaire, and to provide data on resource use. Of the 75 children in the study, only 37 returned completed 17D questionnaires at 6 and 12 months. There were statistically significant differences in utility derived from the 17D questionnaire between baseline and 6 months after surgery: 0.933 (95%CI 0.931, 0.953) to 0.956 (95%CI 0.942, 0.97). It is also reported that at 12 months, the higher utilities remained. Sakki (2021) [57] compared these results with those from an earlier study on children who had received tonsillectomy due to sleep-disordered breathing. They found that there was no statistically significant difference between utilities at any time point between those who had tonsillotomy and the historical group who had tonsillectomy.

The EAG note that utility data from only 49% (37/75) of participants are reported in the article, with no information on those participants who do not have utility data at all three time-points. The study considers a sub-group of the children within the scope of this assessment, those eligible for tonsillotomy due to sleep-disordered breathing, and the severity of sleep-disordered breathing in the study population is not reported.

Table 12 Studies with the potential to provide utility estimates for the EAG model

Author, year	Country	Study design	Study objective	Population	Sample size	HRQoL instrument	Health states	Limitations/ Conclusions
Adult population								
Kataoka, 2017 [59]	Japan	Cross-sectional	Compare QoL in people with OSA depending on presence of locomotive syndrome or no	Mixed population treated with CPAP from 1 hospital	1195 (1030 males; 165 females)	EQ-5D	Diagnosed with OSA	One-off snapshot of a mixed population (in terms of severity and treatment received)
Lugo, 2019 [60]	Spain	RCT	In-person vs virtual management of patients with OSA	Suspected OSA and/or refractory hypertension	186 for ITT; 154 per protocol	EQ-5D	Pre-diagnosis vs 1 year later by hospital or virtual testing/treatment	FU included groups of treated and untreated pts. EQ-5D not reported by severity
Walia, 2017 [61]	US	Retrospective	Pre- vs post PAP	Outpatients of sleep clinic; diagnosis of SDB assumed if patient stated use of PAP	2,027	EQ-5D	pre-CPAP utility post-CPAP utility	Diagnosis of OSA based on self-report use of PAP; not reported by severity
Cambron-Mellott, 2022 [62]	UK, France, Germany, Spain, Italy	Retrospective cross-sectional	Mapping of ESS to EQ-5D at one time-point	Self-reported OSA diagnosis, or experience of OSA within last 12 months	OSA without narcolepsy 2,277	EQ-5D-5L	Self-report of OSA diagnosis (or experienced OSA within 12 months)	One-off snapshot of a mixed population, also OSA status self-defined
Sanchez-de-la-Torre, 2015 [63]	Spain	RCT	CPAP compliance after 6 months for sleep unit vs primary care management of treatment	Recently diagnosed and requiring CPAP treatment	210	EQ-5D	Pre/post CPAP treatment by primary care & sleep unit managed	EQ-5D not reported by severity

Author, year	Country	Study design	Study objective	Population	Sample size	HRQoL instrument	Health states	Limitations/ Conclusions
Huber, 2021 [64]	Switzerland	RCT	Compare performance of health utility measures in OSA pts receiving CPAP (auto adjusting or fixed)	Patients with OSAS	208	EQ-5D (value set not reported); SF-6D (UK value set)	Baseline and mean change after 3, 12 &24 months	Not reported by AHI severity cut-offs
McMillan, 2015 [51]	UK	RCT	CPAP +BSC vs CPAP	Newly diagnosed, older pts	278	EQ-5D and SF-6D	Pre/1yr post CPAP or BSC treatment	Older population; results not reported by severity
Wimms, 2020 [65]	UK	RCT	CPAP +BSC vs CPAP	Newly diagnosed mild OSA	233	EQ-5D and SF-36	Before/after 3 months CPAP or BSC for mild OSA	Only mild severity
Sharples, 2014 [66]	UK	RCT	MAD vs BSC	Mild-moderate OSA	74	EQ-5D and SF-6D	Before/after treatment	Results not presented by severity
Skirko, 2020 [55]	US	RCT and observational studies	Develop and validate a utility scoring algorithm for a sleep apnoea–specific quality-of-life instrument	Newly diagnosed and untreated	500	SF-6D	Mild, moderate and severe	Mild OSA 0.60 (0.09), moderate OSA 0.61 (0.08) severe OSA 0.60 (0.08)
Rizzi 2014 [67]	Brazil	Pre-post CPAP use		AHI > 20; Naïve to CPAP	95	SF-6D	Before and 1 year after CPAP use	More severe disease
Ylitalo-Heikkila 2018 [68]	Finland	Prospective survey	Review HQRoL in people with rhinologic disease	Adults requiring rhinologic services at a single hospital	337	15D	Diagnosed with OSA	Single utility

Author, year	Country	Study design	Study objective	Population	Sample size	HRQoL instrument	Health states	Limitations/ Conclusions
Kuik et al 2023 [69]	Netherlands	Prospective cohort	Assess subjective outcomes for patients having maxillomandibular advancement surgery	Patients with severe OSAHS or treatment-refractory undergoing maxillomandibular advancement surgery	30	EQ-5D-3L	Pre- and post-surgery	Patients undergoing maxillomandibular surgery.
Pinczel et al 2023 [70]	Australia	Case series from a RCT	Long-term outcomes associated with surgery for OSAHS	Moderate-severe OSAHS who have failed CPAP	36	EQ-5D-3L	Pre- and post-surgery	Population already failed CPAP
Population of children								
Sakki, 2021 [57]	Finland	Prospective before/after tonsillotomy study, with comparison with historical tonsillectomy group	Impact on HRQoL and resource use for tonsillotomy vs tonsillectomy	5-11 year old children undergoing tonsillectomy for sleep-disordered breathing. Mean 6.7 years old; 45% female	75	17D (value set parents of 8-11 year olds in Finland)	Before/after tonsillotomy	N=37 Mean utility improved from 0.933 (95%CI 0.931, 0.953) to 0.956 (95%CI 0.942, 0.97) at 6 months; Higher utility remained at 12 months

Abbreviations: 15D 15-Dimension; 17D 17-Dimension; AHI Apnoea Hypopnoea Index; BSC best supportive care; CPAP continuous positive airway pressure; EQ-5D EuroQol-5 dimensions; ESS Epworth Sleepiness Scale; HRQoL health-related quality of life; MAD mandibular advancement device; PAP positive airway pressure; QoL quality of life; RCT randomised controlled trials; SDB sleep-disordered breathing; SF-6D short-form 6-dimensions;

5.3 Overview of economic evidence in the company submissions

One company (ResMed) submitted economic evidence for the NightOwl device. The company submitted a decision tree model in Excel comparing NightOwl with home respiratory polygraphy. The decision tree model is similar to that used in NG202 in that it separates the cohort by severity of OSA. However, while the NG202 model combines moderate and severe OSA, the NightOwl decision tree keeps mild, moderate and severe OSA separate, by applying severity-specific sensitivities and specificities. As in the NG202 model, the NightOwl decision tree assumes that individuals with moderate or severe OSA who are misdiagnosed with no OSA would have a second test. The second test is assumed to be of the same type as the first test, with a result that is independent of the first test. The diagnostic performance of NightOwl is informed by diagnostic accuracy data reported by van Pee et al 2023 [24] (see Table 13). Epidemiological estimates and performance of home respiratory polygraphy are taken from the NG202 economic report. The company use a cost of £90 (excluding VAT) per test for the NightOwl device and an additional personnel cost of £23.13 (described as 15 minutes of nurse time and 7.5 minutes of consultant time). The home respiratory polygraphy sleep study is assumed to cost £266 based on outpatient costs from the NHS National Schedule of Reference Costs (HRG DZ50Z).

Table 13 Diagnostic performance of home respiratory polygraphy and the NightOwl device

Parameter		Home respiratory polygraphy	NightOwl
AHI 5	Sensitivity	0.95	0.93
	Specificity	0.58	0.72
AHI 15	Sensitivity	0.84	0.91
	Specificity	0.89	0.76
AHI 30	Sensitivity	0.84	0.91
	Specificity	0.89	0.76
References		NICE guideline NG202: Economic analysis report [52]	van Pee et al 2023 [24]
Source: Model submitted by ResMed for the NightOwl device Abbreviation: AHI Apnoea Hyponoea index			

Results of the model (Table 14) suggest that NightOwl is cheaper than home respiratory polygraphy, with an estimated saving to the NHS of £171 per person. NightOwl is also estimated to have better diagnostic performance than home respiratory polygraphy, with lower false positive and false negative rates. Variation in the cost of home respiratory polygraphy has the largest impact in the company's sensitivity analyses (one-way deterministic with parameters decreased/increased by 20%).

The manufacturer also estimated the carbon footprint associated with each device based on estimates of greenhouse glass emissions for outpatient appointments, patient collection of

equipment in-person and postage. They assumed that patients collect the home respiratory polygraphy equipment in person from the hospital but receive the NightOwl device through the post, and that patients would return the disposable device for recycling through the post. One-way sensitivity analysis was conducted with parameters decreased/increased by 20%.

Table 14 Results from the decision tree model submitted by ResMed for NightOwl

Outcome	NightOwl	Home respiratory polygraphy	Incremental
Total cost per diagnosis	£118.80	£289.43	-£170.63
Total CO_2e emissions, kg	0.440	86.13	-85.69
True positive	0.797	0.792	0.005
False positive	0.050	0.076	-0.026
True negative	0.023	0.028	-0.005
False negative	0.130	0.104	0.026
Source: Model submitted by ResMed for the NightOwl device Abbreviation: CO_2e carbon dioxide equivalent			

The NightOwl economic model has several important limitations:

- Omission of failure rates for both arms, and the impacts of this on costs. Although it is accepted that a repeat within 10 nights would not lead to any additional costs for the test, as the NightOwl device has a 10 night battery length, there would be additional resource use implications.
- Sensitivity and specificity evidence are from the study by Van Pee which evaluated the reusable version of NightOwl, not the disposable version which the company have indicated will be commercialised in the UK. The EAG note that these estimates are likely to be more favourable to NightOwl than those we believe should be used: NightOwl disposable from the Lyne 2023 study [25].
- A higher cost for home respiratory polygraphy is used than that assumed in the EAG model (£266 rather than £212), see section 5.7.13 below.
- No explanation is given for the 15 minutes of nurse time and 7.5 minutes of consultant time. We assume that this is meant to cover time for the download of data, and the review and preparation of the diagnostic report at the hospital. It is also unclear how the cost of £23.13 for staff time was derived.
- No costs were included for posting the device to the patient's home.
- No consideration of treatment costs or longer-term impacts.
- No consideration of health outcomes beyond the diagnostic accuracy.
- Does not meet NICE reference case for economic analysis.

5.4 Decision problem for economic modelling

5.4.1 Population and subgroups

The NICE scope specifies the population of interest as people suspected to have OSAHS who are considered suitable for a home sleep study, including adults (over 16 years old) and children (2-16 years). As stated in section 2.1 above, the EAG reviews of clinical and diagnostic evidence were analysed separately for the adult and child populations, as not all included devices are indicated for use in children or young people. We followed a similar approach for the economic assessment, considering the adult and child populations separately.

Adult population (over 16 years of age)

The EAG developed a model to estimate costs and health outcomes for adults referred to specialist sleep services with symptoms suggesting the presence of OSAHS, and who are considered suitable for a home sleep study (see sections 5.5 to 5.10 below). We aimed to reflect characteristics of this population in routine NHS practice, including age, gender, and risk factors for cardiovascular events. The characteristics of our base case cohort are largely the same as in the economic model developed for NICE guideline NG202,[52] see section 5.7.1 below. We report scenario analyses with higher or lower baseline cardiovascular risk than in the base case. Due to a lack of data, we have not been able to assess cost-effectiveness separately for any of the adult subgroups highlighted in the protocol (see section 4.2.2 above). This includes the subgroup of people from black, Asian and minority ethnic backgrounds requested in the NICE scope, for whom there is a potential equality issue related to the use of light-based assessment methods.

Children and young people (2-16 years of age)

Whereas the approach to modelling the cost-effectiveness of home-based devices for diagnosing OSAHS in adults is clear (largely based on the existing cost-effectiveness model for the NICE guideline for adults, NG202)[52] the situation in children is more challenging.

A model for children would need to be structurally different to that for adults, due to differences in diagnostic and treatment pathways for children and adults. As discussed in the BTS guideline for diagnosing paediatric sleep-disordered breathing,[9] the population of children at risk of OSAHS is heterogeneous, with different diagnostic pathways for children without comorbidities (for whom adenotonsillectomy is the most common treatment) and subgroups of children with a wide range of other conditions (for whom appropriate

interventions vary according to clinical context and symptoms). There is also uncertainty over the long-term impacts of OSAHS in children, the extent to which observed associations with adverse developmental and health outcomes are related to comorbid conditions, and whether the long-term impacts are reversible with treatment for OSAHS.[9] As with adults, there is also a potential equality issue for children from black, Asian and minority ethnic backgrounds related to use of light-based assessment methods. Evidence to estimate key parameters for a children's model is currently sparse, including evidence on the diagnostic accuracy of the novel home-testing devices (section 4.3) and the impact of OSAHS on utility (section 5.2.2).

See section 5.11 below for a more detailed discussion of the current challenges to modelling the use of novel home-devices to diagnose OSAHS in children. The EAG does not consider that a decision model is likely to resolve uncertainty over the cost-effectiveness of home-based assessment with novel devices in children at this time. However, there is research in progress, notably the ongoing UK trial of AcuPebble in a paediatric population (NCT04031950 2019) [37 71 72] that is expected to improve the evidence base for children and could potentially support development of an economic model (see section 4.3). We propose a model structure and parameter sources for the cohort of children with symptoms of OSA who are under consideration for adenotonsillectomy. We also discuss the potential for modelling and data sources for children with comorbidities.

5.4.2 Intervention and comparators

Six novel home-testing devices for diagnosing OSAHS are named in the NICE scope: AcuPebble SA100, Brizzy, NightOwl, Sunrise, WatchPAT 300, WatchPAT ONE.
The EAG model for the adult population is designed to evaluate all six devices in comparison with home respiratory polygraphy (that does not include any of the named novel devices). We have also included home pulse oximetry as a comparator, should home respiratory polygraphy be limited.

5.4.3 Framework for economic analysis

The economic analysis follows the NICE reference case, as specified in the NICE process and methods manual 2022.[73]

- The model uses a 'lifetime' time horizon to reflect the consequences of misdiagnosis and sub-optimal treatment for OSAHS
- Health outcomes are estimated as QALYs, with utilities estimated from EQ-5D-3L data with NICE-recommended UK general population values. For some utility

parameters, a mapping algorithm has been used to estimate EQ-5D-3L utility from Epworth Sleepiness Scale (ESS), based on an analysis of individual patient data from a trial of CPAP (as in McDaid et al 2009 [54] and NG202 economic model [52])

- Costs are estimated from an NHS and personal social services (PSS) perspective. Unit costs are taken from standard national and NHS sources.
- Standard rates of discounting for time preference over costs and QALYs are applied, as recommended by NICE (currently 3.5% per year for costs and QALYs).

5.5 EAG approach to economic evaluation for the adult population

5.5.1 Overview

The EAG developed a health economic model to assess the cost-effectiveness of novel home-testing devices in diagnosing and assessing the severity of OSAHS for adults. This model was adapted from the existing economic model used to inform NICE guidelines on the diagnosis and management of OSA in people >16 years of age [52], as summarised in section 5.1.3 above.

We made a number of adaptations to the NG202 model to capture differences between the novel devices and conventional home sleep study diagnostics, as well as differences between the six novel devices. We have also sought to update relevant parameters from the NG202 economic model where possible, including health-related utility values for cardiovascular events, treatment costs, risks for road traffic accidents, and general mortality. Further details are provided below.

5.5.2 The diagnostic pathway: the decision tree

The decision tree is described in detail in section 5.6.1 below. It is designed to capture the short-term impacts of the diagnostic pathway by dividing the modelled cohort into 12 subgroups depending on their true OSAHS severity (no OSAHS, mild OSAHS, moderate OSAHS, severe OSAHS) and the diagnosis obtained from the modelled sleep study (no OSAHS, mild OSAHS, moderate to severe OSAHS). Moderate and severe OSAHS are combined to simplify the model, so that available published evidence on diagnostic accuracy can be used (see section 5.6.1 for further details).

The structure of the EAG decision tree is based on that used in the NG202 economic model and reflects the accuracy of the different types of sleep studies. We adapted the NG202 decision tree to account for potential differences between the novel and comparator home sleep studies in failure rates and times to treatment. An impact on time to treatment may

arise from the assumption that individuals with moderate to severe OSA who are diagnosed as having no OSA, would have a second sleep study due to on-going symptoms. To address the potential impact of treatment delay, our decision tree has a time horizon of 12 months, with an assumption that all people who are offered treatment would start treatment within this period. As some treatments are associated with utility improvements, diagnostic strategies resulting in longer delays would accrue fewer QALYs in the first year. The treatments used in the model are conservative management (e.g. lifestyle advice), mandibular advancement devices (MAD) and continuous positive airway pressure (CPAP). The costs and QALYs associated with the diagnostic sleep studies and any delays in treatment are captured.

Inputs to the decision tree are:

- The prevalence of OSAHS in our cohort of adults with suspected OSAHS, stratified by level of severity (see section 5.7.2). In the base case analysis, prevalence estimates are based on those used in the NG202 model, which are similar to the mid-range of estimates in our included diagnostic accuracy studies.
- The sensitivity and specificity of home-based sleep studies with the six named novel devices (section 5.7.3) and two comparators (section 5.7.4). Sensitivity and specificity estimates are required at two diagnostic cut-offs to divide the cohort by OSAHS severity: any OSAHS (AHI ≥ 5), and moderate to severe OSAHS (AHI ≥ 15).
- Failure rates and times to diagnosis and treatment associated with the home-based sleep studies using novel devices and comparators (sections 5.7.5 and 5.7.6).
- Treatment options depending on the diagnosed severity of OSAHS (section 5.7.11).
- Utilities associated with having OSAHS and with treatment for OSAHS. The impact of treatment on utility is determined by the appropriateness of the treatment, given the true severity of OSAHS (section 5.7.10).
- Costs associated with the novel device and comparator home-based sleep studies, including costs of the first study and any additional studies required (sections 5.7.12 and 5.7.13). In addition to the cost of the devices and any consumables, estimates are required for NHS staff costs to prepare the devices, train patients in their use, review the data, and to get the device to (and from) the patient’s home
- Costs of different treatments (see section 5.7.15).

5.5.3 Long-term consequences: the Markov model

The structure of the EAG Markov model follows that in the NG202 economic model, which in turn is essentially the same as that used in TA139 for the evaluation of CPAP for the treatment of OSAHS.[54]

It is driven by the impacts of treatment (or a lack of treatment) on the utility associated with OSAHS symptoms, and the risk of cardiovascular events and RTAs for people with OSAHS. For individuals with no OSAHS, the Markov model effectively has two states: alive and dead. For those with OSAHS, the Markov model has 12 states: alive, dead, five acute cardiovascular event states and five post-cardiovascular event states. The cardiovascular events modelled are stroke, myocardial infarction (MI), transient ischaemic attack (TIA), unstable angina and stable angina. A simplifying assumption is that individuals only experience one type of cardiovascular event. From any of the alive states, individuals are also at risk of a slight, serious or fatal road traffic accident (RTA). The model uses a 12 month cycle, and a lifetime horizon.

Parameters for the Markov model include:

- The risk of cardiovascular events, and probability of the five specific events (stroke, MI, TIA, unstable angina, stable angina), section 5.7.7.
- The risk of a slight, serious or fatal RTA (section 5.7.8).
- The impact of treatment on i) utility, ii) the risks of cardiovascular events and iii) the risk of RTAs. The impact of treatment depends on the treatment received and the true severity of OSAHS (section 5.7.11).
- The risk of death from a cardiovascular event, and the risk of death in the years following a cardiovascular event (section 5.7.9).
- Loss of utility associated with mild, moderate or severe OSAHS (section 5.7.10).
- Utility impacts of treatment, which is dependent on the type of treatment and the severity of OSAHS (section 5.7.11)
- Costs associated with treatment, cardiovascular events and RTAs (sections 5.7.15, 5.7.16 and 5.7.17).

For the NG202 model, many of the parameter estimates the same as in the TA139 economic model. [54] We therefore sought to update many of these parameters. Further details on the Markov model are in section 5.6.2.

5.6 Model structure

The model comprises a decision tree which divides the cohort into 12 subgroups depending on the true severity of OSAHS and the diagnosed severity as determined by the device used in the home-based sleep study, and a Markov model. The two components are linked by decisions on the management of individuals in the 12 different subgroups. See section 5.7 below for the full set of model input parameters and section 5.8 for a list of model assumptions.

5.6.1 Decision tree

The NG202 model did not assume a time frame for the decision tree. Instead, the diagnostic sleep studies and start of treatment were assumed to occur at the point the Markov model started. One adaptation we have made to the NG202 model is to assume that the decision tree covers a time frame of 12 months from referral for a home sleep study. We have done this to capture any differences there may be between the modelled diagnostic pathways in the time to treatment (for patients offered treatment). Thus, diagnostic pathways that require additional sleep studies, due to misdiagnosis or failure of the first study, take longer to obtain a diagnosis, and hence to start treatment. The model can capture the costs of additional sleep studies and loss of utility due to delayed treatment.

The structure of the decision tree is illustrated in Figure 3 below. The first node in the decision tree uses estimates of the prevalence of OSA in our defined population to categorise the cohort depending on whether individuals truly have OSA (AHI ≥5). Those with OSA are then divided by severity: with mild OSA defined by 5 ≤ AHI < 15, moderate OSA by 15 ≤ AHI <30 and severe OSA by AHI ≥ 30.

In addition to estimates of the prevalence and severity of OSAHS, the decision model uses estimates of the sensitivity and specificity of each sleep study at two diagnostic cut-offs to estimate the proportion of the cohort in the 12 subgroups. At the AHI = 5 cut-off, individuals with AHI ≥ 5 are diagnosed as having OSAHS, and those with AHI <5 deemed to not have OSAHS. For the second cut-off, individuals with AHI ≥ 15 are diagnosed as having moderate-severe OSAHS.

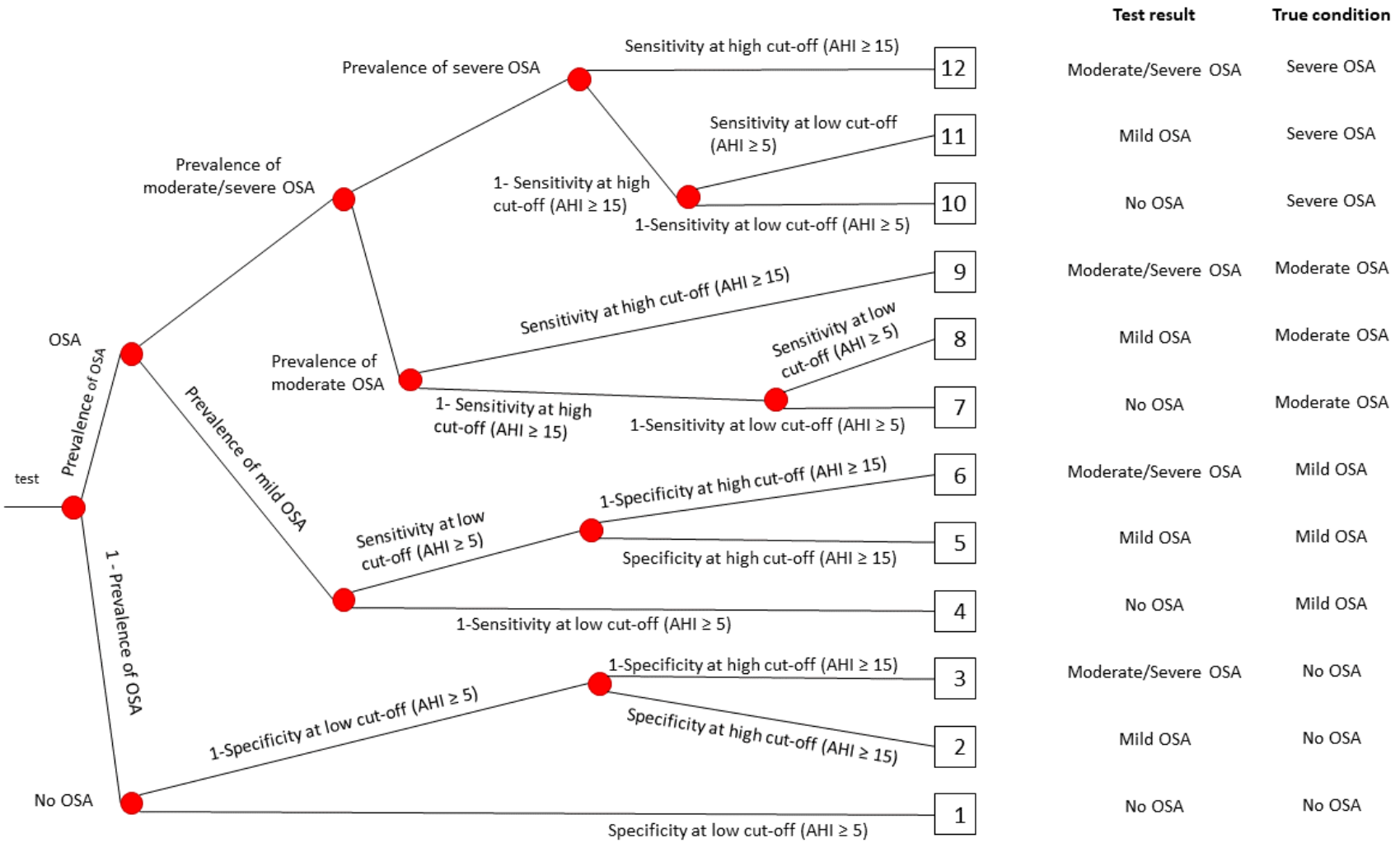
To estimate the proportion of the cohort who are correctly diagnosed as having moderate-severe OSAHS, the sensitivity of the sleep study at the AHI = 15 cut-off is used (see boxes 9 and 12 in Figure 3). In those with moderate-severe OSAHS who are misdiagnosed, it is assumed that applying the sensitivity estimates at the AHI =5 cut-off estimates the

proportion misdiagnosed with mild OSAHS (boxes 8 and 11 in Figure 3). The remainder are assumed to be misdiagnosed as having no OSAHS (boxes 7 and 10 in Figure 3).

In the proportion of the cohort who truly have mild OSAHS, the sensitivity at the AHI = 5 cut-off is initially applied to estimate the proportion diagnosed with no OSAHS (i.e. false negatives, box 4 in Figure 3), and those who truly have OSAHS. To estimate the proportion who are truly diagnosed with mild OSAHS, the estimate of specificity at the AHI = 15 cut-off is applied (box 5 in Figure 3). The remaining proportion is assumed to reflect those who truly have mild OSAHS who are incorrectly diagnosed with moderate-severe OSAHS (box 6 in Figure 3).

To determine the proportion of patients who are true negatives, i.e. are correctly diagnosed as not having OSAHS, the specificity estimate at the AHI = 5 cut-off is applied (box 1 in Figure 3). To calculate the proportion of patients who are incorrectly diagnosed as having mild OSAHS when they do not have OSAHS, the estimate of specificity at AHI = 15 is applied (box 2 in Figure 3). The remaining proportion are assumed to be diagnosed with moderate-severe OSAHS (box 3 in Figure 3).

This approach to subdividing the cohort based on the sensitivities and specificities at low and high thresholds is a simplification based on the summary results available from the diagnostic accuracy studies. Ideally, we would want to model using 4 x 4 contingency tables, which would show all three levels of severity of OSAHS and no OSHAS, and how individuals have been misdiagnosed by the sleep studies. However, as we do not have such data for all novel devices and comparators, all analyses presented in the base case use the simplification noted above. In scenario analysis, we reparameterise the decision trees using data from the 4x4 contingency tables for respiratory polygraphy, AcuPebble, NightOwl, WatchPAT 300 and WatchPAT ONE.



Test result	True condition
Moderate/Severe OSA	Severe OSA
Mild OSA	Severe OSA
No OSA	Severe OSA
Moderate/Severe OSA	Moderate OSA
Mild OSA	Moderate OSA
No OSA	Moderate OSA
Moderate/Severe OSA	Mild OSA
Mild OSA	Mild OSA
No OSA	Mild OSA
Moderate/Severe OSA	No OSA
Mild OSA	No OSA
No OSA	No OSA

Figure 3 Illustration of the decision tree

Source: Adapted from NG202 Economic Report[52], Figure 1

All patients are modelled to receive an initial diagnostic sleep study. It is assumed that there are two reasons why an individual may require a second sleep study:

1. The initial diagnostic sleep study failed to provide a diagnosis. This could be due to failure of the sleep study equipment, or insufficient data collection due to the patient's lack of sleep, for example. We assume that in this situation, a repeat sleep study would take place rapidly, with no substantive delay in diagnosis.
2. That patients who are misdiagnosed by the sleep study as having "no OSA", but truly have moderate or severe OSA (i.e. false negatives, boxes 7 and 10), will continue to be symptomatic, and so will have a second sleep study. Here we assume a delay of one month in time to diagnosis.

Although point 2 above was modelled in the original NG202 model, there was no accounting for sleep study failure rates in NG202. We have added failure rates to the EAG version of this model.

In both of the cases above, the second sleep study is assumed to be of the same type as the first (i.e. if the initial sleep study is with AcuPebble, the second sleep study will be with AcuPebble), except when the first sleep study is conducted using pulse oximetry. The second sleep study would then be the comparator respiratory polygraphy. In the base case analysis, it is assumed that no more than two sleep studies would be undertaken in these situations. For the case where the first sleep study does not provide data to make a diagnosis, this is informed by expert opinion noting that should a sleep study fail, it is very unlikely that a second sleep study would fail, especially when members of staff discuss the failure with the patients. Moreover, for both cases, experts highlighted the current limited options for providing in-hospital sleep studies in England, which would be the alternative type of sleep study. In a scenario analysis, we assume that should a second home-based sleep study continue to misdiagnose an individual with no OSA when they truly have moderate-severe OSA, they would be invited for an in-hospital PSG. We also conduct a scenario analysis assuming that a proportion of patients with mild OSA who are misdiagnosed as having "no OSA" (box 4) will have a second sleep study.

Estimates of failure rates for each type of sleep study drive the proportion of patients within each diagnostic pathway who are modelled to have a second sleep study for reason 1 above. See section 5.7.5 below for estimates of sleep study failure rates. Estimates of sensitivity for each type of sleep study drive the proportion of patients within each diagnostic

pathway who are modelled to have moderate-severe OSA misdiagnosed as no OSA (i.e. reason 2 above).

As highlighted above, we have adapted the NG202 model so that the decision tree covers a time period of one year, which is equal to one Markov cycle. It is assumed that, within this 12 month period, all individuals within the cohort would have had their sleep study, or studies, and for those offered treatment, they would have commenced treatment.

As part of the adaptations made by the EAG, we assume a time to diagnosis for each type of sleep study, and a time to treatment (for those having treatment) which does not depend on the type of sleep study modelled. The time taken to undertake a second sleep study depends on the reason (i.e. a failure or misdiagnosis), and the type of device. See section 5.7.6 for more detail. The model captures any costs associated with additional sleep studies due to failures and misdiagnosis, as well as the impacts of these additional sleep studies in terms of lost utility. This is achieved as it is assumed that treatment has a direct impact on utility (see section 5.7.11). Individuals starting appropriate treatment earlier will gain more benefits than those starting treatment later. The costs of any treatment within the 12 months are also captured in the decision tree.

5.6.2 Markov model

At the end of the decision tree, patients enter one of 16 Markov models according to their underlying severity of OSAHS, their diagnosed severity and the treatment they are receiving. This is presented in Table 15 below, where the subgroup number corresponds to the numbered boxes in the decision tree (Figure 3). The Markov model has yearly cycles and incorporates a half-cycle correction.

Patients who do not have OSAHS (subgroups 1-3) are assumed to have standard population mortality rates; consequently, they are simulated in Markov models using national lifetables for England and Wales. Thus, the Markov models for these cohorts contain only two health states, alive and dead.

Individuals who truly have OSAHS, are assumed to be at risk of non-fatal and fatal cardiovascular events and road traffic accidents, as well as other cause mortality. The main purpose of the Markov model is to estimate the impacts of treatment decisions (informed by the diagnostic pathway) on the utility, and risk of cardiovascular events and RTAs, alongside related costs, for people with OSAHS. People with OSAHS who are correctly diagnosed and

treated will benefit from an improvement in utility, a potential reduction in cardiovascular event risk (depending on their severity and the treatment they receive), and a potential reduction in the risk of RTAs (again, depending on their severity and the treatment they receive). Further details on the assumed impacts of treatment are given in section 5.7.11.

Table 15 Treatment options depending on diagnosed severity of OSAHS

Subgroup	True severity	Diagnosis	Treatment
1	No OSAHS (AHI <5)	No OSAHS	No treatment
2		Mild OSAHS	Conservative treatment Conservative management + CPAP Conservative management + MAD
3		Moderate/ severe OSAHS	Conservative management + CPAP Conservative management + MAD
4	Mild OSAHS (5 ≤ AHI < 15)	No OSAHS	No treatment
5		Mild OSAHS	Conservative treatment Conservative management + CPAP Conservative management + MAD
		Moderate/ severe OSAHS	Conservative management + CPAP Conservative management + MAD
7	Moderate OSAHS (15 ≤ AHI < 30)	No OSAHS	No treatment
8		Mild OSAHS	Conservative treatment Conservative management + CPAP Conservative management + MAD
9		Moderate/ severe OSAHS	Conservative management + CPAP Conservative management + MAD
10	Severe OSAHS (AHI ≥ 30)	No OSAHS	No treatment
11		Mild OSAHS	Conservative treatment Conservative management + CPAP Conservative management + MAD
12		Moderate/ severe OSAHS	Conservative management + CPAP Conservative management + MAD

Abbreviations: AHI Apnoea Hypopnoea Index; CPAP continuous positive airway pressure; MAD mandibular advancement device
Source: Adapted from NG202 Economic Report [52], Table 4

The subgroups in Table 15 for individuals who do have OSA (subgroups 4-12) are simulated in the Markov model to be at risk of cardiovascular events, road traffic accidents (RTAs), and death. An illustration of the Markov model is given in Figure 4. Twelve distinct health states are used to model the long-term consequences in OSA. Individuals enter the Markov model in the 'OSAHS' state. They are then at risk of an acute cardiovascular event, which could be stable angina, unstable angina, myocardial infarction (MI), transient ischemia attack (TIA) or stroke. A risk of death is associated with each cardiovascular event, and those with a fatal event transition to the Dead state. Those surviving an acute cardiovascular event leave the acute event state after one model cycle and enter the related post event state. Individuals are assumed to stay in the post cardiovascular event state until they die and transition to the

Dead state. The model assumes that once an individual has a cardiovascular event, they cannot have a second event.

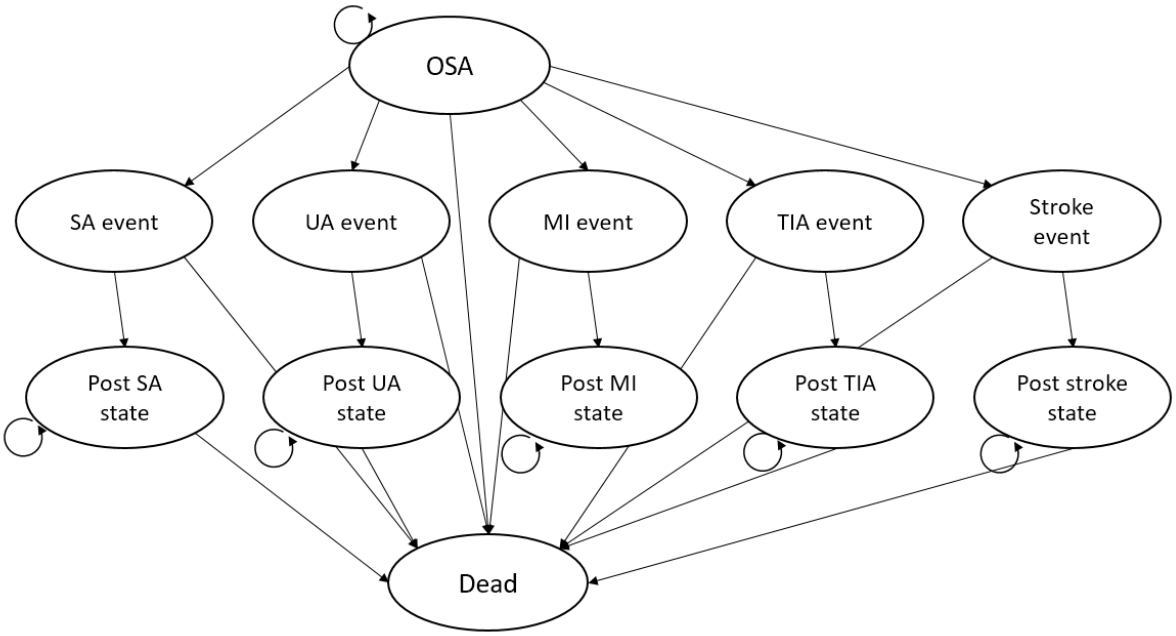


Figure 4 Illustration of the Markov model

Abbreviations: SA = Stable Angina; UA = Unstable Angina; MI = Myocardial Infarction; TIA = Transient Ischemic Attack

Source: Adapted from NG202 Economic analysis report, Figure 2 [52]

The health state 'OSAHS' is the start of the model. Where there are arrows between different states, transition probabilities are assumed between these states at every cycle. Thus, from the OSAHS state transitions can be made to any of the acute cardiovascular event states or the dead state. States in which individuals can remain for more than one model cycle are indicated by arrows circling back to the same state. Note that the five acute cardiovascular event states do not have arrows circling back, as it is assumed individuals only remain in these states for one model cycle, i.e. one year. Dead is an absorbing state, from which individuals cannot leave.

The risk of a RTA is assumed throughout the model, so can occur within any alive health states, but is not shown in Figure 4. Individuals can have a slight or serious RTA, which is associated with relevant costs and utility impacts, but does not impact on any transition probabilities. In the case of a fatal RTA, and other cause mortality, individuals transition to the Dead state. The transition probabilities for the different cardiovascular events depend on the assumed characteristics of the cohort (see section 5.7.1), the model cycle (i.e. age), whether individuals are treated, and the effectiveness of, and adherence to, treatment. The

probabilities for a RTA depend on the model cycle (i.e. age), sex, whether individuals are treated, and the effectiveness of, and adherence to, treatment (See section 5.7.11).

The only adaptation the EAG has made to the structure of the NG202 Markov model is to allow for the modelled cohort to be a mixed population of males and females. In the NG202 model, analyses could only be obtained for a cohort of males or females. To allow for a mixed population, we have added sex-specific risks and utilities to the Markov model.

5.7 Model parameters

The values of all model parameters used in the EAG base case, probabilistic and one-way sensitivity analysis are shown in Appendix 9b Table 78.

5.7.1 Cohort characteristics

In the NG202 economic model, the cohort characteristics were based on those used in a previous NICE technology appraisal.[54] The mean age of individuals in the diagnostic accuracy studies reported in section 3 above is approximately 50 years. This is the age that was assumed in NG202, and so we also assume the cohort is aged 50 when they enter the model. Very few other population characteristics were available from the diagnostic accuracy studies (see section 4.2 above), so we use baseline characteristics adapted from the NG202 economic model for our base case analysis (Table 16). We assume that 70% of the cohort are male. This is based on the proportions reported in the diagnostic accuracy studies informing this model: 57% [25], 55% [74], 71% [19] and "predominantly male" [35].

Table 16 Model cohort characteristics

Cohort characteristic	Base case cohort		Low risk cohort		High risk cohort	
	Not treated with CPAP	Treated with CPAP	Not treated with CPAP	Treated with CPAP	Not treated with CPAP	Treated with CPAP
Age	50 years	50 years	50 years	50 years	50 years	50 years
% male	70%	70%	30%	30%	70%	70%
Smoking status	Non-smoker	Non-smoker	Non-smoker	Non-smoker	Heavy smoker	Heavy smoker
Diabetes	Type 2	Type 2	None	None	Type 2	Type 2
Cholesterol ratio	5.2	5.2	5.2	5.2	5.2	5.2
Systolic blood pressure	130	128	130	128	130	128
Chronic kidney disease	No	No	No	No	Yes	Yes

Abbreviation: CPAP continuous positive airway pressure
Source: Adapted from NG202 Economic analysis report, Table 6 [52]

In scenario analyses, we assess the impact of changing the assumptions on the cohort characteristics to reflect i) a cohort assumed to be at a lower risk of cardiovascular events, and ii) a cohort assumed to be at a higher risk of cardiovascular events compared to that assumed in the base case analysis. These scenarios are based on assumptions about baseline cardiovascular risk factors taken from the TA139 assessment[54] for base case, low risk and high risk cohorts, with and without CPAP (NG202 Economic report [52] Table 6).

The limited information on patient characteristics and the lack of subgroup analysis from clinical and diagnostic studies (section 4.2) has prevented estimation of the cost effectiveness of novel devices for the subgroups of interest noted in the NICE scope (e.g. People with COPD / neuromuscular disorders / People from black, Asian and minority ethnic backgrounds and pregnant women and pregnant people).

5.7.2 Baseline prevalence

The model requires an estimate of the prevalence of OSAHS within the population of adults suspected of having OSAHS, and the distribution of mild, moderate and severe disease in those with OSAHS. In the NG202 economic model, these estimates were obtained from a meta-analysis of prevalence reported in the diagnostic studies included in their review (NG202 Economic report [52] Table 7). They estimated an overall OSAHS prevalence (AHI ≥ 5) of 82% in the diagnostic studies. Of those with OSAHS, 32% were in the mild category (5 ≥ AHI < 15) and 68% in the moderate to severe category (AHI ≥ 15). Within the moderate to severe category, 60% were estimated to have severe disease (AHI ≥ 30). These prevalence estimates are similar to those in the diagnostic studies included in our review (section 4.2.2). We therefore use the prevalence estimates as used in NG202 but assess the impact of changing these estimates in sensitivity analyses.

5.7.3 Diagnostic accuracy of novel devices

Data informing the diagnostic accuracy of novel home testing devices are taken from the clinical effectiveness review (section 4). To inform our evaluation, the ideal accuracy study would:

- have a comparative accuracy design, allowing for all novel devices and comparators to be evaluated together in a model,
- be used in the patient's home to reflect the intended setting, and
- use a gold standard diagnostic test that perfectly discriminates between the presence and severity of OSAHS.

The results of the clinical effectiveness review show that many of the included studies are not close to our ideal study design.

- We do not have any studies where one novel device is compared to another novel device.
- Only two devices have been evaluated for diagnostic accuracy in the home-setting (Devani 2021[19] for AcuPebble, Kelly 2022 [27] for Sunrise). A second study of AcuPebble,[20], and a second study of Sunrise[26], plus studies for Brizzy[21] and NightOwl[25] have evidence for their accuracy in a clinic setting. No diagnostic accuracy evidence was identified from the systematic review for WatchPAT 300 and WatchPAT ONE.
- There is no perfect test for the diagnosis of OSAHS. Laboratory-based PSG has been used as a reference standard in many diagnostic accuracy studies for home RP,[75] [43] and is widely regarded as being the gold standard test for OSAHS. However, it has inherent limitations which can impair the patient's sleep quality (e.g. discomfort from being attached to monitoring equipment).

Of the studies providing evidence on the diagnostic test accuracy of current versions of the novel devices, four used laboratory-based PSG as the reference standard (Sanchez Gomez (2024)[20] for AcuPebble, Pepin et al (2020)[26] for Sunrise, Lyne et al (2023)[25] for NightOwl and Martinot et al (2017)[21] for Brizzy), one used home-based PSG (Kelly et al (2022)[27] for Sunrise), and one used home-based RP (Devani et al (2021)[19] for AcuPebble). From previous systematic reviews, we know that home-based RP is not equivalent in accuracy to PSG [43] [75]. Therefore, to assume equivalence between home RP and PSG, and take the estimates of sensitivity and specificity from the Devani et al (2021) study for AcuPebble, would over-estimate AcuPebble's diagnostic accuracy. We therefore use accuracy data from Sanchez Gomez (2024)[20] for AcuPebble, even though this study is based in a sleep clinic and is smaller than that reported by Devani et al (2021).

Due to the lack of evidence on the diagnostic accuracy of WatchPAT 300 and WatchPAT ONE we examined the available evidence for predecessor versions of this novel device. Our systematic review identified two studies reporting the diagnostic accuracy WatchPAT 200-U: Tauman et al 2020 [30] and Pillar et al 2020 [31]. Both studies evaluate WatchPAT 200U in the sleep laboratory, simultaneously with PSG. The study by Pillar et al focusses on the ability of the device to distinguish between OSA and central sleep apnoea, while Tauman et al focus on the use of the device to diagnose OSA in people with a previous diagnosis of atrial fibrillation. We use the diagnostic accuracy data for WatchPAT 200U reported in Tauman et al to inform the modelling of WatchPAT 300 and WatchPAT ONE.

Diagnostic accuracy in the home-setting

Devani et al (2021)[19] report the diagnostic accuracy of AcuPebble used in the home, with home RP, using the Embletta device, as the reference standard. This is different to the reference standard used for other novel devices, which makes it difficult to compare diagnostic accuracy results for AcuPebble with those for other novel devices. We therefore use accuracy data from Sanchez Gomez (2024)[20] which was conducted in the clinic. In scenario analyses, we use the accuracy data as reported in Devani for AcuPebble.

Of the two studies included in the systematic review which report on the diagnostic accuracy of Sunrise, our base case analysis uses data from the Pepin et al 2020 [26] study, done in the sleep laboratory setting. Although the other study (Kelly et al 2022)[27] is based in the home setting, it is very small (n=31 participants) compared to the Pepin et al 2020 study (n=376 participants) and uses home PSG as the reference standard. We note that both studies use post-hoc optimisation of thresholds, which is likely to overestimate accuracy of the devices, as highlighted in the clinical effectiveness risk of bias assessment (see Section 4.4). We use data from Kelly et al 2022 [27] in a scenario analysis, which relies on an assumption that home PSG is equivalent to laboratory PSG. The estimates of sensitivity and specificity at the required diagnostic cut-offs for the economic model are reported in Kelly 2022 but they did not report confidence intervals (see Table 17 for EAG confidence interval estimates).

Table 17 Sensitivity and specificity estimates used in the model for home studies

Device	Reference standard	Cut-off used in model	Sensitivity (95% CI)	Specificity (95% CI)	Source
AcuPebble [a,b]	Home RP	AHI ≥ 5	0.92 (0.84, 0.96)	0.96 (0.88, 1.00)	Devani et al 2021[19]
		AHI ≥ 15	0.93 (0.82, 0.98)	0.97 (0.91, 0.99)	
Sunrise [b]	Home PSG	AHI ≥ 5[c]	0.88 (0.69, 0.97)	1.00 (0.54, 1.00)	Kelly et al 2022 [27] 95% CIs calculated by EAG
		AHI ≥ 15[d]	1.00 (0.79, 1.00)	0.75 (0.45, 0.92)	

[a] using accuracy estimates based on AHI cut-off with desaturation ≥ 3%; [b] only used in scenario analysis; [c] using accuracy estimates based on MM-ORDI 9.53 cut-off; [d] using accuracy estimates based on MM-ORDI 12.65 cut-off;
Abbreviations: AHI Apnoea Hypopnoea Index; CI confidence interval; MM-ORDI mandibular movements obstructive respiratory disturbance index; PSG polysomnography; RP respiratory polygraphy

Diagnostic accuracy in the clinical setting

Studies evaluating the disposable version of NightOwl, AcuPebble SA100, Brizzy, Sunrise and WatchPAT 200U were conducted in a clinical setting with laboratory PSG as the reference standard. Studies either reported sensitivity and specificity estimates of the novel devices at the required diagnostic cut-offs (AcuPebble, Brizzy, Sunrise), reported the 4x4 contingency table data for diagnostic accuracy (AcuPebble, NightOwl), or were estimated by the EAG from a scatter plot of the index test and reference standard scores (WatchPAT 200U). As the 4x4 contingency table data for the accuracy of AcuPebble, NightOwl and WatchPAT 200U were available, they were modelled in a scenario analysis to demonstrate the impact of the simplified decision tree model (as noted in section 5.6.1 above). The sensitivity and specificity estimates at the required diagnostic cut-offs are shown in Table 18.

For NightOwl, two additional studies based in the clinic reported on the reusable version of NightOwl (Massie et al 2018 and van Pee 2020), as opposed to the disposable version to be launched in the UK (and used in Lyne et al 2023). Since the company confirmed to NICE that the only difference between the reusable and disposable NightOwl devices is whether the battery can be re-charged, and that the sensors and software are identical, we conduct scenario analyses using data from Massie et al 2018 and van Pee et al 2020.

For WatchPAT 200U, only sensitivity and specificity estimates at the AHI ≥ 15 diagnostic cut-off are presented in Tauman et al. However, the authors report a scatterplot of AHI values from WatchPAT 200U compared to laboratory PSG (Figure 1 in Tauman et al [30]), from which we were able to extract data points (using Engauge Digitiser) to create the 4x4 contingency table. Using these data we were able to replicate the sensitivity and specificity estimates at the AHI ≥ 15 diagnostic cut-off, as reported in Tauman et al. However, there is some uncertainty to this approach due to the fact that Tauman et al report that there were 101 participants in the study and we could only extract 100 data-points. Because of this, the sensitivity and specificity estimates at the diagnostic cut-off of AHI ≥ 5 may not be accurate.

There are some uncertainties in the use of data from Tauman et al 2020 to inform the model for WatchPAT 300 and WatchPAT ONE. In particular, we are making the following assumptions:

- The accuracy of the earlier version (WatchPAT 200U) is the same as that for WatchPAT 300 and WatchPAT ONE, which the company has confirmed
- The accuracy of WatchPAT 300 is the same as the disposable WatchPAT ONE device, which the company has confirmed

- The population in Tauman et al of people with a previous diagnosis of atrial fibrillation is generalisable to the general population of people suspected of having OSAHS (as defined in the scope for this assessment).

Table 18 Sensitivity and specificity estimates used in the model for clinic studies

Device	Cut-off used in model	Sensitivity (95% CI)	Specificity (95% CI)	Source
AcuPebble	AHI ≥ 5 [a]	[redacted]	[redacted]	Acurable (2023) [15] Sanchez Gomez (2024) [20]
	AHI ≥ 15 [a]	0.93 (0.77, 0.99)	0.97 (0.85, 1.00)	
Brizzy	AHI ≥ 5 [b]	0.93 (0.87, 0.97)	1.00 (0.51, 1.00)	Martinot et al 2017
	AHI ≥ 15 [c]	0.89 (0.80, 0.94)	1.00 (0.83, 1.00)	
NightOwl	AHI ≥ 5	0.93 (0.85, 0.98)	0.77 (0.55, 0.92)	Lyne et al 2023
	AHI ≥ 15	0.89 (0.76, 0.96)	0.82 (0.68, 0.91)	
NightOwl	AHI ≥ 5 [d]	0.98	0.80	Massie et al 2018, 95% CIs not reported
	AHI ≥ 15 [e]	0.97	0.83	
NightOwl	AHI ≥ 5 [a]	0.93 (0.89, 0.97)	0.72 (0.54, 0.91)	Van Pee et al 2020
	AHI ≥ 15 [a]	0.91 (0.85, 0.96)	0.76 (0.65, 0.87)	
Sunrise	AHI ≥ 5	0.91 (0.89, 0.92)	0.94 (0.91, 0.97)	Pepin et al 2020
	AHI ≥ 15	0.92 (0.90, 0.94)	0.84 (0.81, 0.87)	
WatchPAT 300	AHI ≥ 5 [a]	0.96 (0.90, 1.00)	0.25 (0.01, 0.81)	Calculated by EAG from Figure 1 of Tauman et al 2020[f]
	AHI ≥ 15 [a]	0.88 (0.79, 0.94)	0.63 (0.38, 0.84)	Tauman et al 2020[f]

WatchPAT ONE	AHI ≥ 5 [a]	0.96 (0.90, 1.00)	0.25 (0.01, 0.81)	Calculated by EAG from Figure 1 of Tauman et al 2020[f]
	AHI ≥ 15 [a]	0.88 (0.79, 0.94)	0.63 (0.38, 0.84)	Tauman et al 2020[f]

[a] Using accuracy estimates based on 3% desaturation rate; [b] using accuracy estimates based on MM-RDI 5.9 cut-off; [c] Using accuracy estimates based on MM-RDI 13.5 cut-off; [d] Using accuracy estimates based on REI 5 cut-off; [e] Using accuracy estimates based on REI 15 cut-off; [f] WatchPAT 200U

Abbreviations: AHI Apnoea Hypopnoea Index; CI confidence interval; MM-RDI mandibular movements respiratory disturbance index; REI respiratory event index

Finally, to consider the uncertainties and limitations in the diagnostic accuracy estimates for the novel devices as discussed above and earlier in section 4 we conduct a 'worst case' scenario analysis using the lower-bound of the 95% CIs for accuracy estimates in the model.

5.7.4 Diagnostic accuracy of home respiratory polygraphy and oximetry

To inform the accuracy of home RP, the NG202 economic model used pooled accuracy estimates from a systematic review of eight studies evaluating home respiratory polygraphy with at least four channels.[75] The eight included studies were published between 2002 and 2017, used in-hospital or in-laboratory PSG as the reference standard, and evaluated eight different devices: Somnocheck, WatchPat 200, Embletta, ApnoeaScreen-I, Breas SC20, MediByte, StarDust-II and Nox-T3. The number of participants in the studies ranged from 43 (StarDust-II) to 366 (Breas SC20) and the studies were conducted in Canada (n=2), Brazil (n=2), Spain (n=2), China (n=1) and the USA (n=1). In their evaluation of the quality of the overall evidence, the NG202 evidence review authors rated the pooled estimates as being of very low quality "due to risk of bias, serious inconsistency, and serious imprecision". It is reported in the NG202 evidence review that committee members noted that technology will have improved over time, but that there was no clear time-point for when evidence could be considered inappropriate, thus no restrictions on year of study were placed in the NG202 evidence review.

We sought more recent evidence on the diagnostic accuracy of home RP. We consulted the comprehensive systematic review by Khor et al (2023) [43] but could not identify more recent studies fully meeting our criteria: home-setting, reporting sensitivities and specificities at the diagnostic cut-offs of AHI ≥ 5 and AHI ≥ 15. Thus, we took a pragmatic approach, discussing the advantages and limitations of studies from Khor's review to reach a consensus the best

available candidate study to inform our estimates of the sensitivity and specificity of home RP.

In our base case analysis, we use data from Xu et al 2017.[44] This study was chosen as it evaluates RP in the home-setting, is compared to in-laboratory PSG, uses a named device (Nox-T3, that one of our experts noted as being representative of what is currently used in England), and is one of the most recent studies that we considered. The latter two points go some way to addressing the comments raised in the NG202 economic analysis report,[52] that technology has improved over time, with recent studies more likely to reflect current practice than older studies. Xu et al 2017 [44] was conducted in China and included 80 adults (18-80 years old) who had been referred for OSA investigation. Only those who had not had previous sleep studies, or been treated, for OSA were included. Participants were excluded if they had no access to a telephone, no ability to return to the clinic for follow-up, prior diagnoses of CSA, obesity hypoventilation syndrome, narcolepsy, and COPD among others, people doing shift work or regular jet lag, and a "clinically unstable medical condition". Before the home sleep study, participants were given instruction and demonstration in the clinic on how to apply the sensors and perform the recording. Within a week after the home Nox-T3 sleep study, participants had an in-laboratory PSG and simultaneous Nox-T3 sleep study. One person declined to do the home sleep study, so was excluded, leaving 79 participants. The authors state that home RP was conducted first to reflect the situation where participants had no previous experience of using any sleep study kits, and to assess the ability of participants to successfully complete the sleep study at home. The accuracy results for Nox-T3 are based on initial automatic scoring followed by manual editing "by an experienced PSG technologist" who was blind to whether the Nox-T3 study was at home or in the clinic, and also blind to results from the PSG. Hypopnoea was based on a ≥ 4% desaturation rate. Individuals who had an unsuccessful first sleep study at home were given the home RP kit, after their PSG visit, to repeat the home RP on a subsequent night. Five individuals had an unsuccessful initial home sleep study, all were due to the loss or absence of the oximetry signal. Three of these did a second study which was successful, leading to 77 participants having diagnostic data from home RP and laboratory PSG. Four of the 80 in-laboratory RP sleep studies were unsuccessful due to an absence of, or inability to interpret, the oximetry signal.

We considered a second potential candidate study, Pereira et al (2013),[76] which was included in the systematic review for NG202 and the Khor et al (2023)[43] systematic review. The study included 128 consecutive patients from a single clinic in Canada, who had been referred for assessment of OSA. People with known COPD, congestive heart failure or

uncontrolled asthma were excluded. The home sleep study (using MediByte) was conducted on two consecutive nights, with results from the first night used, unless it was unsuccessful, and then the second night sleep study results were used. Prior to the sleep study, individuals received demonstration on the use of the device. Results from the home sleep study were compared to those from an in-laboratory PSG sleep study. It is reported that the PSG occurred after the home sleep study, but there are no details on how long after the home study this occurred. The home sleep study and PSG were manually scored, with the scorer being blind to the results of the other type of sleep study. Hypopnoea was based on a ≥ 3% desaturation rate. Thirteen individuals did not have sufficient recordings from the first night of the home sleep study, and so their second night results were used.

Although the data from Xu et al 2017 and Pereira et al 2013 that are used in the economic model are from use of RP in the home setting, the reference standard of PSG was conducted on a different night in a laboratory, thus both studies are limited by the lack of simultaneous measurement. A further limitation is that the RP signal scoring is based on estimated recording time whereas PSG signal scoring is based on directly observed total sleep time. It has been found that the former is associated with lower AHI scores and under-diagnosis of OSA. We conducted scenario analyses replacing sensitivity and specificity estimates from Xu et al with those from Pereira et al 2013 and those used in the NG202 economic analysis.[52] Table 19 provides the sensitivity and specificity estimates for home RP from Xu et al 2017, Pereira et al 2013 and the NG202 economic analysis.[52]

In the NG202 economic model, accuracy data for home pulse oximetry were taken from four studies published between 1993 and 2010.[75]. For two of the studies, the pulse oximetry device is not named, one study evaluated Flow Wizard and the other study evaluated Biox 3740. In all studies, the reference standard was in-laboratory PSG. Two studies contributed estimates of sensitivity and specificity for diagnosing OSA in terms of ODI≥5, while data from three studies were pooled to obtain sensitivity and specificity estimates for ODI≥15. The authors of the NG202 diagnostic evidence review reported the level of evidence for these estimates as low or very low based on QUADAS-2.

The EAG sought more recent evidence on the accuracy of home oximetry, including consideration of a recent systematic review on different sleep studies.[43] However, we could not identify any usable estimates from any of the included studies meeting either of our criteria of: the setting of the evaluation being in the home and not the clinic; and requiring sensitivity and specificity estimates at cut-offs of AHI or ODI ≥ 5, and AHI or ODI ≥15.

Therefore, we use the accuracy estimates for oximetry that were included in the NG202 economic model (see Table 19 below).

In our base case analysis, when a second test is required due to a misdiagnosis, we assume that the result of the second test is independent to the result of the first test. In scenario analyses, we explore estimates of correlation between the two tests.

Table 19 Accuracy estimates for RP and oximetry used in the model

Test and threshold	Sensitivity (95%CI)	Specificity (95%CI)	Source
Base case analysis			
Respiratory polygraphy			
AHI ≥ 5	0.95 (0.87, 0.99)	0.69 (0.39, 0.91)	Xu et al 2017[44]
AHI ≥ 15	0.93 (0.81, 0.99)	0.85 (0.69, 0.95)	
Oximetry			
ODI≥5	0.52 (0.08, 0.93)	0.96 (0.15, 1.00)	NG202 Evidence Review D[75]
ODI≥15	0.35 (0.13, 0.65)	0.99 (0.95, 1.00)	NG202 Economic Analysis Report[52]
Scenario analyses – Alternative data sources for respiratory polygraphy			
Data source: Pereira et al 2013			
AHI≥5	0.87 (0.80, 0.93)	0.67 (0.35, 0.90)	Pereira et al 2013[76] NG202 Evidence Review D[75]
AHI≥15	0.77 (0.67, 0.86)	0.95 (0.83, 0.99)	
Data source: NG202 economic analysis			
AHI≥5	0.94 (0.90, 0.97)[a]	0.58 (0.40, 0.74)[a]	NG202 Economic Analysis Report[52]
AHI≥15	0.84 (0.60, 0.96)[a]	0.89 (0.71, 0.97)[a]	

[a] The 95%CIs are not reported in the NG202 Economic Analysis Report, instead we report the 95%CI are given in the NG202 Evidence Review D report[75]
Abbreviations: AHI Apnoea Hypopnoea Index; CI confidence interval; ODI oxygen desaturation index; RP respiratory polygraphy

5.7.5 Test failure rates

As stated in section 5.6.1, the EAG model accounts for sleep studies failing to provide sufficient data to make a diagnosis, and therefore requiring a second sleep study. Only those failures that are likely to require a repeat sleep study in practice, at a cost to the NHS, are included in the model.

Estimates of failure rates for the respective novel devices and types of sleep studies are shown in Table 20 below. A number of sources were available to estimate the failure rates associated with home RP. This included published studies identified from our systematic review, 6-month data from the Newcastle Regional Sleep Service (personnel communication from James Oliver), and expert opinion. In the base case analysis, we chose the estimate from the Newcastle Regional Sleep Service as it provides real-world estimates of home RP failures in current UK practice (the data were collected between October 2022 and March 2023) and based on 1000 sleep studies (54/1000, 5.4%). We note that clinical experts consulted by the EAG quoted a likely failure rate for home respiratory polygraphy of 5%. In sensitivity and scenario analyses we explore the impact of varying this assumption, [redacted]. Failure rates for home oximetry were also taken from the Newcastle Regional Sleep Service, again due to our assumption that this reflects current practice in the UK. This estimate is based on data collected from 194 sleep studies during October 2022 and March 2023. For RP and oximetry, it is reported that failures recorded in the Newcastle Regional Sleep Service data, were due to a lack of adequate data, for instance, signals missing, recordings being too short, or being unable to gain useful information from the test to draw conclusions.

The test failure rates for Brizzy, and NightOwl are taken from the relevant diagnostic accuracy studies in the systematic review of clinical effectiveness (see section 4.5.6). The studies by Devani et al (2021) and Sanchez Gomez et al (2024) of AcuPebble both report failure rates. As the study by Devani is larger and conducted in the home setting, we use these data to inform our model. Note that only device one failure is included in the model, as all other reasons given by Devani 2021 are either related to their study protocol (and so would not occur in clinical practice, e.g. patients logging themselves out of the device) or do not incur any additional costs for a repeat sleep study (e.g. patients forgetting to do the sleep study would be able to re-do the following night at no additional cost). Failure rates for Sunrise are available from the Kelly et al (2022) study, and the Alsaif et al (2023) unpublished manuscript. The failure rate for Sunrise reported from the Alsaif manuscript on the full study dataset is [redacted]. We note the failure rate for the comparator (home RP) was [redacted] Both studies (Alsaif 2023 and Kelly 2022) were conducted in the home [redacted] Due to data from Kelly being publicly available from a peer-reviewed journal we use this data in the base case analysis. However, we conducted a scenario analysis assuming [redacted] from the Alsaif et al (2023)

unpublished manuscript (see Section 5.10.2). Failure rates for WatchPAT 300 and Watch ONE are taken from Mueller et al 2022 (see Table 20). In a scenario analysis, we assume that a proportion of the cohort may have a third sleep study with laboratory-based PSG when a second home-based sleep study continues to misdiagnose an individual with moderate-severe OSA as having no OSA. As we assume, for the purposes of modelling, that PSG is a perfect test, we estimate a failure rate of 0%. We acknowledge that this is a simplifying assumption.

Table 20 Estimates of failure rates used in the model

Device/test type	Base case value	Scenario analysis value	Source
Home respiratory polygraphy	5.4% (54/1000)	[redacted]	Base case value: Newcastle Regional Sleep service (personnel communication James Oliver) Scenario analysis value: Alsaif et al 2023
Home oximetry	4.6% (9/194)		Newcastle Regional Sleep service (personnel communication James Oliver)
AcuPebble SA100	0.6% (1/176)		Devani et al 2021 [19]
Brizzy	4.0% (4/100)		Martinot et al 2017 [21]
NightOwl	11.5% (13/113)		Lyne et al 2023 [25]
Sunrise	10.5% (4/38)	[redacted]	Base case value: Kelly et al 2022 [27] Scenario analysis value: Alsaif et al 2023
WatchPAT 300	3.28% (2/61)		Mueller 2022[29]
WatchPAT ONE	3.28% (2/61)		Mueller 2022[29]

5.7.6 Time to diagnosis and treatment

The systematic review of clinical effectiveness did not identify any studies reporting data on the time to diagnosis or time to treatment associated with use of any of the novel devices. We sought the opinion of a number of clinical experts on the likely time to treatment for individuals offered treatment. There was large variation in the responses, but some consensus that it would most likely be within 12 months of being referred for a sleep study. We therefore assume that it would take 3 months to receive a diagnosis, and a further 3

months to commence treatment, should treatment be offered. In the base case analysis, this assumption applies to all home-testing sleep studies. In one-way sensitivity analyses, we assess the impact of increasing time to diagnosis and time to treatment to 6 months, and decreasing time to diagnosis and time to treatment to 1.5 months. However, these sensitivity analyses are not based on any data. The EAG obtained a manuscript submitted for publication (Alsaif et al 2023), which was made available to us as academic in confidence by the study authors. The manuscript provides final results of the study, superseding interim results presented in a conference abstract. This manuscript reported data on time to treatment decision. We therefore undertook a scenario analysis which assumes a [redacted] using these unpublished data (Alsaif 2023), see Section 5.10.2.

As noted in the model description (see section 5.6.1), we assume that if an individual who truly has moderate or severe OSA, but is initially diagnosed with no OSA, they would have a second sleep study due to on-going symptoms. In the base case analysis, we assume that for individuals who have a second sleep study for this reason, their time to treatment would be delayed by one month. Therefore, diagnostic pathways for novel devices or comparators that are more likely to misdiagnose people with moderate to severe OSA as having no OSA, will have a delay in their time to treatment. This delay to treatment translates into a decrement in utility as appropriate treatment is assumed to improve utility.

5.7.7 Cardiovascular risk

As in the NG202 model, we assume that OSA leads to an increased risk of cardiovascular events. We take the same approach as in NG202 and use estimates from QRISK3 [78] to inform transition probabilities within the Markov model from the initial 'OSAHS' state to the acute cardiovascular event states. QRISK3 is a prediction model for 10-year cardiovascular risk. It is based on UK data from general practices, and uses risk factors including age, BMI, ethnicity, systolic blood pressure, smoking, the ratio between total cholesterol and high density lipoprotein cholesterol, and comorbidities including type I or II diabetes, atrial fibrillation and chronic kidney disease. It was derived using data from almost 8 million individuals aged 25-84 years whose general practice is part of the QResearch database. The prediction models were validated in over 2.5 million individuals, 25-84 years old.

The risks of CV events in the EAG model are estimated based on our cohort demographics and the risk factor assumptions in Table 16 (section 5.7.1 above). The raised CV risk for

people with OSAHS with and without CPAP treatment is approximated using the following assumptions:

- In the base case, men and women with OSAHS are assumed to have a CV risk the same as a non-smoker of the same age with type 2 diabetes.
- In the high-risk scenario, men and women with OSAHS are assumed to have the same CV risk as a heavy smoker of the same age, with type 2 diabetes and chronic kidney disease.
- In the low-risk scenario, men and women with OSAHS are assumed to have the same CV risk as non-smokers of the same age, who do not have diabetes, nor chronic kidney disease.
- All of the above groups are assumed to have a cholesterol ratio of 5.2, and systolic blood pressure of 130 mmHg without treatment, reduced to 128 mmHg with treatment.
- Other risk factors were left at their default values.

As in the NG202 economic model, we assume that the age-specific probability of a cardiovascular event from QRISK3 is constant over the 10 years and is a good estimate of the probability of an event in 5 years. For example, the 10 year predicted risk for a 60 year old can be transformed into an annual risk for a 65 year old. Thus, an annual rate is calculated, while adjusting for mortality in that year. It is then converted into an annual probability of a cardiovascular event to be applied 5 years later, and so is adjusted for mortality occurring 5 years later.

QRISK3 does not distinguish between the risk of different cardiovascular events, and so when an event occurs in the model, the probability of it being stable angina, unstable angina, MI, TIA, stroke, fatal CVD event or fatal CHD event is assumed. The NG202 model uses the relative distributions published in Ward 2007[79] for stable angina, unstable angina, MI, TIA, stroke, fatal CHD and fatal stroke. These data are estimated from the Markov model implemented by Ward et al 2007[79] and are from sources dating back to the 1990s. As discussed in the economic evaluation for NG202, incidence rates for CV events have changed over time, for example Davies et al 2007 [80] found that incidence of CVD risk decreased for men and women over the period 1996-2005, but it is unclear whether the distribution of CV events has changed during this time.

The EAG sought more recent evidence of the distribution of CV events. However, using the sources identified we could not extract the required CV event subgroups. Therefore, we use

the distribution of events used in NG202, noting this is a potential limitation of the model, but also that the NG202 state that British Heart Foundation statistics indicate that the distributions used in the model are “approximately correct”, given the example that coronary heart disease is approximately twice as common as stroke.

5.7.8 Road traffic accident (RTA) risk

In the NG202 model, the risk of an RTA for the general population is first estimated (see paragraph below). It is assumed in the model that individuals with OSAHS who are appropriately treated have an RTA risk equivalent to the general population. The increased risk of an RTA for people with OSAHS who are misdiagnosed or not appropriately treated is then adjusted. See section 5.7.11 for the assumed impact of untreated, or inappropriately treated, OSAHS on RTA risk.

To estimate the risk of a RTA in the NG202 model, data on the age and sex-specific probability of having a license in England were applied to the population of England to obtain the total number of drivers in England by age and sex. The number of (slight, serious and fatal) road traffic accidents where the driver was a casualty was divided by the total number of drivers to obtain a probability of a slight, serious or fatal driver accident at each age and for males and females. The EAG have updated the estimates reflecting accidents recorded in 2021[81].

5.7.9 Mortality

As stated in the description of the Markov model (see section 5.6.2), the mortality for individuals who do not have OSAHS is modelled using the age- and sex- specific general population lifetables for England and Wales (2021).[82] For individuals who do have OSAHS, non-cardiovascular mortality rates are estimated by subtracting the proportion of IHD and CVD deaths from general mortality rates using ONS 2018-2020 cause-specific data.[83] Mortality following an RTA is estimated as described in section 5.7.8 above.

Cardiovascular mortality is estimated from the distribution of events in the implementation of transition probabilities from QRISK3 (see section 5.7.7). For the post cardiovascular event states, mortality is based on standardised mortality rates for each of the five cardiovascular events. In the NG202 model these estimates were taken from a previous NICE guideline (NICE hypertension guideline 2019), with data sources from many years ago. The EAG sought to update these data sources, however found limited evidence within a UK population.

A number of sources were identified to inform the risk of death after stroke compared to the general population, however none were very recent. Dennis et al 1993[84] report on the risk of death following a first stroke in the Oxfordshire Community Stroke Project. Individuals had a mean age of 72 years and were followed-up for a minimum of 2 years, but no more than 6.5 years. Age- and sex- stratified mortality rates for the general population in Oxfordshire, were used as the comparator. Across all age groups, individuals who survived the first 30 days after experiencing a stroke had a 2.3 (95%CI 2.0, 2.7) times greater risk of dying than those in the comparator group.

Clarke et al 2003[85] also use data from the Oxfordshire Community Stroke Project to estimate the risks of death over 10 years in 290 patients who had previously experienced a TIA. Individuals included in the analyses had a mean age of 69 years, and 62% were male. There were 147 deaths and using age- and sex-adjusted mortality rates for England and Wales, Clarke et al reported a standardised mortality ratio (SMR) of 1.05 (95%CI 0.89, 1.23) for those having had a TIA. The low SMR from Clarke et al 2003 may be explained by the fact that individuals were recruited after a median of 3.8 years after the initial TIA event. Thus, anyone dying within this time period would not be included.

Rutten-Jacobs et al 2013[86] report on mortality following a first TIA, ischemic stroke or intracerebral haemorrhage for 959 individuals aged 18-50 years from the Netherlands. Patients who had had an event were identified from a prospective registry based at a single centre between 1980 and 2010. Twenty-seven percent (n=262) of individuals had a TIA, 63% (n=606) had an ischemic stroke and 9.5% (n=91) had an intracerebral haemorrhage and were followed up for a mean of 11.1 years (SD 8.7). To estimate the SMRs, mortality rates for individuals who survived the first 30 days after the event were compared to general population mortality rates stratified by age, sex and calendar year. Although these SMRs are more up to date than those used in previous versions of the NG202 model, they relate to a young Dutch population, therefore may not reflect the risk of death for an older UK population.

Ellis et al 2019 [87] report on the risk of death after an acute coronary syndrome event (MI or unstable angina) in New Zealand with a 12-year follow-up. Estimation of the standardised mortality ratio was based on 721 patients admitted with an ACS, and age-, sex- and Māori ethnicity adjusted mortality rates for the general population in New Zealand. A SMR of 1.3 (95% CI 1.2, 1.5) was reported.

The study by Plakht et al 2017 [88] is set in Israel and estimates standardised mortality rates after an MI over a 10 year follow-up of 2671 individuals. It is a retrospective analysis of data from individuals within a single medical centre who survived the initial hospital stay after the MI. At the time of the MI patients had a mean age of 66.6 years, and 67% were male. The authors report that age-, sex- and ethnicity/religion-specific mortality rates for the general population were used as the comparator. A SMR of 2.07 (95%CI 1.93, 2.23) is reported for those individuals having their first MI, a SMR of 2.91 (95%CI 2.57, 3.29) is reported for those having recurrent MI.

The more recent studies identified are not based in the UK, with one study including a much younger population than is modelled. Moreover, use of these more recent studies could lead to SMRs that may not have face validity across the different CV events, e.g. a SMR of 2.6 for TIA from Rutten-Jacobs et al 2013 being greater than a SMR for MI of 2.07 from Plahkt et al 2017 (see Table 21). Thus, we use the SMRs that were employed in the NG202 model for our base case analysis but look at alternative estimates in one-way sensitivity analyses.

Table 21 Base case and alternative SMRs for cardiovascular events

CV event	Base case SMR (95%CI)[a]	Source	Alternative SMR (95%CI)	Source
Stable angina	1.95 (1.65, 2.31)	Rosengren 1998 [89]	-	-
Unstable angina	2.19 (2.05, 2.33)	NG94 [90]	1.3 (1.2, 1.5)[b]	Ellis 2019[87]
MI	2.68 (2.48, 2.91)	Bronnum-Hansen 2001 [91]	1.3 (1.2, 1.5)[b] 2.07 (1.93, 2.23)	Ellis 2019[87] Plakht 2017[88]
TIA	1.4 (1.1, 1.8)	Dennis 1990 [92]	1.05 (0.89, 1.23) 2.6 (1.8, 3.7)	Clarke 2003[85] Rutten-Jacob 2013[86]
Stroke	2.72 (2.59, 2.85)	Bronnum-Hansen 2001 [93]	2.3 (2.0, 2.7) 3.9 (3.2, 4.7)	Dennis 1993[85] Rutten-Jacobs 2013[86]

[a] as used in NG202; [b] MI or unstable angina

5.7.10 Utilities

To model the utility of individuals without OSAHS and provide baseline utilities for individuals with OSAHS, we apply the age- and sex-specific utilities as obtained from Ara and Brazier 2010.[94] These utilities were used in the NG202 model.

In the NG202 model, for people with OSAHS, utility is assumed to depend on severity of OSAHS. This was done using the approach employed by McDaid 2009 [54] where severity, based on AHI, was estimated by the Epworth Sleepiness Scale (ESS). The ESS was subsequently mapped to EQ-5D to estimate utility associated with mild, moderate and

severe OSA, as defined by AHI (see Table 22). The mapping was conducted using individual participant data for a study evaluating the effectiveness of CPAP in a sample of 94 individuals with OSA in the UK.[54]

To update this analysis, we searched the QoL literature to identify studies directly estimating utility in people suspected or diagnosed with OSA by severity (see section 5.2). As noted in section 5.2.2, we found just one article reporting on the QoL of people with OSA that provided utility estimates (SF-6D) by AHI severity[55]. However, this study reported minimal difference in the utilities, with mild and severe having utility of 0.60 and moderate having utility of 0.61. This does not reflect the literature, which suggests that worse HRQoL is associated with increasing severity of OSA. Thus, we follow the approach taken in the NG202 economic model, where AHI is linked to ESS, which is in turned mapped to the EQ-5D. The general population EQ-5D utility estimates from Ara and Brazier are adjusted to reflect utility for mild, moderate and severe OSA. This is achieved by using mapping published by McDaid et al 2009 of mean baseline ESS values to EQ-5D values. The EQ-5D values obtained from the mapping, were then divided by the general population utility for 50 year old individuals (i.e. the age of the cohort entering the model, 0.876 for males and 0.855 for females). The resulting multiplier was then applied to EQ-5D values from the general population to estimate EQ-5D values for males and females with mild, moderate and severe OSA (Table 22).

The utilities for cardiovascular events in the NG202 economic model were taken from previous analyses which dated back many years. We therefore sought publications reporting EQ-5D values from the UK population to update the utility values for CV events.

Table 22 Utility multipliers for mild, moderate and severe OSAHS

Severity of OSAHS	Mean ESS (as reported in McDaid 2009)	Mean EQ-5D (as reported in McDaid 2009)	Utility multiplier for males (50 year old)	Utility multiplier for females (50 year old)
Mild	9	0.805	0.919	0.942
Moderate	13	0.766	0.875	0.897
Severe	16	0.737	0.842	0.862
Abbreviations: ESS Epworth Sleepiness Scale; EQ-5D Euro-Qol 5-dimensions				

Pockett et al 2018 [95] is a longitudinal UK study evaluating EQ-5D-3L in individuals after MI, unstable angina or stroke following an acute coronary syndrome event. Individuals were identified prospectively (n=1350 responders) and retrospectively (where the event had occurred prior to the start of the study, n=753 responders). Those identified retrospectively

did not provide responses at one month follow up and may not have responded to later surveys, depending on the time since their event. Respondents had a mean aged of 68.3 years (67.9% male) and were asked to complete the EQ-5D at 1, 6, 12, 18 and 24 months after discharge. Pockett 2018 undertook linear regression analyses and report EQ-5D regression coefficients for various baseline characteristics and events. We use these regression coefficients to estimate EQ-5D values for males and females for MI, unstable angina and stroke. We use the values reported at one month to reflect the health related quality of life of individuals within the year of the event health state, and values reported at 12 months to reflect quality of life in the post event states. We note that for individuals experiencing a stroke, this was assumed to occur after an acute coronary syndrome event, and this is not relevant to our model (as individuals are assumed to experience one CV event only).

Luengo-Fernandez et al 2013 [96] followed-up individuals from the UK who had experienced a stroke (n=748) or TIA (n=440). Follow-up with the EQ-5D-3L started one month after the event, then at 6, 12, 24 and 60 months after the event. Luengo-Fernandez 2013 also recruited a matched-control group, based on a representative sample of the general population from the 2006 Health Survey for England. Participants were matched on age, sex, medical history and marital status, with those individuals with a stroke event having a mean age of 75 years, and those with a TIA having a mean age of 73 years, just under half of the population were male. We used the reported mean EQ-5D values for TIA and stroke at the month 1 and month 12 time-points.

Rieckmann et al 2020 [97] reported a cross-sectional study on the HRQoL of 1263 individuals presenting with suspected angina as part of a European pilot study (which included sites in the UK). Patients were assigned a type of angina as determined during a nurse interview (prior to further diagnostic assessment). Individuals had a mean age of 61.1 years, with 54% males. EQ-5D-3L values for typical and atypical angina are reported. The publication of the full trial indicates that the EQ-5D data would have been collected within about a month of the event. There is no follow-up EQ-5D (the full trial only reports on EQ-5D VAS). The mean utility reported in Rieckmann et al 2020 for stable angina (0.64, SD 0.21) is lower than that for unstable angina (0.68, SD 0.20), but was not found to be statistically significantly different. We note that participants in this study did not have a formal diagnosis of stable or unstable angina when the EQ-5D values were estimated, that there are no follow-up data and that these values lack face validity when compared to estimates for other CV events from papers described above. To deal with the issue of face validity of the published utility values, we assume that stable angina in the first and subsequent years is similar to unstable

angina in subsequent years. This assumption is made in previous versions of this model, where those utilities for event and post-event stable angina were the same and equal to the unstable angina post-event state. In an adaptation to the NG202 model, we apply sex-specific cardiovascular event utilities where available (Table 23).

To estimate the utility associated with a serious RTA, we take the approach used in NG202, which used data from the Health Outcomes Data Repository[98]. EQ-5D data were available for 56 patients at six weeks after an RTA which required an inpatient hospital stay. Based on these data, it was assumed in the NG202 model that a serious RTA would lead to a utility of 0.62 in the first year after the accident, but that utility would return to the pre-accident value after that year.

For a slight RTA, data from a 12 month observational study of EQ-5D for 560 individuals who had experienced acute whiplash was used[99]. The authors reported an improvement of 0.0851 in EQ-5D at 12 months after the incident. It is therefore assumed that a slight RTA is associated with a decrease in utility of 0.0851 in the first year, but that in subsequent years the pre-accident utilities apply (Table 23).

Table 23 Cardiovascular event and RTA utilities used in the model

Event	Utility	Comment	Source
Slight RTA (new event)	-0.0851 (disutility)	Only assumed for year of event. Pre-accident utilities apply in subsequent years	Pink et al 2014 [99]
Serious RTA (new event)	0.62		McDaid et al 2009 [54], Currie et al 2005 [98]
Stable Angina (New Event)	0.834 (males) 0.776 (females)	Assumption	
Stable Angina (Post Event)	0.834 (males) 0.776 (females)	Assumption	
Unstable Angina (New Event)	0.724 (males) 0.649 (females)		Pockett et al 2018 [95]
Unstable Angina (Post Event)	0.834 (males) 0.776 (females)		
MI (New Event)	0.767 (males) 0.692 (females)		
MI (Post Event)	0.873 (males) 0.815 (females)		
TIA (New Event)	0.78		Luengo-Fernandez et al 2013 [96]
TIA (Post Event)	0.78		
Stroke (New Event)	0.631 (males) 0.556 (females)		Pockett et al 2018 [95]

Stroke (Post Event)	0.676 (males) 0.618 (females)		
Abbreviations: MI myocardial infarction; RTA road traffic accident; TIA transient ischemic attack			

5.7.11 Treatment

The EAG assume that individuals diagnosed with OSA are offered some form of treatment, and that this could be conservative management (lifestyle advice), mandibular advancement devices (MAD) or CPAP. The model assumes that treatment for OSA impacts on individuals in 3 potential ways: i) direct improvement in utility, ii) reduced risk of cardiovascular events, iii) reduced risk of road traffic accidents.

Treatments offered for OSAHS

The type of treatment offered to individuals depends on their diagnosed severity of OSA. Based on expert opinion we assume that for individuals who are diagnosed with mild OSA, 75% would receive conservative management only, 20% would receive CPAP, and the remaining 5% would have mandibular advancement devices. For individuals diagnosed with moderate or severe OSA, it is assumed that 90% would receive CPAP, and 10% would receive mandibular advancement devices.

In terms of the type of CPAP patients would receive, the NG202 guidelines state that fixed-level CPAP with telemonitoring should be considered, with auto-CPAP recommended in certain circumstances. Clinical experts who we consulted indicated that very few centres would be offering fixed-level CPAP, with the majority providing auto-CPAP. To reflect current practice, we assume in base case analyses that auto-CPAP is offered to patients. In scenario analyses, we assess the impact on the cost-effectiveness results of assuming fixed-level CPAP. We further assume that the mandibular advancement devices individuals would use are customised devices.

Adherence to OSAHS treatments

It is assumed that adherence to these treatments options is not 100%. In the NG202 model, treatment compliance for CPAP is taken from a 10-year study in England of 639 patients who received CPAP treatment for sleep apnoea.[100] The study reports data on drop outs by severity, in terms of ODI (mild ODI ≤ 15, moderate 15 ≤ ODI 30, severe ODI > 30), and these estimates are used in the model (see Table 24). The NG202 model assumed that after 10 years of use, those individuals who are still using CPAP would continue to do so. Furthermore, individuals who receive CPAP but do not have OSAHS (false positives), are assumed to stop CPAP treatment in the first year, due to a lack of an effect on their

symptoms. From Table 24 it can be seen that of those individuals with mild OSAHS who are still on treatment at the end of year 4, 98% will continue treatment in year 5.

Table 24 Proportion of patients who comply with treatment in the given year

Year	Mild	Moderate	Severe
1	0.88	0.90	0.95
2	0.90	0.95	0.97
3	0.95	0.95	0.99
4	0.97	0.97	0.98
5	0.98	0.98	0.99
6	0.98	0.97	0.97
7	0.99	0.99	1.00
8	0.98	0.99	0.98
9	0.91	0.96	1.00
10	1.00	1.00	0.91

There is a lack of long-term compliance data for mandibular advancement devices. The TOMADO trial report some data on compliance, depending on the type of MAD used, but the treatment period in the study was only 4 weeks.[66] We take the approach used in the NG202 model to estimate compliance to MADs, which was to assume that compliance would be the same for CPAP and MADs. In scenario analyses, we assess the impact of different assumptions on compliance, for instance using the short-term data from TOMADO, but also using expert opinion for the use of CPAP in population with mild OSAHS.

Direct effects of treatment on utility

As stated above, the model assumes that treatment for OSA has the potential to impact on utility (as well as impacting on CV event risk and RTA risk). Three treatments for OSA are assumed in the model: conservative management, CPAP and mandibular advancement splints. From our systematic review of HRQoL, we found no evidence reporting the direct impact of these treatments on utility, by OSA severity. Thus, we follow the previous approach of mapping effects observed on the ESS to EQ-5D using the mapping described by McDaid et al 2009.[54]

Treatment effects on ESS

As part of the NG202 evaluation, the authors undertook a systematic review of the evidence for CPAP treatment in individuals with mild OSA. Six studies met their inclusion criteria comparing CPAP to placebo/sham or standard care (i.e. conservative management). To investigate whether conservative management is associated with a treatment effect on ESS, studies comparing CPAP to conservative management were reviewed separately to those

comparing CPAP to placebo or sham. The NG202 authors expected that any treatment effect of conservative management would mean that CPAP was estimated to be less effective compared to conservative management, than when CPAP is compared to placebo/sham. The NG202 authors report that they actually found CPAP to be more effective when compared to conservative management, rather than when compared to placebo/sham. According to the NG202 review author, NG202 committee members believed it was unreasonable to assume that conservative management would lead to a decrease in utility, compared to no treatment. It was therefore assumed that treatment with conservative management would not impact on ESS.

Of the six studies identified in the NG202 review of CPAP in populations with mild OSAHS, three had restricted their population to only those with mild OSA. All three reported on the impact of CPAP treatment compared to placebo or standard care on the ESS. A meta-analysis, conducted by the authors of the NG202 evidence review, resulted in a CPAP-treatment estimate on ESS of -2.87 (95% CI -3.62, -2.11) for CPAP.

For the impact of CPAP in individuals with moderate or severe OSA, the NG202 authors used estimates reported in McDaid et al 2009.[54] Combined with the results of their meta-analysis of CPAP effect in a mild OSA population, the CPAP treatment effects assumed in the NG202 economic model[52] indicate that CPAP treatment is more effective in patients with mild OSA than in patients with moderate OSA (see Table 25).

We sought alternative estimates of the impact of CPAP on ESS. McMillan et al 2014 [51] report on the treatment of OSA with CPAP in an older UK population (>64 years old). They estimated a reduction in ESS for CPAP compared to best supportive care of 2.1 (95%CI 3.0, 1.3) at 3 months, and 2.0 (95%CI 2.8, 1.2) after 12 months. However, McMillan et al 2014 did not report a treatment effect by OSA severity, and report that at baseline participants had a mean ODI 11.1.

Table 25 Effect of CPAP treatment compared to best supportive care or placebo/sham

Severity	As used in NG202			Feltner et al 2022 meta-analysis		
	Mean (95%CI) difference in ESS	Mean improvement in EQ-5D [a]	Source	Mean (95%CI) difference in ESS	Mean improvement in EQ-5D [a]	Comments
Mild	-2.87 (-3.62, -2.11)	0.028	NG202 Evidence Report E, with McDaid mapping	-1.91 (-2.61, -1.2)	0.019	Includes studies with mild-only, and mild-moderate populations
Moderate	-2.04 (-2.99, -1.09)	0.023	McDaid et al 2009[54]	-2.21 (-2.92, -1.51)	0.021	Includes studies with moderate-only, and moderate-severe populations
Severe	-3.41 (-4.56, -2.26)	0.033	McDaid et al 2009[54]	-3.08 (-3.71, -2,45)	0.030	Includes studies with severe-only populations

Abbreviations: CI confidence interval; EQ-5D Euro-Qol 5-dimensions; ESS Epworth Sleepiness Scale

[a] EQ-5D estimates obtained using McDaid et al 2009 [54] mapping of ESS to EQ-5D

Feltner et al 2022 [101] conducted numerous systematic reviews to inform the US Preventative Services Task Force recommendations on screening for OSA in adults. One of these reviews focussed on the impacts of CPAP on ESS in individuals diagnosed with OSA. Feltner et al 2022 report that 47 studies met their inclusion criteria and had sufficient data to perform meta-analysis. Although treatment estimates for mild, moderate and severe OSA can be assumed from the meta-analyses conducted by Feltner et al 2002 (see Table 25), due to variability in the severity of study populations, some studies considered a range of severities at baseline. However, we note that there is some face validity with the impact of CPAP on ESS increasing as baseline OSA severity increases. Due to this face validity and the comprehensiveness of the systematic review by Feltner et al 2022, we use their estimates of CPAP treatment effectiveness on ESS in our base case analysis. We note that Feltner et al 2022 do not report any studies evaluating the impact of CPAP on EQ-5D.

In the NG202 economic model, the effectiveness of oral devices was taken from the UK-based TOMADO trial in individuals with mild-moderate OSA.[66] In this cross-over trial, three types of oral devices were evaluated: self-moulded, semi-bespoke and bespoke. Sharples et al 2014 report on the impact of the oral devices on EQ-5D, and on the ESS (n=83). The ESS estimates are reported directly in Sharples et al 2014, however the EQ-5D values are based on and assumed QALY increase over 4 weeks for each oral device over no treatment. This is then converted to a utility, by multiplying the incremental QALY by 52/4. As in the NG202 economic model, we use the EQ-5D utility estimates reported in TOMADO to estimate the effectiveness of oral devices. Note that TOMADO only included individuals with mild-moderate OSA. Due to a lack of RCT evidence on the effectiveness of oral devices in individuals with severe OSA, and the NG202 committee members believing such devices would have little impact on those with severe disease, the NG202 economic model assumed no impact of oral devices in individuals with severe OSA.

A systematic review conducted by Feltner et al 2022 [101] on the impact of oral devices found 6 RCTs reporting quality of life outcomes. However, the authors state that due to heterogeneity, inconsistency and issues of bias, their conclusion is that these studies are insufficient to make recommendations for the use of oral devices. Ten studies were identified that provided sufficient data on the impact of ESS to be combined in meta-analysis. All included studies with mixed-severity populations, either mild-moderate or mild-severe. Importantly, there were no studies only including individuals with severe OSA. Overall, oral devices were associated with a reduction in ESS of 1.67 (95%CI 2.09, 1.25). There is no discussion about the type of oral devices evaluated within these studies. Due to the greater detail available in the TOMADO trial and its direct relevance to the UK, we follow

the approach taken in the NG202 economic model and assume direct impacts of treatment on EQ-5D from TOMADO for those with mild or moderate OSA (Table 26). Individuals with severe OSA are assumed to have no benefit from the use of oral devices.

Table 26 EQ-5D and ESS impacts of oral devices (TOMADO, Sharples et al 2014)

Type of device	Incr QALY (SE) over 4 weeks vs no treatment	Estimated incremental EQ-5D	ESS: Mean difference (95%CI) vs no treatment	Mapped EQ-5D (from ESS values)
Self-moulded	0.00094 (0.00105)	0.012	-1.51 (-.2.29, -0.73)	0.015
Semi-bespoke	0.00088 (0.00123)	0.011	-2.15 (-2.99, -1.31)	0.021
Bespoke	0.00177 (0.00147)	0.023	-2.37 (-3.22, -1.53)	0.023
Abbreviations: EQ-5d Euro-Qol 5-dimension; ESS Epworth Sleepiness Scale; Incr incremental; QALY quality-adjusted life-year; SE standard error				

Treatment effects on blood pressure

In the NG202 economic model [52] it is assumed that treatment with CPAP for individuals with moderate or severe OSA leads a reduction in systolic blood pressure. This then translates into a reduction in the risk of cardiovascular events via QRISK3 (see section 5.7.7 above). Based on a meta-analysis, the NG202 economic model authors report that CPAP treatment is associated with a 1.06 mmHg reduction in blood pressure. They assume that conservative management would not have any impact on cardiovascular risk. Furthermore, they assumed that mandibular advancement devices would not impact on cardiovascular risk either, regardless of the underlying severity of OSA.

The EAG reviewed studies reporting on the impact of CPAP and mandibular advancement devices on cardiovascular risk factors (see Table 27). McMillan et al 2014 [51] reported a statistically significant reduction in systolic BP for the best supportive care group compared to the CPAP group of 3.7 mmHg (95%CI 0.2, 7.3) at 12 months. The reported baseline characteristics of the studies suggest that most of the included studies concerned individuals with severe OSA.

Feltner et al 2022 [101] report three systematic reviews evaluating the impact of CPAP on blood pressure. The most recent, Labarca et al 2021,[102] meta-analysed blood pressure data from six studies of CPAP compared to control arms in individuals with OSA and resistant hypertension. They found non-statistically significant reductions in systolic and diastolic blood pressure for the CPAP arm compared to the control arm. Feltner et al 2022 also

highlight a systematic review by Patil et al 2019,[103] which sought effectiveness evidence for all types of PAP, including CPAP. Across all patient subgroups (resistant hypertensive, hypertensive, normotensive or mixed), statistically significant reductions in blood pressure were observed for individuals in the PAP treatment arms compared to those in the control arms. However, studies varied in the severity of OSA of individuals included in the study populations. The third systematic review by Zhang et al 2016 [104] included five studies that restricted their populations to those unlikely to have symptoms (i.e. “populations with minimally symptomatic, asymptomatic, or non-sleepy OSA”, Feltner et al 2022 [101]). This systematic review found non-statistically significant impacts on systolic blood pressure.

Based on these systematic reviews, we assume that treatment with CPAP leads to a reduction of 2 mmHg in systolic blood pressure for individuals with moderate or severe OSA. This is mainly based on the results of the systematic review by Patil et al 2019. [103] Compared to these results, our assumption is conservative, but we note that the results from Labarca et al were not statistically significant. In sensitivity analyses, we assess the impact on the cost-effectiveness of novel devices by reducing this assumed improvement on systolic blood pressure.

Sharples et al 2014 [66] considered systolic blood pressure as an outcome in the TOMADO UK trial of mandibular advancement devices for individuals with mild to moderate OSA. They reported very little impact on systolic blood pressure for any of the three devices they evaluated over the treatment period of 4 weeks.

A systematic review of the impact of MAD on blood pressure was identified by Feltner et al 2022 (De Vries et al 2018 [105]). When the authors pooled available evidence from five RCTs (including the TOMADO study), they found a non-statistically significant reduction in systolic BP when compared to inactive controls: mean change of -1.55 mmHg (95%CI -3.92, 0.82). The five studies contributing to these estimates covered the range of severity. We assume that mandibular advancement devices do not impact on the risk of cardiovascular events for individuals with OSA. Conservative management is also assumed to have no impact on cardiovascular risk.

Table 27 Evidence on the impact of treatment for OSAHS on systolic blood pressure

Author Study type	Intervention, Study period	Mean difference (95% CI) mmHg	Comments
Positive Airway Pressure			
McMillan et al 2015[51] RCT	CPAP, 12 months	3.7 (0.2, 7.3)	Results are in favour of control arm
Labarca et al 2021[102] Systematic Review	CPAP, 12-24 weeks	-2.34 (-6.94, 2.27) [a]	Meta-analysis of six studies
Patil et al 2019[103] Systematic Review	PAP	-2.76 (-4.31, -1.20) [a]	Meta-analysis of 12 RCTs
Zhang et al 2016[104] Systematic Review	CPAP	-0.51 (-3.39, 2.38)	Meta-analysis of five studies Population defined as those unlikely to have symptoms
Mandibular advancement devices			
Sharples et al 2014[66] RCT	MAD 4 weeks	"very little impact"	
De Vries et al 2018[105] Systematic Review	MAD, 4 weeks to 41 months	-1.55 mmHg (-3.92, 0.82)	Meta-analysis of five RCTs (including Sharples 2014)

[a] Daytime blood pressure;
Abbreviations: CI confidence interval; CPAP continuous positive airway pressure; MAD mandibular advancement device; PAP positive airway pressure; RCT randomised controlled trial

Treatment effects on RTA risks

In the NG202 economic model[52], the impact of treatment on RTA risks was based on a meta-analysis of RTA rates before and after the start of CPAP, as reported in McDaid et al 2009.[54] Using the assumption that individuals with OSA who are treated with CPAP would have the same RTA risk as individuals in the general population, the probability of a RTA was increased using the resulting OR of 0.168 from McDaid to reflect an increase in risk for those not treated with CPAP. The model also assumed that treatment with mandibular advancement devices would also contribute to a reduction in risk of RTAs. However, conservative management was not assumed to impact on RTA risk. To maintain consistency with the NG202 model, we assume in the base case that there is an impact on the risk of an RTA from both CPAP and MAD, but not for conservative management.

However, we note that more recent evidence for an effect of CPAP on RTAs is less convincing. Feltner et al 2022[101] identified 3 RCTs assessing the impact of CPAP on RTAs, having a follow-up of 12-52 weeks. In one RCT (Lewis et al 2017 [106]) no RTAs were observed in either treatment arm: CPAP plus usual care (n=106) or usual care alone

(n=106). A second study[101] identified by Feltner, observed 10 RTAs in the CPAP arm (n=558) and 11 RTAs in the control arm (n=547)[107]. The third study identified was McMillan et al 2014 [51], who observed 2 RTAs in the CPAP arm (n=140) and 1 RTA in the control arm (n=138) over a 52 week period. None of these trials reported a statistically significant impact on the risk of RTA with CPAP. In scenario analyses we assess the impact of removing any treatment effect on risk of RTAs.

Summary of treatment effects

The treatment effects assumed in the model depend on the treatment received and the underlying true severity. Table 28 details the type of treatment offered to individuals depending on their true disease status and their diagnosis, alongside the assumed impacts of any treatment on utility, risk of CV events and RTAs.

Table 28 Treatment and impact depending on true and diagnosed severity

True severity	Diagnosed severity	Treatment (assuming adherence)	Impact of treatment on utility	Impact of treatment on CV risk	Impact of treatment on RTA risk
Mild	Mild	CM only	None	None	None
		CM + MAD	Improvement	None	Reduces risk
		CM + CPAP	Improvement	None	Reduces risk
	Moderate	CM + MAD	Improvement	None	Reduces risk
		CM + CPAP	Improvement	None	Reduces risk
	Severe	CM + MAD	Improvement	None	Reduces risk
		CM + CPAP	Improvement	None	Reduces risk
Moderate	Mild	CM only	None	None	None
		CM + MAD	Improvement	None	Reduces risk
		CM + CPAP	Improvement	Reduces risk	Reduces risk
	Moderate	CM + MAD	Improvement	None	Reduces risk
		CM + CPAP	Improvement	Reduces risk	Reduces risk
	Severe	CM + MAD	Improvement	None	Reduces risk
		CM + CPAP	Improvement	Reduces risk	Reduces risk
Severe	Mild	CM only	None	None	None
		CM + MAD	Improvement	None	None
		CM + CPAP	Improvement	Reduces risk	Reduces risk
	Moderate	CM + MAD	Improvement	None	None
		CM + CPAP	Improvement	Reduces risk	Reduces risk
	Severe	CM + MAD	Improvement	None	None
		CM + CPAP	Improvement	Reduces risk	Reduces risk

Abbreviations: CM conservative management; CPAP continuous positive airway pressure; CV cardiovascular; MAD mandibular advancement device; RTA road traffic accident

5.7.12 Costs of novel device testing

Even though four of the six novel home-testing devices are re-usable, all devices are priced per sleep study. There are differences between the devices in terms of additional costs if more than a one-night sleep study is required, and additional costs for any consumables per sleep study. Note that additional sleep studies may be required due to misdiagnosis (a patient with moderate or severe OSA is misdiagnosed as having no OSA) or due to a failure of the sleep study to provide a diagnosis.

Acurable (AcuPebble) and Itamar (WatchPAT) provide a cost per sleep study dependent on the number of sleep studies done per year (with AcuPebble), and per week (with WatchPAT). Based on the experience of clinical experts consulted by the EAG, the number of sleep studies per week currently ranges from 20 to 100. To calculate the cost per sleep study, we therefore assume in our base case that approximately 50-60 sleep studies would be done per week, which corresponds to approximately 3000 sleep studies per year.

Where it has been stated by companies that devices can be posted, we assume the cost of a small parcel to cover the postage (£2.99 for up to 2kg). As some devices are re-usable, so need to be returned to the clinic, it is assumed that they are posted back, incurring an additional postage cost of £2.99. Whether patients are posted their device or collect it from the hospital, we assume both would incur the cost of 10 minutes of band 4 staff admin time (£6.17). Thus, this cost is not included for the first sleep study undertaken by a patient as it applies to all. However, for any further postage or collection of devices for doing a re-test, due to failure to make a diagnosis or on-going symptoms, this cost is incurred.

The costs for each novel device are now described. Following these descriptions, the component and total costs for a successful one-night sleep study are shown in Table 30. In Table 31, the total costs for a repeat sleep study due to failure to obtain a diagnosis or misdiagnosis of individuals with moderate-severe OSA as not having OSA, are presented. The component costs associated with the second sleep studies are shown in Table 74 and Table 75.

Cost of AcuPebble SA100

The cost of a sleep study with AcuPebble is based on the volume of sleep studies conducted at the unit using AcuPebble. See Table 29 for the costs per sleep study by volume assumed for a one-night sleep study and the additional costs if a two-night, or more than two-night, sleep study were conducted.

Table 29 Costs of a sleep study with AcuPebble by volume of sleep studies assumed

Volume of sleep studies assumed per year (per week)	Cost for one-night	Additional cost for second night	Additional cost for third or more nights
1250 (25)	£49	£16	£0
3000 (60)	£44	£14	£0
5000 (100)	£40	£13	£0

Source: company submission and response to EAG questions

As we assume in our base case analysis that 50-60 sleep studies per week would be undertaken, we take the costs per sleep study assuming there are 3000 sleep studies per year with AcuPebble. The only consumable required for AcuPebble is the adhesive. This is costed as £100 per 100 adhesives, thus there is an additional cost of £1 per sleep study. We assume that the cost for a one night study would be the cost of the sleep study plus costs for postage, preparation and cleaning of device and adhesives. As the healthcare professional receives a notification once a sleep study has been completed, we assume that the data are reviewed before the patient returns the device. Therefore, should the data recorded be insufficient to make a diagnosis, for any reason, we assume that the patient would be asked to repeat the sleep study on the following night. This would not incur any additional costs for this sleep study. There is functionality in the mobile application for AcuPebble to detect any issues and ask the patient to repeat the sleep study. We therefore conduct two analyses for AcuPebble, one which assumes that this functionality is enabled (base case analysis), and that there is no input for the clinician; and one where this functionality is disabled (scenario analysis), and so time for checking the data by a member of staff is incurred (5 minutes of sleep specialist time to have reviewed the data from the first night, and a 10 minute phone call with the patient to discuss and request a repeat study).

According to the company submission, 10-15 minutes of support time for preparation of the device, including cleaning and putting details on the web or mobile application, is needed for AcuPebble. We assume 10 minutes of band 4 time for these activities, resulting in a cost of £5.67. No training of patients in the use of the device by NHS staff is needed for AcuPebble, according to the company submission. The company submission also includes estimates of the time taken for a sleep specialist to review the report automatically generated: 5-10 minutes. In the base case analysis, we assume it would take 20 minutes of band 6 sleep physiologist time to review the report (£17.67). In scenario analyses, we assume that this would take 10 minutes. The company state that the device can be posted to patients at home. We therefore assume a postage cost of £5.98 to cover postage to the patient and

postage back to the clinic after the sleep study. In alternative analyses we assume that the patient would collect and return the device in person, therefore incurring no postage costs.

We assume that individuals with moderate-severe OSAHS who are misdiagnosed with no OSA after an initial sleep study with AcuPebble would have a second sleep study using AcuPebble. If this second study is within a 2-week period, this would incur the cost of the second night (i.e. £14), however if it were more than 2 weeks after the initial sleep study, it would incur the cost of a new sleep study (£44). Additional costs for a second sleep study would include postage, adhesive, preparation and cleaning of the device, organisation of postage to or collection by the patient, as well as time taken to review data and prepare a report.

Cost of Brizzy

Brizzy report a cost per clinical question of £44. This cost covers any additional testing to answer the same clinical question. Thus, there are no additional costs should the patient require further testing due to misdiagnosis or failure to obtain a diagnosis, nor are there any additional costs for consumables. The company state that any future testing “due to new symptoms or developments” would constitute a new clinical question. In the company submission, it is mentioned that the device could be collected in person or sent. We conduct analyses assuming postage costs of £5.98 and without postage costs to reflect the patient collecting and returning the device in person.

The company state that after every 80-100 uses, the device would require 5 minutes to calibrate the sensor. We treat this as negligible and assume no cost for calibration. However, 5 minutes of support time is assumed to clean the device for the next patient (£2.83). The company further state that no training of patients in the use of the device by NHS staff is required. The device needs to be returned to the clinic for the data to be uploaded and a report to be generated. We assume in base case analyses, that 20 minutes of a band 6 sleep physiologist’s time would be required for this (£17.67). In scenario analyses, we assume this would take 15 minutes as indicated by the company.

As the device needs to be returned to the clinic to upload data, should a sleep study fail to make a diagnosis, we include costs for postage (where assumed), a sleep physiologist’s time to review the data/report (5 minutes), and a 10 minute phone call to the patient to request a second sleep study. We also include costs for preparing and sending out another device for the repeat sleep study (15 minute of support time, £13.25).

We assume that individuals with moderate-severe OSAHS who are misdiagnosed with no OSA after an initial sleep study with Brizzy would have a second sleep study using Brizzy. Since the sleep study costs for Brizzy are per clinical question, a repeat test would have no costs associated with the device itself. However, there would be additional costs for preparation of another device, organisation of postage to or collection by the patient, and postage (if assumed), as well as time to review data and prepare a report.

Cost of NightOwl

NightOwl is a disposable device and ResMed report a cost of £90 for a sleep study, with no consumables, no need for any preparation of the device, nor any training to be given to the patient by NHS staff. In their submission, ResMed assume 15 minutes of nurse time to prescribe the NightOwl, and 7.5 minutes of consultant time to prepare the report. For consistency with our assumptions across the novel devices, in the base case, we assume 20 minutes of band 6 sleep physiologist for report preparation, but do not include the costs for prescribing (as it is assumed that all patients incur this cost, and it would be cancelled out in the model). In scenario analyses, we assume 7.5 minutes of sleep physiologist time for report preparation, as indicated by the company. As the NightOwl can be posted, but is disposable, we assume a postage cost of £2.99 (assuming that any postage costs for returning the device for recycling are met by the company). In alternative analyses, we assume that the patient would collect the device in person.

The device has 10 days of battery power, thus if the sleep study should fail it can be repeated at no additional device cost. However, we assume that the sleep physiologist would take 5 minutes to identify that the study has not recorded sufficient information, and a 10 minute phone call to the patient to request the repeat study.

We assume that individuals with moderate-severe OSAHS who are misdiagnosed with no OSA after an initial sleep study with NightOwl would have a second sleep study using NightOwl. The company have confirmed that the device would be kept by the patient to be used for any further sleep studies within 3 years (up to 10 nights). Thus no additional device costs are incurred for these repeat sleep studies.

Cost of Sunrise

The list price for Sunrise is £75 per device. However, the company advised that the cost per sleep study would depend on the volume of devices ordered: 5-9 = £73, 10-49 = £68, 50-99 = £65, 100+ = £62. As stated above, we assume that between 50-60 sleep studies would be conducted per week, therefore as the Sunrise device is disposable, we assume that clinics

using this device would be ordering in batches of 100 or more. Thus, we use the cost of £62 per sleep study in the model. The company report that there are no consumables requiring payment in addition to the cost per sleep study, and that no training or preparation of the device is required. The device can be sent to patients' homes, but as it is a disposable device, the cost of postage is £2.99 (with the company stating that patients can use a pre-paid envelop to return the device for recycling, at no cost to the patient or NHS). In the base case, we assume it would take 20 minutes for a band 6 sleep physiologist to review the data. In scenario analyses, 10 minutes is assumed as indicated by the company.

In base case analyses, we assume that should a sleep study fail, the full cost of a new device would be incurred to undertake a second sleep study.

[redacted] As the device is single use only, in the event of a repeat sleep study, we include costs associated with postage (£2.99), 15 minutes of sleep physiologist time to identify the failed study and a phone call to the patient to request a repeat (£13.25), plus an additional 10 minutes of band 4 time to organise the postage or collection of the additional device (£5.67).

We assume that individuals with moderate-severe OSAHS who are misdiagnosed with no OSA after an initial sleep study with Sunrise would have a second sleep study using Sunrise. This second sleep study would incur the cost for the device, in addition to costs to organise postage to or collection by the patient, postage (where assumed) and data review.

Cost of WatchPAT 300

The list price for WatchPAT 300 is £50 per sleep study. The manufacturer of WatchPAT 300, Itamar, also provide a cost based on the number of sleep studies undertaken by the unit per week: 20-39 £55 per sleep study; 40-69 £50 per sleep study, and 70 or more £45 per sleep study. As we assume approximately 50-60 sleep studies are done per week, we use the cost of £50 per sleep study. In alternative analyses, we assume a cost per study of £45 for WatchPAT 300 (based on 100 sleep studies per week), and also a cost per sleep study of £55 (based on 25 sleep studies per week).

The company state that there are consumables associated with the device, a finger probe, but that this does not result in additional costs to the NHS, as the cost is per sleep study. Preparation of the device is required before it can be given to patients. The company

submission reports that this includes a few minutes to add adhesive and a new finger probe, plus physical connection to a computer “to initialise the device for a new patient”. Including a few minutes for cleaning of the device, we assume that preparation would take 10 minutes for a band 4 member of staff, leading to a cost of £5.67. The company also indicate that training on use of the device would take 5 minutes in person, or a video is available to patients. We therefore assume that in the scenario where the patient collects the device in person, they would have 5 minutes training by a band 4 member of staff. However, should the device be posted to the patient, it is assumed that the patient would not require training from NHS staff, but access the online training. As the device can be posted, we assume the cost of £5.98 to cover the postage of the device to the patient’s home, and back to the clinic. In the base case, we assume it would take 20 minutes for a band 6 sleep physiologist to review the data. In scenario analyses, 10 minutes is assumed as indicated by the company.

As the WatchPAT 300 needs to be returned to the clinic for data download, should a sleep study fail, this would only be detected at that point. The company state that the NHS only pays for “successful (technically valid)” sleep studies, thus there is no additional sleep study or device cost to the NHS for a failed sleep study. However, we assume that a failed study would incur the costs of 5 minutes of band 6 sleep physiologist time to download the data and identify the sleep study as a failure, plus 10 minutes for a phone call with the patient to discuss and request a repeat sleep study. Where it we assume that the device is posted, we include £5.98 for the postage to the patient’s home and return to the clinic. We also include the cost of 20 minutes for a band 4 to prepare the device and organise re-sending, or collection, of the device for the repeat sleep study (£11.33).

We assume that individuals with moderate-severe OSAHS who are misdiagnosed with no OSA after an initial sleep study with WatchPAT 300 would have a second sleep study using WatchPAT 300. This second sleep study would incur the cost for the device, preparation of the device, in addition to costs to organise postage to, or collection by, the patient, postage (where assumed) and data review.

Cost of WatchPAT ONE

This is a disposable device which has a list price of £80 per sleep study. The manufacturer also offer a cost of £90 per sleep study when the device is sent directly from the manufacturer (rather than from the sleep clinic). Based on the experience of experts consulted by the EAG, we assume that the sleep clinic would supply the devices to the patients, therefore we assume a cost of £80 per sleep study. The company state that there are consumables associated with the device, a finger probe, but that this does not result in

additional costs to the NHS, as the cost is per sleep study. Before the device is given to patients, a “quick registration” of the device is required, which we assume to take 5 minutes of time for a band 4 member of staff (£2.83). According to the company, NHS staff training of patients in the use of the device is not required. The mobile application provides instructions on how to use the device. We include postage costs of £2.99 (where assumed). As the device is disposable, the company state that they will fund the postal costs for patients to return the device to them for recycling. Once a sleep study is completed, the data are downloaded automatically. In the base case, we assume it would take 20 minutes for a band 6 sleep physiologist to review the data. In scenario analyses, 10 minutes is assumed as indicated by the company.

A poster (Storey et al 2022) identified from the clinical effectiveness review reports on the costs per appointment for WatchPAT ONE compared to Nox-T3, among other outcomes. The costs are stated to include equipment, room, staff and postage costs, and are reported to be £73.16 for WatchPAT ONE. This estimate is lower than our estimate of approximately £100 for the sleep study. However, it would appear that the estimate from Storey is associated with NHS contact with the patient prior to them undertaking the sleep study, so does not include time taken to review data. Further differences are difficult to identify as there are limited details in the poster, with the unit costs are not being reported.

Should a sleep study fail, an additional device would need to be sent to the patient. There is no additional sleep study or device cost to the NHS for a failed sleep study with the WatchPAT ONE. However, we assume that a failed study would incur the costs of 5 minutes of band 6 sleep physiologist time to download the data and identify the sleep study as a failure, plus a 10 minute phone call with the patient to discuss and request a repeat sleep study. Where we assume that the device is posted, we include £2.99 for the postage to the patient’s home. We also include the cost of 15 minutes for a band 4 to prepare the device and organise re-sending, or collection, of the device for the repeat sleep study.

We assume that individuals with moderate-severe OSAHS who are misdiagnosed with no OSA after an initial sleep study with WatchPAT ONE would have a second sleep study using WatchPAT ONE. This second sleep study would incur the cost for the device, preparation of the device, in addition to costs for organisation of postage to, or collection by, the patient, postage (where assumed) and data review.

Volume of sleep studies

In scenario analyses we assume that there are 25 sleep studies conducted per week (approximately 1250 per year), and 100 sleep studies conducted per week (approximately 5000 per year). These assumptions impact on the costs associated with AcuPebble and WatchPAT 300. Although costs associated with the Sunrise device depend on the volume of the order, we still assume that clinics would order at least 100 of the disposable devices at a time. Thus, the costs for Sunrise would not change from that discussed above.

The impact of assuming a different volume of sleep studies with AcuPebble would be to reduce total costs for initial and repeat sleep studies by £4 when assuming 5000 studies per year and increase total costs by £5 when assuming 1250 studies per year. The impact of assuming a different volume of sleep studies with WatchPAT 300 would be to reduce total costs for initial and repeat sleep studies by £5 when assuming 100 studies per week and increase total costs by £5 when assuming 25 studies per week.

Table 30 Component and total costs for each type of sleep study, per successful one-night study (assuming 50-60 studies per week)

	Novel devices						Comparators	
	AcuPebble SA100	**Brizzy**	**NightOwl**	**Sunrise**	**WatchPAT 300**	**WatchPAT ONE**	**Oximetry**	**Respiratory polygraphy**
Cost components								
Device per one night sleep study	£44	£44	£90	£62	£50	£80	£0.53	[a]
Consumables per one night sleep study	£1	£0	£0	£0	£0	£0	£0[b]	[a]
Preparation of device	£5.67	£2.83	£0	£0	£5.67	£2.83	£2.83	[a]
Training of patients	£0	£0	£0	£0	£0 posted £2.83 collected [c]	£0	£5.67	[a]
Postage	£5.98	£5.98	£2.99	£2.99	£5.98	£2.99	£5.98	[a]
Data and report review/ preparation based on 20 minutes	£17.67	£17.67	£17.67	£17.67	£17.67	£17.67	£8.83[d]	[a]
Total costs								
Assuming device posted and 20 minutes to review data	£74.31	£70.48	£110.66	£82.66	£82.15	£103.49	£18.18	NA
Assuming device collected/returned in-person and 20 minutes to review data	£68.33	£64.50	£107.67	£79.67	£76.17	£100.50	£17.86	£212.36

[a] All cost components are assumed to be captured in the NHS unit cost for home respiratory polygraphy; [b] Batteries included in sleep study cost; [c] It is assumed that if the device is collected in person, face-to-face training is provided; [d] 10 minutes to review data

Table 31 Total costs for a repeat sleep study due to failure to obtain a diagnosis or misdiagnosis (assuming 50-60 studies per week)

	Novel devices						Comparators	
	AcuPebble SA100	**Brizzy**	**NightOwl**	**Sunrise**	**WatchPAT 300**	**WatchPAT ONE**	**Oximetry**	**Respiratory polygraphy**
Failed sleep study[a]								
Assuming the device is still with the patient	£1 if automated £14.25 otherwise	NA	£13.25	NA		NA	NA	NA
Assuming device posted to/from patient	NA	£27.73	NA	£83.91	£30.56	£24.74	NA	NA
Assuming device is collected/returned in person	NA	£21.75	NA	£80.92	£24.58	£21.75	£212.36[b]	£212.36
Repeat sleep study due to misdiagnosis [c]								
Assuming device posted to/from patient	£79.98	£32.15	£23.33	£88.32	£84.98	£109.16	NA	NA
Assuming device is collected/returned in person	£74.00	£26.17	NA	£85.33	£79.00	£106.17	£212.36 [b]	£212.36

[a] **does not include the costs of** 20 minutes to review data and prepare report (as this is accounted for in the initial sleep study costs); [b] cost of home respiratory polygraphy; [c] assume repeat test for misdiagnosis is not within 2 weeks

5.7.13 Costs of home oximetry and respiratory polygraphy

In the NG202 economic model,[52] the costs associated with home oximetry were based on the Nonin device. After consulting with clinical experts, we also assume that the costs for this device are reflective of those for home oximetry. To update the costs of this device from 2020 costs as assumed in the NG202 economic model, we sought prices from the NHS Supply Chain. However, we were not able to access this information, and have therefore inflated the costs of devices to cost year 2021/22 using the NHS Cost Inflation Index (pay and prices) and used updated resource use costs.[108]

In the NG202 economic model it was assumed that a band 2 health care assistant would spend 15 minutes with the patient to advise how to use the Nonin device at home, and that a consultant would spend 15 minutes reviewing data and preparing the report. In contrast to these assumptions, and after consideration by our clinical experts, we assume that a band 4 member of staff would provide 10 minutes of training, and a band 6 sleep physiologist would take 10 minutes to review the results and prepare a report. We also assume that the equipment would need cleaning (5 minutes). According to the experts we consulted, there is variation in how pulse oximetry is delivered to the patient: in some clinics the patient will collect and return the device in person, while in others, the device is posted. We therefore undertake analyses assuming the device is posted, and assuming the device is collected and returned in person. Should a sleep study with home oximetry fail, we assume that a home respiratory polygraphy sleep study would be undertaken, and so incur the cost of home respiratory polygraphy in the model.

We assume that individuals with moderate-severe OSAHS who are misdiagnosed with no OSA after an initial sleep study with home oximetry would have a second sleep study using home respiratory polygraphy. This second sleep study would incur the cost for the home respiratory polygraphy. Detail on the costs for a successful home pulse oximetry study are shown in Table 30 above, alongside costs for the novel devices.

As in NG202, we assume that the cost of a home respiratory polygraphy sleep study, including the NHS costs for reviewing data, are reflected in the National Schedule of NHS Costs estimate for an outpatient respiratory sleep study £212.25 (DZ50Z). We had planned to estimate the costs associated with a particular device, as for home oximetry, for a sensitivity analysis. However, as we could not access cost information from the NHS Supply Chain, we are unable to do this. In scenario analyses, we assume a lower cost for an outpatient respiratory sleep study using a weighted average of the Respiratory Medicine

Service (n=20,649 procedures at £175.11 each) and Respiratory Physiology Service (n=21,460 procedures at £204.37 each) descriptions: £190.02.

5.7.14 Costs of hospital polysomnography

As in NG202, we assume that the cost of an in-hospital PSG sleep study is reflected in the National Schedule of NHS costs estimate for a Non-Elective Inpatient Short Stay, Respiratory Sleep Study (DZ50Z): £1,818.13. Use of hospital polysomnography is only assumed in scenario analyses.

5.7.15 Costs of treatment

Cost of conservative management (lifestyle advice)

As per the NG202 guidelines, we assume that all patients diagnosed with OSAHS receive lifestyle advice. This is reflected in the cost of a respiratory medicine consultant-led outpatient appointment: £194.30 (National Schedule of NHS Costs - Year 2021/22). In sensitivity analyses, we assume that the cost of conservative management consists of a letter to the patient, rather than an outpatient appointment.

Cost of CPAP

We base the costs of CPAP treatment on those used in the NG202 economic model,[52] updated to 2021/22 costs using the NHS Cost Inflation Index for devices and PSSRU 2021/22 Unit Costs of Health and Social Care [108]. In the NG202 model, five different CPAP strategies were assumed and costed. The 2021/22 costs for these five strategies (for the first year and on-going years) are shown in Table 32.

The device and consumables component includes annuitised costs for the CPAP device, mask, humidifier, hose, filter and chamber. For auto-CPAP strategies, staff time for auto-titration are also included. In the first year, the staff cost component consists of a respiratory consultant-led outpatient appointment for initial education and set-up of the device, and a non-consultant follow-up appointment at 3 months. In subsequent years, the cost of an outpatient appointment for an annual review is included. In the fixed-CPAP strategies, we follow the approach in NG202 where it is assumed that 18% need re-titration [109] within the first year. Telemonitoring costs comprise staff costs, and a telemonitoring access cost (from NG202 economic model). Further details on these cost components are given in Table 76 Appendix 8.

Table 32 Annual costs for CPAP treatment

Type of CPAP	Device and consumables	Staff (excluding retitration)	Retitration	Tele-monitoring access	Total cost
Auto-CPAP: Year 1	£159.74	£327.00	NA	NA	£486.75
Auto-CPAP: On-going years	£159.74	£132.71	NA	NA	£292.45
Auto-CPAP with telemonitoring: Year 1	£159.74	£327.00	NA	£30	£516.75
Auto-CPAP with telemonitoring: On-going years	£159.74	£132.71	NA	£30	£322.45
CPAP with autotitration: Year 1	£140.88	£327.00	£10.64	NA	£478.52
CPAP with autotitration: On-going years	£140.88	£132.71	£0	NA	£273.59
CPAP with telemonitoring: Year 1	£140.88	£327.00	£3.18	£30	£501.07
CPAP with telemonitoring: On-going years	£140.88	£132.71	£0	£30	£303.59

Cost of mandibular advancement devices

We assume that patients would be offered customised mandibular advancement splints. The costs for these are based on those used in the economic model for the NG202 guidelines.

We could not confirm the costs of devices assumed in the NG202 model; therefore we focus on the customised device used in the TOMADO study.[66] Sharples et al report that the manufacture of the device was estimated to take 7 hours for a grade 6-8 NHS maxillofacial technician. We assume the hourly rate for band 8d, as in Sharples et al. This is estimated as £46.23,[108] leading to a manufacture cost of £323.59 (£46.23 x 7). Sharples et al further report that the cost of consumables was negligible, and therefore no additional manufacture costs were assumed. We annuitised costs assuming a lifetime of 2 years for the device.

To obtain the impression for the customised mandibular advancement splint, a maxillofacial consultant-led appointment is assumed. As in NG202, we assume that the patient would have a 3 month and then annual consultant-led appointment, plus an additional sleep study

within the first year to assess effectiveness of the splint. For the customised splint it is assumed that such appointments would be with a maxillofacial consultant, rather than respiratory medicine. This follows the assumptions made in the NG202 economic model.

The total costs in the first year are £716.82 plus the cost of an additional sleep study, which is assumed to be of the same type used for diagnosis (i.e. one of the six novel devices, home oximetry or home respiratory polygraphy). The cost for subsequent years is £338.10.

5.7.16 Cost of cardiovascular events

In the base case analysis, we take costs of cardiovascular events as in the NG202 economic model[52], inflated to 2021/2022), which were derived from the economic model for the NICE guideline on hypertension (NG131). See Table 33 below.

5.7.17 Cost of road traffic accidents

Data from the most recent Department for Transport estimates (2021) [81] on the total costs per fatal, serious and slight accident, are divided by the total number of these accidents to estimate the NHS cost per RTA event (see Table 33). Due to the cost perspective for NICE being NHS and PSS, in the base case analysis, we only consider the NHS costs for a RTA. However, as in NG202, we also estimate the associated police costs and combine these with the NHS costs in a scenario analysis.

Table 33 Cardiovascular event and road traffic accident costs

Event	Value	Source
Cardiovascular events (costs per year for each state)		
Stroke	£19,169	NCG and SSNAP 2016 [108], NG202[52]
Post stroke	£7,277	NCG and SSNAP 2016 [108], NG202[52]
TIA	£1,902	Danese 2016 [110], NG202[52]
Post TIA	£639	Danese 2016 [110], NG202[52]
MI	£5,057	Danese 2016 [110], NG202[52]
Post MI	£837	Danese 2016 [110], NG202[52]
Stable angina	£1059	NHS Ref Costs 2021/2022 [108]
Post stable angina	£298	Assumed to be as post unstable angina
Unstable angina	£2,545	Danese 2016 [110], NG202[52]
Post unstable angina	£298	Danese 2016 [110], NG202[52]
Road traffic accidents (cost per event)		
Fatal (NHS costs)	£1,119	Department for Transport 2021 [81]

Serious (NHS costs)	£15,930	Department for Transport 2021 [81]
Slight (NHS costs)	£5,777	Department for Transport 2021 [81]
Fatal (NHS and police costs)	£1,910	Department for Transport 2021 [81]
Serious (NHS and police costs)	£18,435	Department for Transport 2021 [81]
Slight (NHS and police costs)	£28,843	Department for Transport 2021 [81]
Abbreviations: MI myocardial infarction; TIA transient ischemic attack		

5.8 Model assumptions

There are three sets of parameters that differ between the novel devices and the comparators (and between novel devices), which cause any differences in costs and outcomes:

- Sensitivity and specificity at the two diagnostic cut-offs (AHI ≥ 5 and AHI ≥ 15)
- Failure rates
- Cost per sleep study using each device.

Although the model allows for differential time to diagnosis and treatment for each novel device and comparator, due to a lack of evidence, these parameters are assumed the same across all devices.

Table 34 lists the key assumptions in the base case EAG economic model.

Table 34 Key base case model assumptions

Model component	Key assumption
Population	The cohort is assumed to include individuals suspected of having OSAHS, who are eligible for a home sleep study
	The mean age is 50 years, and 70% are male
Diagnostic accuracy	The low and high thresholds for sensitivity and specificity used in the decision tree reflect the classification of individuals into OSAHS severity and no OSAHS diagnosis subgroups. As noted, this is a simplification based on a lack of more detailed evidence of the classification of individuals from the published diagnostic accuracy studies.
	The setting of the sleep study is important. We do not directly compare data from studies where the device was evaluated in the home, with studies where the device was evaluated in a sleep clinic. However, all accuracy data in the base case are from clinic-based studies.

	To compare the accuracy of novel devices and comparators, it is assumed that the reference standard of PSG has perfect accuracy (sensitivity and specificity equal to 1); and the different types of PSG can be considered equivalent across studies.
	Diagnostic accuracy data for Nox-T3 from Xu et al 2017 [44] is representative of the accuracy of home respiratory polygraphy currently used in England.
	Diagnostic accuracy data for WatchPAT 200U from Tauman et al 2020 [30] reflects the accuracy of WatchPAT 300 and WatchPAT ONE
Diagnostic pathway	All individuals offered treatment would commence treatment within 12 months of the start of the model.
	Individuals may have up to two sleep studies: - Due to a failure of the first study to obtain a diagnosis, or - Due to misdiagnosis of people with true moderate to severe OSAHS as having no OSA (and an assumption that they would continue to be symptomatic)
	For people with mild OSAHS who are misdiagnosed as having no OSAHS, it is assumed that they are never correctly diagnosed.
	The second sleep study would be of the same type as the initial sleep study (except for pulse oximetry where the second sleep study would be respiratory polygraphy).
	When only one sleep study is needed, it takes 3 months to obtain a diagnosis and another 3 months to start treatment, when treatment is offered. The need for a second sleep study due to misdiagnosis would delay time to treatment by one month.
Natural history	People with OSA are assumed to be at higher risk of CV events and RTAs than the general population. The risk of CV events is modelled using QRISK3 based on characteristics of the cohort (age and sex) and other CV risk factors, following assumptions in the NG202 guideline economic model. The risk of RTAs is estimated using age- and sex-specific probabilities for a slight, serious or fatal RTA.
	People without OSA are assumed to have general population mortality, and no increase in CV or RTA risk. As treatment of people who do not have OSA is assumed to have no impact, these CV and RTA events are not captured in the model, as they would cancel out across the different diagnostic pathways.

	Individuals can experience only one CV event
	Once patients leave the acute CV event state (after one year), they transition to the post-CV event state. It is assumed they remain in this state until death.
	RTAs can occur in any of the alive states. Having a slight or serious RTA does not affect the transition probabilities of the Markov model.
Utilities	The direct effect of OSAHS on HRQoL (utility) depends on the severity of OSAHS (defined by AHI). A mapping algorithm is used to link scores on the ESS (a proxy for AHI) to EQ-5D utilities. Utility is also adjusted for age and gender (using general population data) and reduced for patients who experience non-fatal cardiovascular events and road traffic accidents.
	Disutility associated with acute whiplash is assumed to reflect reduced utility due to a slight RTA.
Treatment for OSAHS	All individuals diagnosed with OSAHS are assumed to receive lifestyle advice (conservative management). Depending on the diagnosed severity, a proportion will also receive CPAP, and a smaller proportion will receive MAD
	Treatment has the potential to impact on - Utility - Risk of CV events, via a decrease in systolic blood pressure - Risk of RTAs.
	Treatment with either MAD or CPAP is assumed to improve utility for those with OSAHS (regardless of severity). Conservative treatment (lifestyle advice) is not assumed to have any impact on utility, CV risk or RTA risk.
	CPAP leads to a reduction of 2mmHg in systolic blood pressure for people with moderate to severe OSAHS but has no impact on for people with mild OSAHS.
	Those correctly diagnosed with mild or moderate OSA who receive MAD have a reduced risk of RTA events only. MAD has no impact on RTA risk for people with severe OSAHS. Treatment with MAD has no impact on CV risk, regardless of severity of OSAHS.
	Treatment has no impact on people who do not have OSAHS
	After 10 years of adherence to CPAP or MAD treatment, it is assumed that people continue to adhere to treatment for the rest of their life

Treatment for OSAHS	People with OSAHS who remain misdiagnosed after up to two sleep studies are assumed to remain undiagnosed and untreated. Therefore, they continue to be at untreated levels of risk for CV events and RTAs, and have reduced utility.
Costs	The Nonin WristOx2 device is representative of the types of home oximetry devices currently used in practice
	The NHS Reference Cost for an outpatient respiratory sleep study reflects the costs associated with current home respiratory polygraphy studies
	We assume NHS staff time for any preparation or cleaning of devices and any face-to-face training of patients in the use of the device
	When a sleep study fails, where appropriate, we assume costs for NHS staff to identify the study as a failure and speak to the patient to discuss and request that a second study is done
	Costings for CPAP follow the approach used in NG202, which includes annuitised device and consumable costs and staff time
	As assumed in the NG202 economic model, an additional sleep study is costed within the first year to check the devices are working effectively. It is assumed that the same type of sleep study used for diagnosis is used for this purpose
Abbreviations: CPAP continuous positive airway pressure; CV, cardiovascular; ESS Epworth Sleepiness Scale; EQ-5D EuroQol-5 dimensions; MAD mandibular advancement device; PSG polysomnography; RTA road traffic accident;	

5.9 Model validation

Two authors (CH and JL) who were not involved in coding the Excel model conducted a series of quality assurance and validation checks. We used a 'QA log' that has been developed by the EAG over a number of years to identify and record errors and uncertainties in economic models, and to assess the face validity of the results. This includes a list of tests that are potentially useful for model verification and validation, which reflect published guidance (AdViSHE).[111] We have conducted the following checks:

- *Input checks*: verification that the data and parameter values in the model match those in the cited source and in this report.
- *'White box' checks*: manual assessment of formulae and links between settings, data inputs, parameter selection and probabilistic sampling, decision tree and Markov calculations and the collation of results. A copy of the model was annotated to

document the cells checked and identify potential errors and questions to be addressed.

- '*Black box' checks*: changes to model inputs and settings to check that they have the expected effect on intermediate and final results. This included variation of parameters (e.g. changes to prevalence and severity, diagnostic accuracy statistics, utilities and costs), as well as extreme value test (such as setting utilities to 1 and 0 to check that QALYs are the same as life years).
- *Face validity*: consideration of whether intermediate and final outcomes reflect model inputs. For example, checking that the numbers of false negative and false positive results and repeat tests, and are consistent with diagnostic accuracy parameters.

Any issues identified were discussed with the model developers (BG and JP) and resolved, with a correction to the model if appropriate.

5.10 Economic analysis results

The cost-effectiveness results are reported using a number of intermediate outcomes and summary statistics. These include 'pairwise' Incremental Cost-Effectiveness Ratios (ICERs) for each novel device compared to respiratory polygraphy; and pairwise ICERs for each novel device compared to oximetry. The ICER is defined as [112]

ICER = Incremental costs / Incremental QALYs,

where the incremental costs are calculated as

Total costs from novel device – Total costs of RP (or oximetry)

and the incremental QALYs are calculated as

Total QALYs from novel device – total QALYs from RP (or oximetry).

We also summarise the cost-effectiveness results using incremental net monetary benefit (INMB) statistics [112]. These are obtained by multiplying the incremental QALYs by a willingness-to-pay (WTP) threshold before subtracting the incremental costs:

INMB = (Incremental QALYs * WTP threshold) – Incremental costs

As NICE's stated WTP threshold is between £20,000 and £30,000 per QALY gained, we report INMBs at these two thresholds. The INMB has several advantages over ICERs when interpreting cost-effectiveness results for multiple interventions [112]: INMBs avoid confusion

between ICERs in different quadrants of the cost-effectiveness plane; and they avoid the need to find and exclude options subject to ‘simple’ or ‘extended’ dominance before the most cost-effective option can be identified. A positive INMB indicates that the intervention is cost-effective relative to the comparator at the defined WTP threshold. And importantly, INMB statistics can be used to rank interventions in order of cost-effectiveness. An intervention with a higher positive INMB is more cost-effective than an intervention with a lower INMB. The option which gives the greatest positive INMB is the economically optimal choice.

5.10.1 Base case results

Deterministic results

Costs and outcomes associated with the comparators and the novel devices are shown in Table 35 below. Note that these estimates are for a mixed general population aged over 16 being tested for suspected OSAHS, as described in sections 5.4.1 and 5.7.1. We have not been able to produce estimates for people aged 16 and under, or for the subgroups requested in the scope due to a lack of relevant data (sections 2.1 and 4.2.1). Test performance, diagnostic and treatment costs and health outcomes are likely to differ for these groups, and there are potential equality issues that require consideration. See section 6.4 for further discussion.

The diagnostic pathway using oximetry is estimated to be the cheapest, while respiratory polygraphy is the most expensive option. Costs associated with cardiovascular events make up over 60% of the total costs for all devices. Respiratory polygraphy is associated with the highest estimated QALYs (14.142). This is mainly driven by the generally higher sensitivity estimates compared to most of the novel devices. Due to the low estimates of sensitivity associated with oximetry, it has the lowest proportion of true positives, and thus is associated with the longest time to treatment (as people with moderate-severe OSAHS who are misdiagnosed are assumed to have a delay in the start of treatment due to the need for a second sleep study). We emphasise that there is high uncertainty over the relative diagnostic accuracy estimates for all devices and advise caution in interpreting these simple deterministic base case results.

Figure 5 shows the estimated proportions of the modelled cohort who are diagnosed or misdiagnosed by severity of OSAHS. Generally, the novel devices appear to perform well in identifying true OSAHS severity. None of the devices misdiagnose people with moderate or severe OSAHS with no OSAHS. This is due to the base case assumption that such individuals would receive a second sleep study due to on-going symptoms.

It can be clearly seen that oximetry is estimated to lead to greater misdiagnosis of OSAHS across severities. While the relatively poor performance of WatchPAT 300 and WatchPAT ONE at ruling-out OSAHS can be seen in the second and third columns of Figure 5. Low estimates of specificity at the AHI = 5 diagnostic cut-off in the model lead to people who are falsely diagnosed as having OSAHS incurring treatment costs for a year without accruing any health benefits.

WatchPAT 300 and ONE also perform poorly in that they lead to over-diagnosis of the severity of people with mild OSA. This can be seen in the sixth column of Figure 5, where a higher number of individuals with mild OSAHS are diagnosed as having moderate-severe OSAHS by WatchPAT 300 and ONE. This 'over-diagnosis of severity' leads to more people who truly have mild disease being offered CPAP or MAD, than would be the case if they were diagnosed correctly with mild OSAHS. This is because we assume in the model base case that those diagnosed as having mild OSAHS may receive conservative management, CPAP or MAD, while those diagnosed with moderate-severe OSAHS are only treated with CPAP and MAD. Since only CPAP and MAD are assumed to impact on utility in the truly mild group, not conservative management, the greater chance of being offered CPAP or MAD if diagnosed with moderate-severe OSAHS leads to greater utility gains. Please see Table 29 in Section 5.7.11 for details of treatment impacts. These increases in QALYs are also associated with increased costs of CPAP and MAD treatment for those individuals with mild OSAHS who are misdiagnosed as having moderate-severe OSAHS.

There is very little difference between devices in the proportion of the population predicted to experience a CV event (Table 35). This is due to the relatively small impact of CPAP on CV risk estimated in the model. In addition, CPAP treatment only impacts on CV risk in those individuals diagnosed as moderate or severe who truly have moderate or severe OSAHS and are compliant with treatment. The estimated incidence of RTAs is more variable across the devices, ranging from 12.0% for oximetry to 9.4% for WatchPAT 300 and WatchPAT ONE.

In comparison with respiratory polygraphy, all novel devices are associated with lower costs and QALYs (Figure 6), but the incremental QALYs in particular are very small. Pairwise cost-effectiveness results are reported in Table 36, with each novel device compared against respiratory polygraphy and also against oximetry. For AcuPebble and Sunrise, the reduction in QALYs may be considered cost-effective compared to the reduction in costs (i.e. INMB > £0 at the £20,000 and £30,000 per QALY thresholds). For other novel devices, the INMB

estimates are negative at both thresholds. Relative to oximetry, all novel devices have a positive INMB at both £20,000 and £30,000 per QALY gained thresholds.

We have not reported ‘fully incremental’ ICERs in Table **36** because the estimated cost and QALY differences between novel devices are very small and uncertain, as shown in the sensitivity and scenario analysis results presented below. This suggests that comparisons of cost-effectiveness between the novel devices may not be reliable.

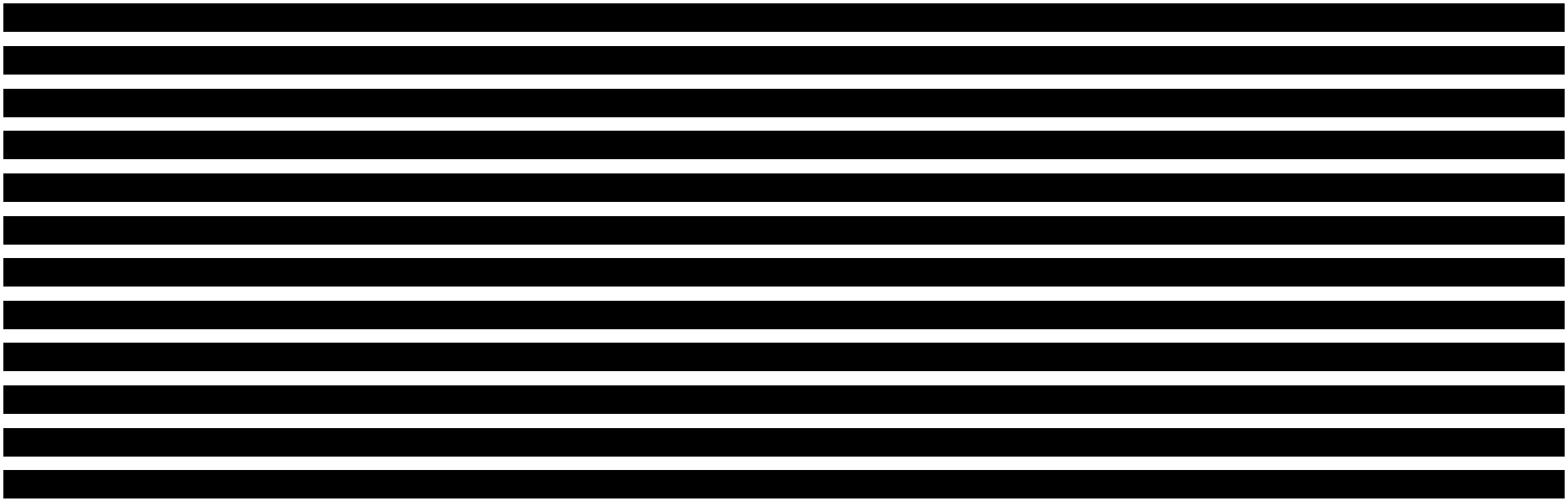


Figure 5 Number of people in the modelled cohort diagnosed with OSAHS by device, shown by true severity of OSAHS

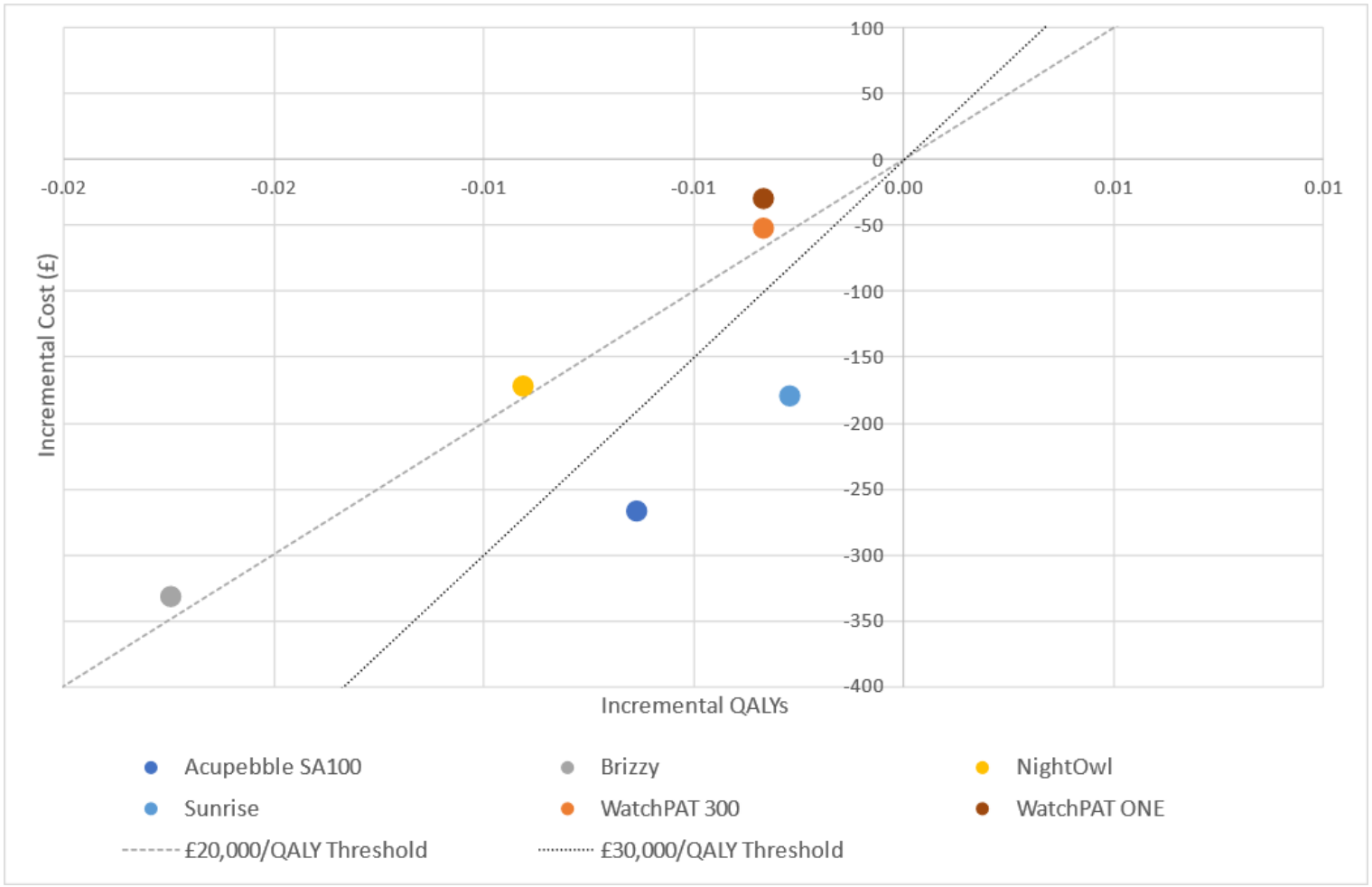


Figure 6 Incremental costs and QALYs for the six novel devices compared to respiratory polygraphy

Table 35 Cost and intermediate outcome results, deterministic base case

	Comparators		Novel devices					
			Clinic					
	Oximetry	Respiratory polygraphy	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Mean costs per person with suspected OSAHS [a]								
Diagnosis	£65	£239	£80	£74	£114	£98	£90	£113
Treatment	£1,903	£2,611	£2,492	£2,425	£2,557	£2,569	£2,713	£2,713
CV	£5,246	£5,241	£5,240	£5,241	£5,241	£5,241	£5,242	£5,242
RTA	£358	£285	£296	£303	£290	£287	£277	£277
Total	**£7,572**	**£8,376**	**£8,108**	**£8,043**	**£8,202**	**£8,195**	**£8,322**	**£8,346**
Intermediate outcomes								
True positives (%) [b]	68.9	80.6	[redacted]	80.1	80.0	79.4	80.8	80.8
True negatives (%) [b]	16.2	12.6	[redacted]	18.1	14.0	17.1	4.5	4.5
False positives (%) [b]	1.9	5.6	[redacted]	0	4.1	1.1	13.6	13.6
False negatives (%) [b]	12.9	1.2	[redacted]	1.8	1.8	2.4	1.1	1.1
Time to treatment [c]	6.17	6.00	[redacted]	6.00	6.00	6.00	6.00	6.00
Long-term outcomes								
Mean QALYs [a]	14.050	14.142	14.135	14.124	14.133	14.139	14.138	14.138
Mean LYs [a]	30.80	30.81	30.81	30.81	30.81	30.81	30.81	30.81
CV events (%)	48.56	48.55	48.54	48.55	48.55	48.55	48.56	48.56
RTA events (%)	11.97	9.59	9.95	10.18	9.76	9.67	9.37	9.37

[a] Discounted; [b] may not add to 100% due to rounding; [c] months

Abbreviations CV cardiovascular; RTA road traffic accident; LYs life-years; QALYs quality-adjusted life-years

Table 36 **Cost-effectiveness results, deterministic base case**

Setting of evaluation	Clinic					
	AcuPebble SA100	**Brizzy**	**NightOwl**	**Sunrise**	**WatchPAT 300**	**WatchPAT ONE**
Compared to respiratory polygraphy						
Incremental cost	-£267	-£333	-£174	-£181	-£53	-£30
Incremental QALYs	-0.006	-0.017	-0.009	-0.003	-0.003	-0.003
ICER (£ per QALY gained)	£42,212[a]	£19,128[a]	£19,227[a]	£67,426[a]	£16,172[a]	£9,163[a]
INMB at £20,000 per QALY gained	£141	-£15	-£7	£127	-£13	-£36
INMB at £30,000 per QALY gained	£77	-£189	-£97	£100	-£46	-£69
Compared to oximetry						
Incremental cost	£536	£471	£630	£623	£750	£773
Incremental QALYs	0.085	0.074	0.082	0.089	0.088	0.088
ICER (£ per QALY gained)	£6302	£6360	£7646	£7020	£8515	£8777
INMB at £20,000 per QALY gained	£1165	£1009	£1017	£1152	£1012	£989
INMB at £30,000 per QALY gained	£2016	£1749	£1841	£2039	£1893	£1870

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Probabilistic results

We explored the impact of uncertainty over the model input parameters in probabilistic sensitivity analysis. The mean and 95% confidence ranges for cost and QALY outcomes from the probabilistic analysis are shown in Table 37. The wide and overlapping confidence ranges illustrate the high degree of uncertainty over the incremental costs and QALYs for each novel device compared with respiratory polygraphy and with oximetry. For example, the incremental costs for WatchPAT 300 compared with respiratory polygraphy range from -£298 to £235 and the incremental QALYs range from -0.040 to 0.033. This uncertainty is reflected in the confidence ranges estimated around the INMBs for the novel devices, for example, comparing Sunrise to respiratory polygraphy, the INMB at £20,000 per QALY gained ranges from -£238 to £572.

A scatter plot of the probabilistic incremental QALYs and incremental costs for each novel device and respiratory polygraphy compared to oximetry is shown in Figure 7. The cost-effectiveness acceptability curves are shown in Figure 8. For low WTP thresholds, oximetry is estimated to be the most cost-effective option. However, as the WTP threshold increases, AcuPebble becomes the most cost-effective option (with a probability of around 50%) at £10,000 per QALY gained, and remains the most cost-effective option across the higher WTP thresholds, but with decreasing probability of being the most cost-effective option.

Table 37 Probabilistic base case analyses

	Comparators		Novel devices					
			Clinic					
	Oximetry	RP	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£7599 (5403, 9825)	£8397 (6145, 10687)	£8133 (5871, 10426)	£8079 (5824, 10405)	£8,229 (5,951, 10,487)	£8226 (5931, 10492)	£8354 (6133, 10587)	£8373 (6125, 10614)
Total QALYs	14.04 (13.07, 14.92)	14.13 (13.25, 14.96)	14.13 (13.26, 14.96)	14.12 (13.24, 14.95)	14.12 (13.23, 14.96)	14.13 (13.26, 14.96)	14.13 (13.25, 14.95)	14.13 (13.25, 14.95)
Compared to respiratory polygraphy								
Incremental cost			-£264 (-475, -64)	-£318 (-524, -119)	-£169 (-395, 35)	-£171 (-319, -2)	-£44 (-298, 235)	-£24 (-286, 237)
Incremental QALYs			-0.006 (-0.047, 0.029)	-0.016 (-0.056, 0.017)	-0.009 (-0.051, 0.025)	-0.002 (-0.027, 0.026)	-0.002 (-0.04, 0.033)	-0.003 (-0.045, 0.036)
ICER (£ per QALY gained)			£ 43,505[a]	£20,199[a]	£19,640 [a]	£108,795 [a]	£21,216 [a]	£ 8,570 [a]
INMB at £20,000 per QALY gained			£143 (-519, 678)	£3 (-652, 482)	-£3 (-697, 487)	£140 (-238, 572)	£2 (-557, 533)	-£32 (-661, 508)
INMB at £30,000 per QALY gained			£82 (-993, 954)	-£154 (-1229, 640)	-£89 (-1204, 727)	£124 (-513, 821)	-£18 (-950, 840)	-£60 (-1131, 837)
Probability cost-effective at £20,000/ QALY			78.90%	58.40%	58.10%	81.60%	55.20%	48.10%
Probability cost-effective at £30,000/ QALY			65.30%	40.00%	46.40%	69.50%	51.50%	46.50%

	Comparators		Novel devices					
			Clinic					
	Oximetry	RP	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Compared to oximetry								
Incremental cost			£535 (168, 992)	£481 (107, 903)	£630 (236, 1028)	£627 (262, 1041)	£755 (329, 1204)	£775 (328, 1197)
Incremental QALYs			0.085 (0.008, 0.223)	0.076 (0.003, 0.209)	0.083 (0.002, 0.204)	0.09 (0.012, 0.224)	0.089 (0.009, 0.224)	0.089 (0.008, 0.212)
ICER (£ per QALY gained)			£6271	£6360	£7613	£6989	£8458	£8750
INMB at £20,000 per QALY gained			£1171 (-196, 3549)	£1031 (-190, 3357)	£1025 (-306, 3249)	£1168 (-186, 3510)	£1031 (-321, 3419)	£996 (-411, 3276)
INMB at £30,000 per QALY gained			£2023 (-64, 5774)	£1787 (-129, 5462)	£1852 (-218, 5294)	£2066 (-59, 5736)	£1923 (-204, 5707)	£1881 (-270, 5408)
Probability cost-effective at £20,000 / QALY			94.30%	92.80%	90.30%	93.70%	88.90%	87.60%
Probability cost-effective at £30,000 / QALY			96.80%	95.90%	94.40%	96.70%	95.00%	93.80%

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio, INMB incremental net monetary benefit, QALYs quality-adjusted life-years

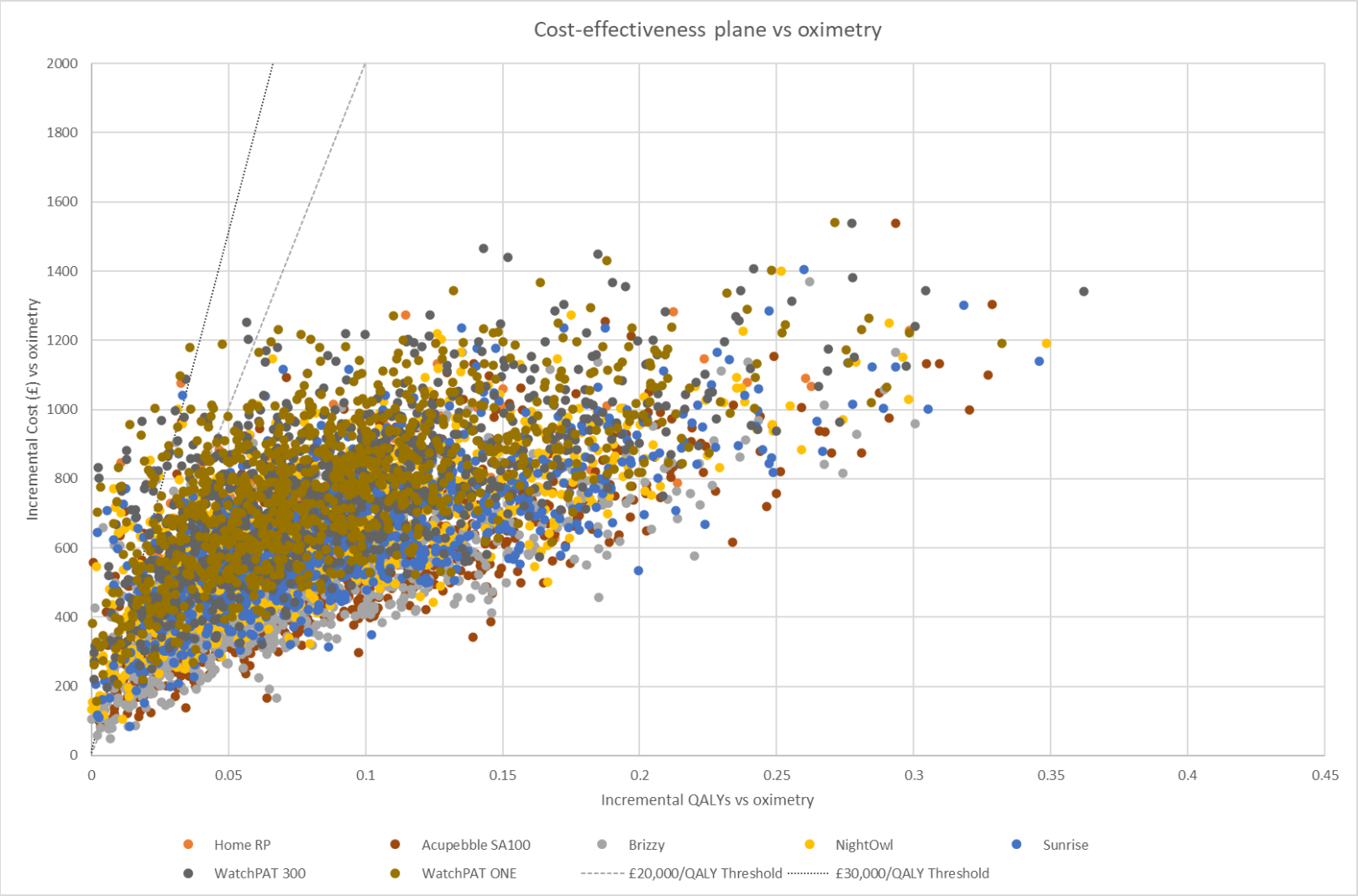


Figure 7 Scatterplot of the incremental costs and QALYs compared to oximetry from the probabilistic analysis

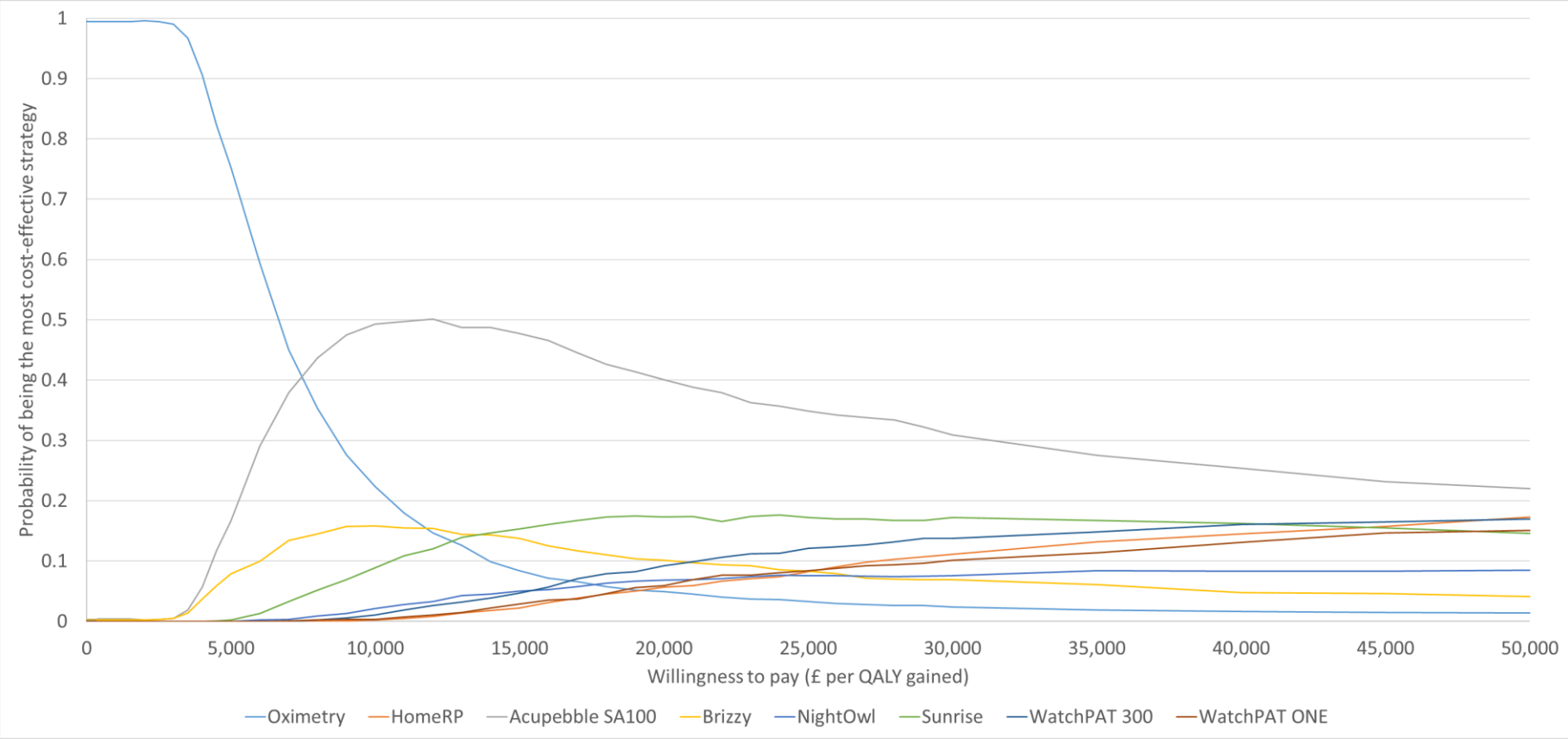


Figure 8 Cost-effectiveness acceptability curves

5.10.2 Scenario analyses

We conducted a range of scenario analyses to assess the impact of different model assumptions on the cost-effectiveness results (Table 38).

Table 38 Scenario analyses assumptions

Assumption	Base case	Scenario analysis	Further details
Alternative accuracy data for respiratory polygraphy	From Xu et al 2017	From Pereira et al 2013 From NG202 economic model	5.7.4
Risk of CV events in modelled cohort	Based on assumed cohort characteristics	Lower risk cohort High risk cohort	Table 16
Correlation between first and second sleep study results in terms of misclassification	No correlation	20% correlation 40% correlation	Section 5.7.6
Sensitivity and specificity of novel devices	Point estimates used as reported or calculated for the articles	Lower 95% confidence limits used as a worst case scenario Upper 95% confidence limits used as a best case scenario	Section 5.7.3
Alternative decision tree parameterisation	Simplification based on low and high diagnostic cut-offs as in NG202 economic model	Using the 4x4 contingency table to inform accuracy for respiratory polygraphy, AcuPebble, NightOwl, WatchPAT 300 and WatchPAT ONE	Section 5.6.1
Accuracy data for AcuPebble	From unpublished AIC data supplied by the company	Accuracy data as reported in Devani et al	Section 4.5.1 and section 5.7.3
Accuracy data for NightOwl	Data source is Lyne 2023	Data source: Massie 2018 and Van Pee 2020	Section 5.7.3
Accuracy data for Sunrise	Data source is Pepin 2020	Data source is Kelly 2022	Section 5.7.3
Third sleep study for false negatives	Diagnosis assumed after maximum of two sleep studies	Those moderate-severe misdiagnosed after two	Section 5.7.14

Assumption	Base case	Scenario analysis	Further details
		sleep studies have a laboratory-based PSG	
Transport of novel devices to/from patient's home	Posted	Collected in person	Section 5.7.12
Transport of oximetry to/from patient's home	Collected in person	Posted	Section 5.7.13
Volume of sleep studies conducted per centre per year or week	Approximately 3000 per year, equivalent to 50-60 per week	Approximately 1250 per year, equivalent to 25 per week. Approximately 5000 per year, equivalent to 100 per week	Affects costs of sleep studies for AcuPebble and WatchPAT 300, section 5.7.12
Lower cost of home RP	£212.36	£190.02	Section 5.7.13
Cost of Sunrise	Assuming 100 or more devices purchased per batch	List price assumed	Section 5.7.12
Functionality of AcuPebble to automatically alert patients on failed sleep study	Assumed enabled	Assumed disabled	Section 5.7.12
Failure rate data for respiratory polygraphy	Data source Newcastle Regional Sleep service	Data source Alsaif et al 2023 manuscript	Section 5.7.5
Failure rate for Sunrise	Data source Kelly 2022	Data source Alsaif et al 2023 manuscript	Section 5.7.5
Additional device cost for repeat of failed Sunrise sleep study	All repeats incur device costs	[redacted]	Section 5.7.5
Time for data review of novel devices	20 minutes	As reported by the companies	Section 5.7.12
Time to treatment	Assumed to be six months regardless of sleep study type	Time to diagnosis for all novel devices and comparators [redacted] and reduced time to treatment [redacted] for	Section 5.7.6

Assumption	Base case	Scenario analysis	Further details
		novel devices (Alsaif unpublished study)	
Cost of RTAs	NHS costs only, and only 1 casualty per RTA	NHS and police costs. >1 casualty per RTA	Section 5.7.17
Cost of conservative management	Consultant-led outpatient appointment	Letter	Section 5.7.15
Type of CPAP	Auto-CPAP	CPAP with autotitration CPAP with telemonitoring CPAP with telemonitoring (1st year only) Auto-CPAP with telemonitoring Lower auto-CPAP cost	Section 5.7.15 and Appendix 8 Table 76
Longer-term treatment impacts	Treatment impacts on CV risk and risk of RTAs	Treatment impacts on CV events only. Treatment impacts on RTAs only. Treatment has no impact on CV events or RTAs.	Section 5.7.11
CPAP for people diagnosed with mild OSAHS	20% receive CPAP, and compliance based on UK study	75% receive CPAP, but 50% discontinue in first year.	Section 5.7.11
Compliance for oral devices	Assumed same as for CPAP	Slightly higher drop-out	Section 5.7.11
Time horizon	Lifetime	10 years	
Impact of false positives	Costs of treatment for 1 years	Costs of treatment for 2 years	Section 5.7.11
Abbreviations: CPAP continuous positive airway pressure; CV cardiovascular; PSG polysomnography; RTA road traffic accident			

Many of the scenario analyses made no difference to the overall conclusions from the base case analysis. This included assumptions of the correlation between results of initial and repeat sleep studies, a lower cost of home RP, incorporating police costs in RTA costs, assuming more than one casualty per RTA, more (and less) expensive CPAP treatment, assuming no impacts on CV events or RTAs from appropriate treatment, reductions in treatment compliance, assuming that people with moderate-severe OSAHS who are still

misdiagnosed after two sleep studies would have a laboratory-based PSG and reduced time to diagnosis and treatment for novel devices compared to respiratory polygraphy and oximetry. The INMB at a £20,000 WTP threshold for all novel devices compared to respiratory polygraphy and oximetry for these scenario analyses are presented in Appendix 9a (Table 77).

A number of scenario analyses lead to changes to the INMB at £20,000 per QALY gained for some novel devices compared with respiratory polygraphy, resulting in the INMB being very close to £0. These scenarios were: assuming a high CV risk cohort (WatchPAT 300 has INMB of £1), assuming a low CV risk cohort (Brizzy has INMB of £0), assuming the estimated time to review given by the companies (NightOwl has an INMB of £5), and assuming that conservative management only incurs the cost of a letter (WatchPAT 300 has an INMB of £2), see Appendix 9a Table 77.

Below, we discuss details of the results for scenario analyses which had larger impacts on the overall conclusions.

Scenario analyses for diagnostic accuracy estimates

The first two scenario analyses presented below used alternative data sources for the accuracy of respiratory polygraphy. Using data from the study by Pereira et al 2013 (Table 39) or that used in the NG202 economic model (Table 40), improves the estimated cost-effectiveness of all novel devices relative to respiratory polygraphy: all novel devices have a positive INMB compared to respiratory polygraphy at both WTP thresholds. This is because the diagnostic accuracy estimates from Pereira and NG202 are less favourable for respiratory polygraphy than the estimates from Xu et al (used in our base case analysis). Note that estimates for the novel devices compared to oximetry also change when alternative data for the accuracy of respiratory polygraphy are used. This is due to the assumption that should a second sleep study be needed after oximetry, it would be home respiratory polygraphy. However, the overall conclusions for novel devices compared to oximetry do not change.

Table 41 shows the results when the lower 95% confidence limits for the sensitivity and specificity of the novel devices are used. Only Sunrise is estimated to be cost-effective, at both willingness to pay thresholds. When the upper 95% confidence limits for the sensitivity and specificity of the novel devices are used, all are assumed to be cost-effective compared to respiratory polygraphy (see Table 42).

When the raw accuracy data from the 4 x 4 contingency tables for respiratory polygraphy, AcuPebble, NightOwl, WatchPAT 300 and WatchPAT ONE are used in the decision tree, WatchPAT 300 is estimated to dominate respiratory polygraphy (Table 43). This is driven by the reduction in performance of respiratory polygraphy to identify people with mild OSAHS as having OSAHS, and the fact that WatchPAT is still more likely to over-diagnose the severity of mild OSAHS. As discussed for the base case results, this is likely to lead to greater QALYs. The improvement in cost-effectiveness with NightOwl is also explained by it being more likely than respiratory polygraphy to over-diagnose severity for patients with mild OSAHS when the 4 x 4 contingency data are used.

[redacted]

When the data from Massie et al 2018 are used to inform the accuracy of NightOwl, it is estimated to dominate respiratory polygraphy: it is associated with greater QALYs at lower cost. This is due to the higher sensitivity estimates at both AHI diagnostic cut-offs for NightOwl from Massie et al, than from Lyne 2023, see Table 44. When van Pee 2020 data inform the performance of NightOwl, NightOwl is estimated to be less costly (by £108) and less effective (by 0.001 QALYs) than respiratory polygraphy, but the reduction in QALYs is estimated to be cost-effective compared to the reduction in costs, see Table 45. This is driven by a slight increase in the sensitivity of NightOwl at the AHI ≥ 15 diagnostic cut-off when van Pee data are used (0.91) compared to when data from Lyne are used (base case analysis, 0.89), and a decrease in the specificity of NightOwl at the AHI ≥ 15 diagnostic cut-off (0.76 from van Pee and 0.82 from Lyne in the base case). This seemingly unintuitive finding is driven by the fact that with more people having mild OSAHS being misdiagnosed as having moderate-severe OSAHS they have a greater chance of receiving CPAP (rather than conservative management), which is associated with utility gains. This pattern can be seen with the scenario analysis where more people with mild OSAHS are assumed to receive CPAP that is assumed in the base case analysis, please see section below.

Table 39 **Scenario analysis using data from Pereira et al 2013 to inform the diagnostic accuracy of respiratory polygraphy**

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,094	£7,496	£8,108	£8,043	£8,202	£8,195	£8,322	£8,346
Total QALYs	14.098	14.036	14.135	14.124	14.133	14.139	14.138	14.138
Compared to respiratory polygraphy								
Incremental cost			£14	-£51	£108	£101	£228	£251
Incremental QALYs			0.037	0.026	0.035	0.041	0.040	0.040
ICER (£ per QALY gained)			£382	Dominant	£3,129	£2,471	£5,673	£6,247
INMB at £20,000 per QALY gained			£730	£574	£582	£716	£577	£553
INMB at £30,000 per QALY gained			£1,102	£835	£927	£1,125	£979	£956
Compared to oximetry								
Incremental cost			£612	£547	£706	£699	£826	£849
Incremental QALYs			0.099	0.088	0.096	0.103	0.102	0.102
ICER (£ per QALY gained)			£6,189	£6,225	£7,337	£6,815	£8,105	£8,332
INMB at £20,000 per QALY gained			£1,366	£1,210	£1,218	£1,352	£1,213	£1,189
INMB at £30,000 per QALY gained			£2,355	£2,089	£2,180	£2,378	£2,232	£2,209
Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years								

Table 40 **Scenario analysis using data used in the NG202 model to inform the diagnostic accuracy of respiratory polygraphy**

Setting	**Comparators**		**Clinic**					
	Respiratory polygraphy	**Oximetry**	**AcuPebble SA100**	**Brizzy**	**NightOwl**	**Sunrise**	**WatchPAT 300**	**WatchPAT ONE**
Total costs	£8,235	£7,531	£8,108	£8,043	£8,202	£8,195	£8,322	£8,346
Total QALYs	14.117	14.043	14.135	14.124	14.133	14.139	14.138	14.138
Compared to respiratory polygraphy								
Incremental cost			-£127	-£192	-£33	-£40	£87	£111
Incremental QALYs			0.018	0.007	0.016	0.022	0.021	0.021
ICER (£ per QALY gained)			Dominant	Dominant	Dominant	Dominant	£4,092	£5,174
INMB at £20,000 per QALY gained			£493	£337	£346	£480	£340	£317
INMB at £30,000 per QALY gained			£677	£410	£502	£700	£554	£531
Compared to oximetry								
Incremental cost			£577	£512	£671	£664	£791	£814
Incremental QALYs			0.093	0.082	0.090	0.096	0.096	0.096
ICER (£ per QALY gained)			£6,232	£6,276	£7,462	£6,897	£8,273	£8,515
INMB at £20,000 per QALY gained			£1,275	£1,119	£1,127	£1,261	£1,121	£1,098
INMB at £30,000 per QALY gained			£2,200	£1,934	£2,026	£2,223	£2,077	£2,054
Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years								

Table 41 Scenario analysis using lower 95% confidence limits for all novel device sensitivity and specificity estimates

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,376	£7,572	£7,961	£8,038	£8,103	£8,185	£8,352	£8,375
Total QALYs	14.142	14.050	14.104	14.110	14.107	14.136	14.126	14.126
Compared to respiratory polygraphy								
Incremental cost			-£414	-£338	-£273	-£191	-£24	-£1
Incremental QALYs			-0.038	-0.032	-0.035	-0.006	-0.016	-0.016
ICER (£ per QALY gained)			£11,045[a]	£10,575[a]	£7,842[a]	£31,633[a]	£1,553[a]	£47[a]
INMB at £20,000 per QALY gained			-£336	-£301	-£423	£70	-£288	-£311
INMB at £30,000 per QALY gained			-£711	-£621	-£771	£10	-£443	-£467
Compared to oximetry								
Incremental cost			£389	£466	£531	£613	£779	£803
Incremental QALYs			0.054	0.059	0.057	0.085	0.076	0.076
ICER (£ per QALY gained)			£7,222	£7,832	£9,375	£7,180	£10,279	£10,589
INMB at £20,000 per QALY gained			£689	£723	£601	£1,095	£737	£713
INMB at £30,000 per QALY gained			£1,227	£1,318	£1,167	£1,948	£1,495	£1,472

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 42 Scenario analysis using upper 95% confidence limits for all novel device sensitivity and specificity estimates

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,376	£7,572	£8,189	£8,123	£8,250	£8,209	£8,248	£8,271
Total QALYs	14.142	14.050	14.150	14.138	14.147	14.143	14.145	14.145
Compared to respiratory polygraphy								
Incremental cost			-£187	-£253	-£126	-£167	-£128	-£105
Incremental QALYs			0.008	-0.004	0.006	0.001	0.003	0.003
ICER (£ per QALY gained)			Dominant	£66,721 [a]	Dominant	Dominant	Dominant	Dominant
INMB at £20,000 per QALY gained			£353	£177	£242	£194	£195	£172
INMB at £30,000 per QALY gained			£435	£139	£300	£207	£228	£205
Compared to oximetry								
Incremental cost			£617	£551	£678	£636	£676	£699
Incremental QALYs			0.100	0.088	0.097	0.093	0.095	0.095
ICER (£ per QALY gained)			£6,187	£6,285	£6,973	£6,862	£7,132	£7,374
INMB at £20,000 per QALY gained			£1,377	£1,202	£1,267	£1,218	£1,219	£1,196
INMB at £30,000 per QALY gained			£2,374	£2,078	£2,239	£2,146	£2,167	£2,144

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 43 Scenario analysis with parameterisation of decision tree for respiratory polygraphy, AcuPebble, NightOwl, WatchPAT 300 and WatchPAT ONE based on the raw 4x4 contingency table data

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,378	£7,572	£8,043	£8,043	£8,258	£8,195	£8,359	£8,382
Total QALYs	14.142	14.050	14.129	14.124	14.137	14.139	14.144	14.144
Compared to respiratory polygraphy								
Incremental cost			-£334	-£335	-£119	-£183	-£19	£5
Incremental QALYs			-0.012	-0.017	-0.005	-0.003	0.003	0.003
ICER (£ per QALY gained)			£27,797[a]	£19,257[a]	£25,645[a]	£68,655[a]	Dominant	£1,711
INMB at £20,000 per QALY gained			£94	-£13	£26	£129	£73	£49
INMB at £30,000 per QALY gained			-£26	-£187	-£20	£103	£100	£77
Compared to oximetry								
Incremental cost			£471	£471	£686	£623	£787	£810
Incremental QALYs			0.079	0.074	0.087	0.089	0.094	0.094
ICER (£ per QALY gained)			£5,936	£6,360	£7,910	£7,020	£8,363	£8,609
INMB at £20,000 per QALY gained			£1,116	£1,009	£1,048	£1,152	£1,095	£1,072
INMB at £30,000 per QALY gained			£1,909	£1,749	£1,916	£2,039	£2,036	£2,013

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 44 Scenario analysis using data from Massie 2018 to inform accuracy of NightOwl

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,376	£7,572	£8,108	£8,043	£8,315	£8,195	£8,322	£8,346
Total QALYs	14.142	14.050	14.135	14.124	14.153	14.139	14.138	14.138
Compared to respiratory polygraphy								
Incremental cost			-£267	-£333	-£61	-£181	-£53	-£30
Incremental QALYs			-0.006	-0.017	0.011	-0.003	-0.003	-0.003
ICER (£ per QALY gained)			£42,211[a]	£19,128[a]	Dominant	£67,426[a]	£16,172[a]	£9,163[a]
INMB at £20,000 per QALY gained			£141	-£15	£281	£127	-£13	-£36
INMB at £30,000 per QALY gained			£77	-£189	£391	£100	-£46	-£69
Compared to oximetry								
Incremental cost			£536	£471	£743	£623	£750	£773
Incremental QALYs			0.085	0.074	0.102	0.089	0.088	0.088
ICER (£ per QALY gained)			£6,302	£6,360	£7,253	£7,020	£8,515	£8,777
INMB at £20,000 per QALY gained			£1,165	£1,009	£1,305	£1,152	£1,012	£989
INMB at £30,000 per QALY gained			£2,016	£1,749	£2,329	£2,039	£1,893	£1,870

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)
Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 45 Scenario analysis using data from van Pee 2020 to inform accuracy of NightOwl

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,376	£7,572	£8,108	£8,043	£8,268	£8,195	£8,322	£8,346
Total QALYs	14.142	14.050	14.135	14.124	14.141	14.139	14.138	14.138
Compared to respiratory polygraphy								
Incremental cost			-£267	-£333	-£108	-£181	-£53	-£30
Incremental QALYs			-0.006	-0.017	-0.001	-0.003	-0.003	-0.003
ICER (£ per QALY gained)			£42,211[a]	£19,128[a]	£106,275[a]	£67,426[a]	£16,172[a]	£9,163[a]
INMB at £20,000 per QALY gained			£141	-£15	£88	£127	-£13	-£36
INMB at £30,000 per QALY gained			£77	-£189	£77	£100	-£46	-£69
Compared to oximetry								
Incremental cost			£536	£471	£696	£623	£750	£773
Incremental QALYs			0.085	0.074	0.090	0.089	0.088	0.088
ICER (£ per QALY gained)			£6,302	£6,360	£7,696	£7,020	£8,515	£8,777
INMB at £20,000 per QALY gained			£1,165	£1,009	£1,112	£1,152	£1,012	£989
INMB at £30,000 per QALY gained			£2,016	£1,749	£2,016	£2,039	£1,893	£1,870

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Other scenario analyses that affect results

When we assume the failure rate for respiratory polygraphy of [redacted] as reported by Alsaif et al (2023), INMB estimates increase for all novel devices relative to RP, and WatchPAT ONE is the only device that is not estimated to be cost-effective compared to respiratory polygraphy at the £20,000 threshold (Table 46).

In the base case analysis, it is assumed that 20% of people with mild OSAHS would receive CPAP. This assumption was based on clinical opinion. Due to variation in clinical opinion on what this proportion would be in practice, we conducted an additional scenario analysis, where we assumed that 75% of people diagnosed with mild OSAHS are treated with CPAP, but that 50% are non-compliant within the first year. In this scenario all novel devices are estimated to be cost-effective compared to respiratory polygraphy at £20,000 and £30,000 per QALY gained. This is driven by more gains in QALYs being estimated in this scenario due to more people with mild OSAHS being treated with CPAP (than is assumed in the base case). People receiving CPAP are assumed to have higher utility than those receiving conservative management.

In scenario analyses where it is assumed that there are no impacts on RTA risk from CPAP treatment, Brizzy and NightOwl are considered cost-effective compared to respiratory polygraphy at the £20,000 WTP threshold, and similarly when we assume no treatment impacts on CV events or RTA (see Table 48 and Table 49). These improvements in cost-effectiveness for Brizzy and NightOwl are also seen in the scenario analysis when greater costs are associated with people who are incorrectly diagnosed as having OSAHS (false positives), see Table 50. In this scenario treatment costs are assumed to be incurred for two (as opposed to one year in the base case), but this treatment has no health benefit to the individual.

When a 10 year time horizon is used, all novel devices are estimated to be cost-effective compared to respiratory polygraphy at both thresholds (Table 51).

Table 46 Scenario analysis assuming failure rate for respiratory polygraphy from the Alsaif et al study

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	[redacted]	£7,572	£8,108	£8,043	£8,202	£8,195	£8,322	£8,346
Total QALYs	[redacted]	14.050	14.135	14.124	14.133	14.139	14.138	14.138
Compared to respiratory polygraphy								
Incremental cost			[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Incremental QALYs			[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ICER (£ per QALY gained)			£46,267 [a]	£20,605 [a]	£22,072 [a]	£77,015 [a]	£23,960 [a]	£16,951 [a]
INMB at £20,000 per QALY gained			£166	£11	£19	£153	£13	-£10
INMB at £30,000 per QALY gained			£103	-£163	-£72	£126	-£20	-£43
Compared to oximetry								
Incremental cost			£536	£471	£630	£623	£750	£773
Incremental QALYs			0.085	0.074	0.082	0.089	0.088	0.088
ICER (£ per QALY gained)			£6,302	£6,360	£7,646	£7,020	£8,515	£8,777
INMB at £20,000 per QALY gained			£1,165	£1,009	£1,017	£1,152	£1,012	£989
INMB at £30,000 per QALY gained			£2,016	£1,749	£1,841	£2,039	£1,893	£1,870

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 47 Scenario analysis with 75% of people with mild OSAHS offered CPAP, but 50% drop-out in first year

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,573	£8,041	£8,329	£8,333	£8,416	£8,381	£8,479	£8,502
Total QALYs	14.164	14.129	14.159	14.159	14.161	14.162	14.162	14.162
Compared to respiratory polygraphy								
Incremental cost			-£244	-£239	-£157	-£192	-£94	-£71
Incremental QALYs			-0.004	-0.004	-0.003	-0.002	-0.001	-0.001
ICER (£ per QALY gained)			£56,527[a]	£55,460[a]	£52,741[a]	£100,130[a]	£63,305[a]	£47,701[a]
INMB at £20,000 per QALY gained			£157	£153	£97	£154	£64	£41
INMB at £30,000 per QALY gained			£114	£110	£68	£134	£49	£26
Compared to oximetry								
Incremental cost			£288	£292	£375	£340	£438	£461
Incremental QALYs			0.030	0.030	0.032	0.033	0.033	0.033
ICER (£ per QALY gained)			£9,474	£9,612	£11,803	£10,354	£13,171	£13,867
INMB at £20,000 per QALY gained			£320	£316	£260	£317	£227	£204
INMB at £30,000 per QALY gained			£624	£620	£578	£645	£560	£536

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 48 Scenario analysis with no treatment impacts on RTA risk

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,624	£7,750	£8,346	£8,273	£8,445	£8,441	£8,577	£8,600
Total QALYs	14.127	14.040	14.121	14.110	14.118	14.124	14.123	14.123
Compared to respiratory polygraphy								
Incremental cost			-£278	-£350	-£179	-£183	-£46	-£23
Incremental QALYs			-0.006	-0.016	-0.009	-0.003	-0.004	-0.004
ICER (£ per QALY gained)			£49,129 [a]	£21,470 [a]	£20,421 [a]	£72,160 [a]	£12,342 [a]	£6,195 [a]
INMB at £20,000 per QALY gained			£165	£24	£4	£132	-£29	-£52
INMB at £30,000 per QALY gained			£108	-£139	-£84	£107	-£66	-£90
Compared to oximetry								
Incremental cost			£596	£524	£695	£691	£827	£851
Incremental QALYs			0.081	0.071	0.078	0.085	0.083	0.083
ICER (£ per QALY gained)			£7,312	£7,393	£8,870	£8,167	£9,925	£10,202
INMB at £20,000 per QALY gained			£1,034	£893	£872	£1,001	£840	£817
INMB at £30,000 per QALY gained			£1,848	£1,601	£1,656	£1,847	£1,674	£1,651

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 49 Scenario analysis with no treatment impacts on CV events or RTAs

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,649	£7,770	£8,371	£8,298	£8,470	£8,466	£8,602	£8,625
Total QALYs	14.121	14.035	14.115	14.105	14.112	14.118	14.117	14.117
Compared to respiratory polygraphy								
Incremental cost			-£278	-£351	-£179	-£183	-£48	-£24
Incremental QALYs			-0.006	-0.016	-0.009	-0.003	-0.004	-0.004
ICER (£ per QALY gained)			£48,858 [a]	£21,748 [a]	£20,936 [a]	£73,185 [a]	£13,523 [a]	£6,943 [a]
INMB at £20,000 per QALY gained			£164	£28	£8	£133	-£23	-£46
INMB at £30,000 per QALY gained			£107	-£133	-£78	£108	-£58	-£81
Compared to oximetry								
Incremental cost			£602	£529	£701	£697	£832	£855
Incremental QALYs			0.080	0.070	0.077	0.083	0.082	0.082
ICER (£ per QALY gained)			£7,515	£7,594	£9,073	£8,367	£10,118	£10,399
INMB at £20,000 per QALY gained			£1,000	£864	£844	£969	£813	£790
INMB at £30,000 per QALY gained			£1,801	£1,560	£1,616	£1,801	£1,635	£1,612

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 50 Scenario analysis assuming false positives continue treatment for 2 years (with no health benefit)

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£8,424	£7,584	£8,121	£8,043	£8,240	£8,205	£8,493	£8,516
Total QALYs	14.142	14.050	14.135	14.124	14.133	14.139	14.138	14.138
Compared to respiratory polygraphy								
Incremental cost			-£303	-£381	-£184	-£219	£69	£92
Incremental QALYs			-0.006	-0.017	-0.009	-0.003	-0.003	-0.003
ICER (£ per QALY gained)			£47,873 [a]	£21,904 [a]	£20,354 [a]	£81,887 [a]	Dominated	Dominated
INMB at £20,000 per QALY gained			£177	£33	£3	£166	-£135	-£158
INMB at £30,000 per QALY gained			£113	-£141	-£87	£139	-£168	-£191
Compared to oximetry								
Incremental cost			£537	£459	£656	£620	£908	£932
Incremental QALYs			0.085	0.074	0.082	0.089	0.088	0.088
ICER (£ per QALY gained)			£6,308	£6,200	£7,965	£6,994	£10,312	£10,574
INMB at £20,000 per QALY gained			£1,165	£1,021	£991	£1,154	£854	£830
INMB at £30,000 per QALY gained			£2,015	£1,761	£1,815	£2,041	£1,735	£1,711

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

Table 51 Scenario analysis with a 10 year time horizon

Setting	Comparators		Clinic					
	Respiratory polygraphy	Oximetry	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Total costs	£2,617	£2,052	£2,385	£2,343	£2,461	£2,445	£2,543	£2,567
Total QALYs	7.251	7.206	7.248	7.243	7.247	7.250	7.250	7.250
Compared to respiratory polygraphy								
Incremental cost			-£233	-£275	-£156	-£173	-£74	-£51
Incremental QALYs			-0.003	-0.009	-0.004	-0.001	-0.001	-0.001
ICER (£ per QALY gained)			£69,630 [a]	£31,161 [a]	£35,133 [a]	£128,006 [a]	£54,103 [a]	£37,199 [a]
INMB at £20,000 per QALY gained			£166	£98	£67	£146	£47	£24
INMB at £30,000 per QALY gained			£133	£10	£23	£132	£33	£10
Compared to oximetry								
Incremental cost			£332	£290	£409	£392	£491	£514
Incremental QALYs			0.042	0.037	0.041	0.044	0.044	0.044
ICER (£ per QALY gained)			£7,833	£7,858	£9,892	£8,834	£11,062	£11,583
INMB at £20,000 per QALY gained			£516	£448	£417	£496	£397	£374
INMB at £30,000 per QALY gained			£940	£818	£831	£940	£841	£817

[a] ICER lies in South-West quadrant of cost-effectiveness plane (intervention less costly and less effective than the comparator)

Abbreviations: ICER incremental cost-effectiveness ratio; INMB incremental net monetary benefit; QALYs quality-adjusted life-years

5.10.3 One-way sensitivity analyses

We undertook one-way sensitivity analyses, with each parameter increased or decreased in turn. The aim of these analyses is to assess the impact each parameter has on the cost-effectiveness results. The base case values and range of variation for each parameter is shown in Appendix 9b Table 78. The parameter ranges were either the 95% confidence intervals for the parameter values, an arbitrary increase/decrease of 20%, or high and low parameter values from alternative sources. Tornado plots, showing the top 20 parameters having the most impact, for each novel device compared to oximetry and respiratory polygraphy at £20,000 and £30,000 per QALY gained, are shown in appendix 9b (Figure 12 - Figure 17). Across all novel devices, the parameters having the most impact in one-way sensitivity analyses were:

- Utilities for mild and moderate OSAHS
- Sensitivity and specificity estimates for the novel devices
- Sensitivity and specificity estimates for the comparators, and
- Prevalence parameters

The lower utility values assumed for mild and moderate OSAHS in the one-way sensitivity analyses were taken from the study by Skirko [55] (0.60 for mild OSA and 0.61 for moderate OSA), while the upper values were based on a 20% increase of the mean value: 0.97 for mild OSA and 0.92 for moderate OSA. Thus, the upper and lower limits represent a relatively large range of values, and so it is not surprising that they have a large impact on the results.

When comparing all of the novel devices to respiratory polygraphy, changes in the sensitivity of respiratory polygraphy or the novel device at the high diagnostic cut-off would lead to changes in the results compared to the base case analysis (see Figure 12 - Figure 17). For example, the INMB for Brizzy vs respiratory polygraphy at £20,000 per QALY gained could range from approximately -£360 to £180 depending on the input value for the sensitivity of Brizzy at the high diagnostic cut-off (Figure 13c).

For WatchPAT 300 and WatchPAT ONE, the specificity at the high diagnostic cut-offs for the novel devices and respiratory polygraphy also has an impact on the results (see Figure 16-Figure 17). This can be explained by the low specificity leading to people with mild OSAHS being diagnosed as having moderate-severe OSAHS. Those correctly identified as having mild OSA may receive conservative management, CPAP or MAD. Those with mild OSA misdiagnosed with moderate or severe are only treated with CPAP and MAD. Note that only CPAP and MAD impact on utility in the truly mild group. With a higher specificity, the number

of people with mild OSA misdiagnosed as having mod/severe OSA will decrease. Which means that fewer people with mild OSA have CPAP or MAD, which reduces their QALYs (and costs). This pattern is also reflected in the scenario analysis where 75% of those with mild receive CPAP.

For all of the novel devices and respiratory polygraphy, the upper and lower values used in these one-way sensitivity analyses are based on the 95%CIs from the sensitivity and specificity estimates. Thus, they represent reasonable extremes to assess. When comparing the novel devices to oximetry, the results are generally consistent, in that the novel devices are seen to be more cost-effective than oximetry, regardless of the parameter inputs.

5.11 EAG discussion of economic evaluation for children

In this section, we discuss the challenges associated with estimating the cost-effectiveness of the novel devices for diagnosis of OSAHS in children, and we outline a potential model structure and set of parameters.

5.11.1 Challenges in modelling for children

Population and clinical pathway

BTS guidelines for diagnosing and monitoring paediatric sleep disordered breathing were published in June 2023 (a month after the final scope for this evaluation).[9] These guidelines recommend diagnostic pathways for children with and without comorbidities.

For children with no comorbidities who are suspected of having moderate or severe sleep disordered breathing, the BTS guidelines recommend that use of sleep questionnaires and clinical examination may be adequate to inform management. However, if there is inconsistency between the findings of the questionnaires/clinical examination and the child's symptoms, sleep studies are recommended. The guidelines are explicit that questionnaires should not be used for the diagnosis of children where mild sleep disordered breathing is suspected. Thus, the guidelines indicate that for children with no comorbidities, diagnostic sleep studies are only relevant where mild sleep disordered breathing is suspected, or where there is inconsistency between symptoms and findings from sleep questionnaires and clinical examination.

For children with comorbidities, sleep questionnaires are not recommended, due to a lack of available evidence. The authors of the BTS guidelines highlight a lack of evidence for the use of pulse oximetry in children with comorbidities and recommend that cardiorespiratory

sleep studies can be considered in children with neuromuscular disorders or Down Syndrome (where there was some evidence), but also in children with other comorbidities.

The BTS guidelines recommend that both home pulse oximetry and home cardiorespiratory sleep studies can be considered for children with and without comorbidities “where the patients and/or carers are deemed appropriate for implementing a home sleep study”. If the data are insufficient to make a diagnosis, consideration should be given to conducting an inpatient sleep study. Throughout the BTS guidelines, recommendations warn about the use of AHI alone for determination of presence and absence of sleep disordered breathing and its severity.

In practice, the paediatric experts we have spoken to report that only pulse oximetry is currently considered for home-based testing in children and cardiorespiratory sleep studies are conducted within the hospital setting. However, a stakeholder has noted that practice varies, and some centres do conduct RP for children in the home setting, with good success rates. This does pose the question of whether novel home-based devices could be effective and cost-effective alternatives to home-based oximetry and hospital-based cardiorespiratory sleep studies.

Thus, there is some uncertainty around the current clinical pathways and the way in which novel home-based devices would fit in to them. It is clear that any modelling would need to reflect important differences according to whether a child had comorbidities or not, and the type of comorbidity. The simplest pathway would be in identification of OSAHS in children without comorbidities, where the use of novel devices as an alternative to pulse oximetry at home could be explored. For children with comorbidities, the case for use of novel home-based devices is still insufficiently developed to define a new potential clinical pathway, particularly as the nature of the diagnosis extends far beyond just ruling OSAHS in or out.

Lack of good quality accuracy data on novel devices in children

Irrespective of the greater complexity of the clinical pathway in children, the diagnostic accuracy (sensitivity and specificity) of the novel home-based devices will be a critical component of the use case. There is consensus that evidence from adults is not automatically generalisable to children because of considerations of anatomy and capacity. Thus, modelling of cost-effectiveness in children is absolutely dependent on credible accuracy data being available in children. Further, the diagnostic cut-offs for mild, moderate and severe OSAHS are different between children and adults, again challenging the transferability of adult data to children.

Our systematic review of the clinical effectiveness evidence for novel devices in children suggests that there has been little research activity in this population to date. Two studies met our inclusion criteria, although only one reported relevant diagnostic accuracy data (see section 4.3 above). Martinot 2022 [35] reported diagnostic accuracy for the Sunrise device compared with PSG for 140 children aged 3-17 years without significant chronic medical conditions. The second study, Martinot 2015,[42] compared the Brizzy device with PSG in 33 children with suspected OSA prior to tonsillectomy, but did not report diagnostic accuracy results. In both studies, the novel device was used simultaneously with sleep laboratory-based PSG, rather than in a home setting. This may be a bigger problem in children than in adults, as the performance of the novel devices when used in a hospital setting may be quite different than in the unattended home setting for children. The absence of relevant test accuracy data is a barrier to credible modelling of the cost-effectiveness of home-based devices for OSAHS in children at the present time.

Information about another study in progress (NCT04031950) examining the diagnostic accuracy of AcuPebble in a paediatric population was included with the Acurable company submission.[37 71 72] This included academic in confidence data on the patient population and a preliminary report on diagnostic performance from one of the two UK centres (see section 4.3 above). Information about the study methodology is currently limited, but it does have the potential to inform a future economic model.

Effective treatment for children

While adults who are diagnosed with OSA would generally be offered CPAP treatment, with good consensus about evidence for effectiveness, the first line treatment for children (who do not have any comorbidities) would be adenotonsillectomy (BTS guidelines).[9] The Childhood Adenotonsillectomy Trial (CHAT) was a pivotal RCT that compared adenotonsillectomy to watchful waiting in 464 children with OSA aged 5-9 years with AHI 2-30 events/hour without prolonged oxyhaemoglobin desaturation, and who were deemed suitable for adenotonsillectomy.[113 114] After 7 months, the primary outcome of attention and executive function, as measured by the Developmental Neuropsychological Assessment, was improved in the adenotonsillectomy arm, but this improvement was not statistically significantly different to that in the watchful waiting arm. However, statistically significant improvements in some secondary outcomes, including AHI, teacher and parent/carer-reported measures of behaviour and quality of life (as measured by the Pediatric Quality of Life Inventory (PedsQL), were observed.

Due to the lack of a statistically significant effect in the primary outcome this trial has raised many questions about who is likely to benefit from adenotonsillectomy. It is noteworthy that polysomnographic abnormalities resolved without intervention in almost half of children (46%) in the watchful waiting arm by 7 months.

One of our experts indicated that children with moderate/severe OSA are referred to ENT surgeons for assessment, but that a proportion are not found to have any obstruction and so lifestyle modifications are discussed, but if that does not work CPAP is offered. For children with mild OSA, there is uncertainty over the effectiveness of treatments, it is often unclear how to proceed. Incorporating the effectiveness of treatment, particularly for mild OSA, thus represents a further challenge to modelling the cost-effectiveness in children. The likelihood of resolving this uncertainty over treatment effectiveness for future modelling may be low.

Longer-term impacts

In adults, evidence indicates that untreated OSA can lead to increased CV and RTA risks. The impact of treatment, particularly CPAP, is to lower these risks. In children, the evidence on the longer-term impacts of untreated OSA is less clear. Experts point to impacts on behaviour, cognition, and educational attainment, but there is uncertainty over the extent and reversibility of these adverse effects and over the longer-term impacts on CV risks and the incidence of accidents.[9]

The uncertainty about the extent and reversibility of longer-term impacts is another challenge to creating a useful model. A further complexity is that some of these impacts would not be included in NICE's preferred perspective of NHS and PSS.

Measurement of utility

The measurement of utility in children comes with known challenges. The NICE Reference Case does not recommend a particular instrument for assessing HRQoL (utility) in children.[73]

We identified one relevant utility study from our systematic review of HRQoL (see section 5.2.2 above). Sakki et al. 2021 [57] reported utility for children with sleep-disordered breathing before and after tonsillectomy from a prospective cohort study in a Finnish hospital. They used the 17D preference-based utility instrument (adapted for children from the 15D instrument), with a Finnish value set. We have not found any relevant utility estimates with a UK general population valuation. We note that the PedsQL has been used in a number of studies in the area of OSA in children.[115-117] However, although mapping from the PedsQL to the EQ-5D is available, it is the adult tariff for EQ-5D that is used.[118 119] A more appropriate

approach to obtaining estimates of utility in children would be to use the CHU9D (Child Health Utility 9D Index), or mapping to that from the PedsQL. We note some controversy over the use of parent-proxies for completing (and valuing) HRQoL questions for children. We identified one study that found little correlation between PedsQL scores for mild OSA between parents and children.[120]

Given the above and after discussion with paediatric specialist committee members and NICE technical staff, we concluded that the best way that we could contribute to an understanding of the potential cost-effectiveness of novel home-based devices in children is to suggest possible model structures. Further we have assessed the availability of key parameters which would be required as a way of informing further research needed to improve our understanding of the potential cost-effectiveness in children.

5.11.2 Potential model structures for children without comorbidities

The BTS guideline recommends separate pathways diagnosing OSAHS in children with and without comorbidities.[9] In this section we suggest a possible model structure for children without comorbidities. We discuss the potential for a model for children with comorbidities in section 5.11.3.

Decision problem

The NICE scope for the current assessment includes children and young people aged 16 and under. None of the novel devices under review are currently indicated for children aged under 2 years (section 4.3.1 above). Unless this changes, it would be appropriate to restrict the age range for a children's model to 2-16 years.

The BTS guideline for diagnosing paediatric sleep-disordered breathing indicates that home-based oximetry or cardiorespiratory sleep studies may be considered for children without comorbidities who are being considered for adenotonsillectomy but for whom there is clinical uncertainty over the diagnosis of OSAHS. Based on the BTS pathways, an appropriate population for this decision problem is:

> Children aged 2-16 years without comorbidities; no strong clinical suspicion of OSA or inconsistent results between clinical examination and sleep questionnaire; and who are potential candidates for surgery and suitable for a home sleep study. (BTS Diagnosis pathway 1 and OSA pathway)

Interventions of interest in this population are the named novel devices indicated for use in children, with available diagnostic accuracy data in an appropriate population.

Comparators should reflect current guidance and routine practice. The BTS guideline recommends that a home sleep study can be used when appropriate for the patient and carer, and that either pulse oximetry or RP may be used; followed by a repeat test (same type as first test or hospital PSG) if data are inadequate or the result is inconsistent with clinical assessment (BTS Home monitoring pathway). As noted above, paediatric experts advised the EAG that in practice cardiorespiratory sleep studies for children are conducted in a hospital setting, although a stakeholder has noted that some centres do provide these tests at home. This suggests that it might be appropriate to include hospital respiratory polygraphy alongside home oximetry and home respiratory polygraphy as comparators in an analysis for the above population.

Key outcomes in this population are 'appropriate' referral for surgery (defined according to true presence and severity of OSAHS), symptomatic relief of OSAHS and associated impacts on health-related quality of life, costs of diagnosis and treatment. If evidence were available, it would be appropriate to include effects on utility for carers and other family members in a reference case analysis.

There may be a benefit of accurate pre-operative sleep assessment to assess and mitigate the risk of surgical complications, although the validity of this is uncertain (BTS guideline).[9]

The persistence of the effects of surgery on quality of life, symptoms and behaviour is uncertain: in the CHAT trial, PSG parameters had normalised by 7 months.[113 114] There is also insufficient evidence to model the extent of effects on neurological development, educational attainment, and potential long-term impact future employment and earnings, and their reversibility with treatment.[9] We therefore suggest that a short time horizon is adopted (1-3 years, say).

Illustrative model structure

Figure 9 below illustrates a possible model structure. As with the adult model (Figure 3), this would start with a decision tree stratifying the population according to the true prevalence and severity of OSA and mapping out the diagnostic process and allocation of patients to treatments.

The mild, moderate and severe OSA categories are defined as in Table 6 of the BTS guideline (based on AHI criteria for OSA, labelled as OAHI):

- Mild OSAHS: OAHI 1 or more to less than 5
- Moderate OSAHS: OAHI 5 or more to less than 10
- Severe OSAHS: OAHI 10 or more

The BTS guideline development group noted that evidence linking AHI values in children to functional outcomes is sparse and advised caution on using AHI alone for decision making.

The decision tree follows the BTS home monitoring pathway, with a first test (one of the intervention or comparator sleep studies), a second test if required due to inadequate data collection or inconsistency between the first test result and the clinical assessment.

Assumptions would be needed to specify the treatments that would be offered depending on the diagnostic results. For this model the main treatment of interest is surgery. A 'watchful waiting' option could be included to reflect the comparator arm of the CHAT trial,[113 114] for the situation where there is uncertainty over the appropriateness of early surgery (e.g. for mild and possibly moderate OSA). A proportion of the cohort could be allocated to CPAP and other treatment options.

The final part of the model would extrapolate outcomes and costs over a defined time horizon. In the adult model, we used a Markov model to predict long-term impacts on cardiovascular outcome and road traffic accidents (Figure 4 above). It is less clear that a long-term extrapolation is needed for the children's model, due to uncertainty over the persistence of the effects of childhood OSAHS and whether surgery has a lasting benefit. A simple extrapolation based on assumptions about the waning of effects from the 7 month follow up in the CHAT trial could be used.

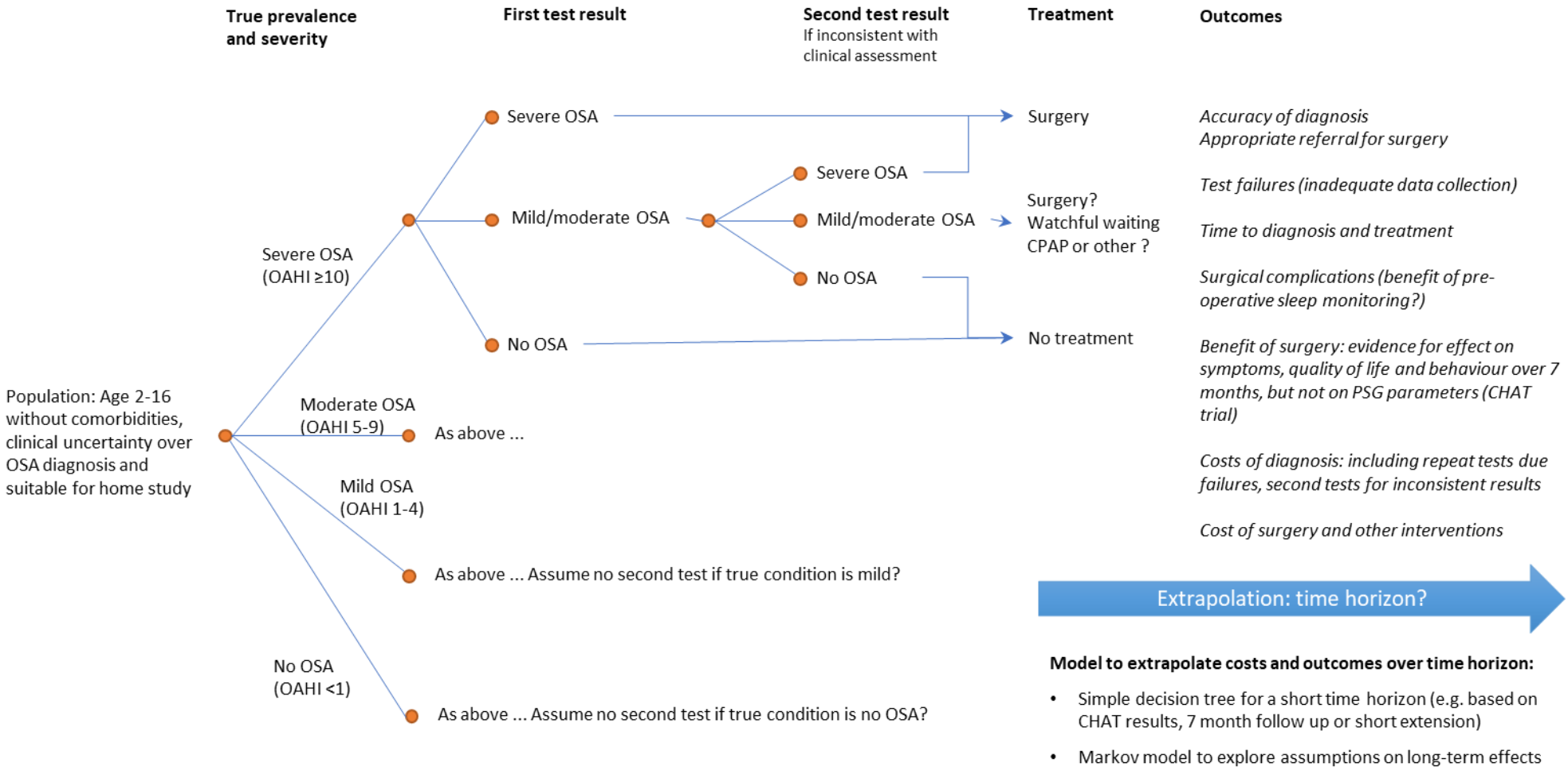


Figure 9 Illustrative model structure for children without comorbidities

Abbreviations: OAHI Apnoea Hypopnoea Index for obstructive sleep apnoea

Parameter requirements

The model in Figure 9 would require evidence or assumptions for the following sets of parameters:

Epidemiology

- Prevalence of OSAHS in children aged 2 to 18 years of age
- Split by severity (mild, moderate, severe)
- Adverse health outcomes: the BTS guideline cites evidence that OSA in children is associated with adverse respiratory and neurocognitive outcomes, with evidence of a dose effect. [9] However, they note debate as to the reversibility of these effects; and suggest that the increased mortality in children with severe sleep disordered breathing is likely to be related to underlying conditions.

Diagnostic performance (accuracy, failure rates)

- Home based pulse oximetry and RP (BTS Online Appendix 8) [9]
- Hospital or home based oximetry and RP (BTS Online Appendix 2) [9]
- Novel home devices: Sunrise device, Martinot et al (2022)[121]; and the AcuPebble study in progress (NCT04031950 2019) [37 71 72] (see section 4.3 above)

Treatment use and effects

- Distribution of treatments by true severity/diagnostic results: This is not explicit in current guidelines (BTS, ENTUK/RCS).[9 122] Assumptions and estimates from clinical experts would probably be required?
- Effects of surgical versus non-surgical treatment: CHAT RCT. [113 114] The Cochrane review (Vennekamp et al 2015 [123]) identified two other small trials, that are not directly relevant to the above decision problem.
- Adverse effects of surgery: unscheduled admissions to intensive or high-dependency care after ENT surgery (BTS guideline Online Appendix 11).[9] Very small mortality risk associated with adenotonsillectomy (Cottrell 2020)[124]

Utilities

- Health-related quality of life (17D utility index score, Finnish value set) for children with sleep-disordered breathing before and after tonsillectomy (Sakki et al. 2021). [57] Prospective cohort study in a Finnish hospital (see Section 5.2.2).

Resource use and costs

- Diagnosis: bottom-up costing for novel devices and comparators in home and hospital setting. Assumptions and unit costs in the EAG cost estimates for adults (5.7.12 to 5.7.14 above) could be adapted for a paediatric population.
- Costs for surgery, including pre- and post-operative care, assessment and monitoring based on assumptions about resource use and HRG/PSSRU unit cost data. This would require expert input.
- Sakki et al. also report hospital costs and patient-reported primary care, hospital and laboratory costs from 3 months prior to surgery and 12 months after, costs in Finnish unit costs (Sakki et al. 2021).[57] These results are unlikely to be transferable to a UK context, but they demonstrate that a large proportion of public sector hospital costs relate to the cost of surgery itself.

5.11.3 Potential model structures for children with comorbidities

The BTS guideline includes a diagnosis pathway for OSA for children with comorbidities, and they recommend use of home-based OSA testing for children with and without comorbidities, when it is suitable for the patient and carers.[9]

However, we suggest that it would not be appropriate to use the same model structure and parameters for children with and without comorbidities, as recommendations for diagnosis and treatment differ between these subgroups. In particular, for children without comorbidities, the main treatment for OSAHS is adenotonsillectomy (BTS Diagnostic Pathway 1), whereas interventions for children with comorbidities differ according to clinical context and symptoms (BTS Diagnostic Pathway 2). [9] A single model structure would therefore not be suitable for all children with comorbidities, as a large number of conditions are associated with OSAHS in children, and different treatments are suitable for different conditions (including airway assessment and surgery, secretion management and initiation and adjustment of positive airway pressure).

It would be possible to adapt the above model structure to assessment of home-based testing for children with OSA and defined subgroups of comorbidities, for whom the diagnosis and treatment pathway is similar. One particular group of interest might be children with conditions that place them at high risk of recurrent respiratory illness, as OSA can further increase this risk and associated hospital admissions and mortality. [9]

6 DISCUSSION

6.1 Discussion of principal findings

6.1.1 Clinical effectiveness evidence

The systematic review of clinical effectiveness identified limited evidence on the performance of novel devices. For some devices only a single study was identified. A large proportion of studies evaluated novel devices in the clinic setting, rather than in the home, thus limiting their generalisability to clinical practice.

The sensitivity and specificity estimates vary across the studies and also within studies at different severity cut-offs. Sensitivity was generally high, in the range 80 to 100%, and fell below 80% in just two studies (at one cut-off each from (Martinot (2017) and Kelly et al (2022))). In contrast, specificity was more variable with estimates ranging from 25% to 100%, with more estimates in the 70% to 80% range than was the case for sensitivity. Confidence intervals in some studies were very wide indicating uncertainty. We urge caution in making inferences about the relative superiority in diagnostic performance between the novel devices.

Limited data are available for other outcomes relevant to the decision problem such as estimates of time to making a diagnosis, and impact on clinical decision making and resources. Some of the suggested benefits of novel home test devices are not necessarily founded in empirical evidence.

Device test failures were reported by many studies, with rates generally less than 10%. It should be acknowledged that some of the factors contributing to failed tests were not anticipated by the study investigators and, with the benefit of hindsight, were preventable. The expectation is that the learning from these instances will have prompted necessary changes to testing protocols, device features and user instructions to avoid similar failures occurring again. If this is the case then novel device failure rates in clinical practice would be lower than those reported in the studies, all other factors being equal.

6.1.2 Cost effectiveness evidence

We adapted an existing economic model used to inform recent NICE guidelines on the diagnosis and management of OSAHS in people ≥ 16 years of age.[52] The adapted model consists of a decision tree to capture the diagnostic outcomes associated with six novel devices and two comparators (home respiratory polygraphy and oximetry). The decision tree

has a time horizon of 12 months to capture any delays to the start of treatment (should treatment be offered). A lifetime Markov model is used to estimate the longer-term impacts associated with the performance of the devices. It models the risks of cardiovascular events and RTAs for people with OSAHS and includes death from other causes for the total cohort.

In the base case analysis, all six novel devices are estimated to be less costly than respiratory polygraphy, but they are also associated with a small estimated reduction in QALYs. For AcuPebble and Sunrise compared to respiratory polygraphy, the reduction in QALYs is considered cost-effective compared to the reduction in costs (i.e. INMB > £0 at the £20,000 and £30,000 per QALY thresholds). Compared to oximetry, all novel devices have a positive INMB at both £20,000 and £30,000 per QALY thresholds.

However, it is important to recognise the high level of uncertainty over the cost-effectiveness results. This is apparent from the probabilistic and scenario analyses. In the probabilistic base case analysis, there are wide and overlapping confidence ranges for the incremental costs and QALYs for each novel device compared with oximetry, which is more pronounced for the comparisons with respiratory polygraphy. For example, the incremental costs for WatchPAT 300 compared with respiratory polygraphy range from -£298 to £235 and the incremental QALYs range from -0.040 to 0.033. This uncertainty is reflected in wide confidence ranges around the INMBs for the novel devices, for example, comparing Sunrise to respiratory polygraphy, the INMB at £20,000 per QALY gained ranges from -£238 to £572.

These results are sensitive to a number of assumptions, as the scenario analyses indicate, including the data source used to estimate the diagnostic performance and failure rates associated with respiratory polygraphy, the proportion of people diagnosed with mild OSAHS who are treated with CPAP, alternative parameterisation of the decision tree (using 4x4 contingency table data), and the impacts associated with false positives. See section 6.3 below for a discussion of the key sources of uncertainty and their effect on cost-effectiveness results. Moreover the data used in the base case analysis to inform the accuracy estimates for novel devices are all derived from a clinical setting, with three based on post-hoc optimisation of thresholds, which is likely to overestimate the accuracy of the devices.

6.2 Strengths and limitations of the assessment

6.2.1 Strengths

The cost-effectiveness model is adapted from one that was used to inform recent NICE guidance on the diagnosis and management of obstructive sleep apnoea/ hypopnoea syndrome and obesity hypoventilation syndrome in over 16s (NG202).[52] The NG202 economic model was developed in consultation with the Guideline Committee, and was itself adapted from the model developed to inform the NICE appraisal on CPAP treatment for people with OSAHS (TA139).[54] We believe that the attention that versions of this model have received, by experts in the field during development and in consultation processes is a strength. We updated parameter values from those used in the TA139 and NG202 models where we could identify more recent, relevant data of better quality. The choice of data for the model parameters was informed by our systematic review of clinical and diagnostic assessments of the novel devices, and economic evidence on cost-effectiveness, resource use and costs and health-related quality of life. We also conducted targeted reviews for other key model parameters. Throughout our adaptation of the model, we consulted with experts, especially on the validity of base case and scenario analysis assumptions

6.2.2 Limitations

The cost-effectiveness analysis is limited by the availability and quality of data for many of the model components. This included:

- Limited diagnostic accuracy data for novel devices evaluated in the home, rather than the clinic – data from a home setting were only available for AcuPebble and Sunrise, and as the AcuPebble study (Devani et al) did not use PSG as the reference standard, and the Sunrise study (Kelly 2022) was very small, neither were used in our base case.
- Lack of evidence on current versions of devices, e.g. WatchPAT 300 and ONE, although the manufacturer has reported that these versions of the WatchPAT device produce identical signals and use the same algorithm as the previous 200U version.
- Inconsistency in reference standards used across devices – in Devani et al (2021) the reference standard was home RP, in Kelly et al (2022) it was home PSG.
- Post-hoc optimisation of diagnostic thresholds within accuracy studies (such as for Sunrise in Pepin et al (2020) and Kelly et al (2022).
- Lack of data on health-related quality of life (utility) associated with different severities of OSAHS.

- Mixed evidence on the long-term impact of OSAHS treatment on risks of cardiovascular events and road traffic accidents, and the impacts of untreated OSAHS on other long-term consequences.
- Limited evidence to update parameter values in NG202 that are informed by studies conducted many years ago.
- Limited data on treatment compliance over time.

These limitations meant that a number of assumptions had to be made. For example, evidence for differences in time to treatment for the novel devices compared to current practice is limited. [redacted]. Until we have further data on the reduction in time to treatment, the magnitude of such a benefit is unclear. Furthermore, we had to make choices on what source of accuracy data to use in the model for the comparators and novel devices (such as for NightOwl). We were also limited by a lack of access to NHS Supply Chain data to update device and treatment costs.

Our analysis does not capture any potential benefits for patients in terms of greater convenience and acceptability with the novel devices, compared with current home or clinic based assessment. This is clearly important to patients and their families, but as the sleep studies occur over just a few nights at most, any attempts to incorporate acceptability in terms of QALYs are not likely to impact on the model results. Acceptability may be captured indirectly via the failure rates for the novel devices (should they be lower than those for the comparators).

Neither does our model account for a potential impact of comorbidities and contraindications to the use of some devices. There is variability of the suitability of devices for people with comorbidities, and a lack of evidence on the likely size of populations excluded from use of some devices.

A key limitation is that our economic model is only relevant to people over the age of 16. We have not been able to estimate the cost-effectiveness of the novel devices for home-based diagnostic assessment of OSAHS in people aged 16 and under. Data on the diagnostic accuracy of the novel devices in children is sparse: with only one published study reporting sensitivity and specificity (Martinot 2022 [35] for the Sunrise device, which used post-hoc optimisation); and one study in progress with interim results that are academic in confidence

(AcuPebble trial)[37 71 72] There are also significant challenges including heterogeneity of the paediatric population at risk of OSAHS, with differences in current diagnostic and treatment pathways for children with and without comorbid conditions, and uncertainty over the impact OSAHS on health-related quality of life and longer-term outcomes and over treatment effectiveness.

After consideration, we concluded that the best way that we could contribute to an understanding of the potential cost-effectiveness of novel home-based devices in children is to suggest possible model structures and sources of evidence. We outline a model structure to evaluate the use of home-based testing with novel devices for OSAHS diagnosis in a subgroup of children who are being considered for adenotonsillectomy and discuss alternative approaches for children with comorbidities.

Finally, we note that we have not been able to estimate the cost-effectiveness of the novel devices for subgroups defined in the scope, which are related to potential equality issues. See section 6.4 below for further discussion of this matter.

6.3 Uncertainties

As highlighted above, there is high uncertainty over the EAG's estimates of cost-effectiveness for the novel devices. The probabilistic sensitivity analysis illustrates the extent of uncertainty related to the model parameter values, but there are other key uncertainties related to limitations and gaps in the evidence base. We have explored some of these uncertainties through scenario analysis, but there are other structural uncertainties that we have not been able to address, such as the impact of different contraindications for the novel devices, and longer term impacts of untreated OSAHS beyond impacts on quality of life, CV events and RTAs.

If alternative sources of evidence are used for the sensitivity and specificity of respiratory polygraphy, including those used in the recent NG202 guidelines, the novel devices all appear to be cost-effective at a willingness to pay threshold of £20,000 per QALY gained. The extent to which the values used in the base case, or the values used in scenario analyses, better reflect the accuracy of home respiratory polygraphy in current practice is not clear. We based our decision to use data from Xu et al 2017 [44] in the base case analysis on the recency of that evaluation and expert advice that the device (Nox-T3) was known to be currently used in practice in the UK. However, Xu et al was conducted in China and only included 80 participants.

Our analyses indicate that when we assume a higher proportion of people diagnosed with mild OSAHS as being on CPAP, use of novel devices becomes more cost-effective compared to respiratory polygraphy and oximetry. This is seen in the analysis of novel devices that 'over-diagnose the severity of mild OSAHS', and in the scenario analysis where a higher proportion of individuals with mild OSAHS are assumed to receive CPAP treatment.

The impacts of false positives (people without OSAHS who are diagnosed as having OSAHS) are unclear. We have assumed that the only impacts in the model for false positives are treatment costs (of one year in the base case, and two years in a scenario analysis), without any health benefits. There may be other costs incurred for false positives, and potentially, impacts on health-related quality of life or even survival, if the misdiagnosis of OSAHS is masking a different condition, which is therefore not being treated.

Finally, a key uncertainty that we have not been able to address is the cost-effectiveness of the novel devices in children. It is unlikely that the clinical or economic results are directly transferable from adults to children, and there are key uncertainties that make development of an economic model for children difficult at this time. We have outlined a model structure and sources of data in the hope that this might contribute to an understanding of the potential cost-effectiveness of novel home-based devices in children and inform further research.

6.4 Equality, diversity and inclusiveness

The scope for this assessment noted a number of potential equality considerations relating to demographic factors, comorbidities that increase the prevalence of OSAHS, and other characteristics that make the technologies difficult to access or use. Subgroup analysis was requested, where data permits, to investigate how the effectiveness and cost-effectiveness of the included tests might vary for people with these characteristics.

This included the subgroup of people from black, Asian and minority ethnic backgrounds, due to concerns that diagnostic technologies that use light-based assessment methods, such as oximetry and/or PPG sensors, may overestimate levels of oxygen in the blood for people with darker skin tones. The studies identified in our clinical and diagnostic review did not report results by ethnicity or race, so we could not adjust our economic analysis for this subgroup. However, we suggest that our conclusions over the cost-effectiveness of the novel home-based tests compared with oximetry are likely to hold for people from black, Asian and minority ethnic backgrounds. Our base case results for a mixed adult population indicated that the novel home tests are likely to provide a cost-effective alternative to

oximetry due to its poor sensitivity (see section 5.10.1), and the sensitivity of oximetry is expected to be worse for people with darker skin tones.

We cannot draw conclusions from our assessment on the relative effectiveness and cost-effectiveness of the different novel home devices and respiratory polygraphy for people from black, Asian and minority ethnic backgrounds. Similarly, we have not been able to estimate effectiveness or cost-effectiveness for other subgroups for whom there are potential equality considerations. These issues therefore remain as a matter for consideration by the committee.

7 CONCLUSIONS

7.1 Implications for service provision

Based on the clinical reviews and economic evaluation, we suggest the following conclusions related to services for people aged over 16 years undergoing home-based testing for suspected OSAHS:

- The estimated cost of the diagnostic pathway is lowest for oximetry and highest for respiratory polygraphy, with the cost for the novel devices lying in between. This is also true for total costs, including costs of OSAHS treatment (if indicated) and costs for care and treatment after cardiovascular events and road traffic accidents, in addition to costs for diagnosis. Estimated total costs are similar for the six novel devices, and differences between these cost estimates are highly uncertain.
- Although oximetry has the lowest cost of the included devices, it has relatively poor sensitivity. In particular, oximetry is estimated to misclassify a high proportion of people with mild OSAHS as not having OSAHS, and a high proportion of people with moderate or severe OSAHS as having mild OSAHS. This implies that patients who would benefit from treatment are not treated or that their treatment may be delayed. All of the novel devices therefore appear to offer relatively good value for money when compared against oximetry.
- In the base case all novel devices are estimated to have lower total costs and to produce fewer QALYs than respiratory polygraphy. For AcuPebble and Sunrise, the reduction in QALYs is estimated to be cost-effective compared to the reduction in costs, and this finding is consistent across the majority of scenario and sensitivity analyses.
- Across the base case and many sensitivity and scenario analyses, Brizzy, NightOwl, WatchPAT 300 and WatchPAT ONE do not appear to be cost-effective compared to respiratory polygraphy. However, some or all of these devices may be considered to be cost-effective in analyses exploring the more important areas of uncertainty, including: the accuracy of respiratory polygraphy, alternative accuracy data for the novel devices, the use of 4x4 contingency data to parameterise the decision tree and the extent of CPAP treatment in people with mild OSAHS.
- We emphasise that there is high uncertainty over the relative diagnostic accuracy estimates for all devices and advise caution in interpreting these results.

7.2 Suggested research priorities

Research to address significant gaps in the availability and quality of evidence on the diagnostic accuracy of sleep studies OSAHS, including:

- The diagnostic accuracy of home-based respiratory polygraphy compared against a laboratory PSG reference standard.
- The extent to which diagnostic accuracy evidence from older versions of devices for home based testing for OSAHS is transferable to new versions.
- The diagnostic accuracy of sleep studies conducted in the home using conventional and novel devices, rather than in a clinic setting.
- Studies of diagnostic accuracy of home-based sleep studies in children and young people under the age of 16, including those with and without comorbidities
- Indirect comparisons between novel and conventional devices and reference standards, with appropriate adjustment for heterogeneity. This would facilitate more robust comparison of results and economic evaluation, but we acknowledge that this is challenging given the high degree of heterogeneity in the current evidence base.
- Alternative study designs for collecting comparative data should be considered, including trials and prospective observational studies. We note that there are also challenges in designing a trial, given heterogeneity of patient populations, variations in practice, and differing opinions on the appropriateness of oximetry.

Further research to provide data to evaluate the clinical and economic effects of home-based sleep studies in children. Key evidence gaps for children include:

- The impact of OSAHS on health related quality of life for children stratified by OSAHS severity. Studies using preference-based utility instruments with a UK general population value set (e.g. using the CHU9D) would help to inform future economic evaluations.
- The relationship between OSAHS in childhood and long-term effects on health outcomes and well-being, and the extent to which these effects can be assumed to be causative and reversible with appropriate treatment.

8 REFERENCES

1. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for Scoring Respiratory Events in Sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *Journal of Clinical Sleep Medicine* 2012;08(05):597-619. doi: doi:10.5664/jcsm.2172
2. National Institute for Health and Care Excellence (NICE). Obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s. NICE guideline [NG202]: NICE, 2021.
3. Bitners AC, Arens R. Evaluation and Management of Children with Obstructive Sleep Apnea Syndrome. *Lung* 2020;198(2):257-70. doi: 10.1007/s00408-020-00342-5
4. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *The Lancet Respiratory Medicine* 2019;7(8):687-98. doi: 10.1016/S2213-2600(19)30198-5
5. Sunrise. Sunrise: User manual version 2.4 (22 APR 2022). Namur, Belgium: Sunrise, 2022.
6. Sunrise. Sunrise - User manual HCP - version 2.5 (07 FEB 2022). Namur, Belgium: Sunrise, 2022.
7. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine Clinical Practice Guideline. *J Clin Sleep Med* 2017;13(3):479-504. doi: 10.5664/jcsm.6506 [published Online First: 2017/02/07]
8. Kirk V, Baughn J, D'Andrea L, et al. American Academy of Sleep Medicine Position Paper for the Use of a Home Sleep Apnea Test for the Diagnosis of OSA in Children. *Journal of Clinical Sleep Medicine* 2017;13(10):1199-203. doi: doi:10.5664/jcsm.6772
9. Evans HJ, Gibson NA, Bennett J, et al. British Thoracic Society guideline for diagnosing and monitoring paediatric sleep-disordered breathing. *Thorax* 2023;78(Suppl 2):s1-s27. doi: 10.1136/thorax-2022-218938
10. Nomics. Brizzy by Nomics. User manual - Manuel d'utilisation - Gebruiksaanwijzing - Gebrauchsanleitung. Angleur, Belgium: Nomics, 2019.
11. Ectosense nv. NightOwl Software: Instructions for Use. Rotselaar, Belgium: Ectosense nv, 2020.
12. Itamar Medical Ltd. WatchPAT™ 300 - EU Declaration of Conformity: Itamar Medical Ltd., 2023.
13. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009 [published Online First: 2011/10/19]
14. Abrahamyan L, Sahakyan Y, Chung S, et al. Diagnostic accuracy of level IV portable sleep monitors versus polysomnography for obstructive sleep apnea: a systematic review and meta-analysis. *Sleep Breath* 2018;22(3):593-611. doi: 10.1007/s11325-017-1615-1 [published Online First: 2018/01/11]
15. Acurable. Phase 1 Virgen Macarena Trial Preliminary Analysis Report, 2023.
16. Nct. A Validation Study of the NightOwl Home Sleep Apnea Test. 2021
17. Alsaif SSD, W.; Steier, J.; Morrell, M.J.; Polkey, M.I.; Kelly, J.L. Mandibular movement monitor provides faster, yet accurate diagnosis, for obstructive sleep apnoea: a randomised controlled study, 2023.
18. Alsaif. S, Douglas. W, Steier. J, et al. S121 Mandibular movement monitor for the diagnosis of obstructive sleep apnoea: clinical application. *Thorax* 2022;77(Suppl 1):A75-A75. doi: 10.1136/thorax-2022-BTSabstracts.127
19. Devani N, Pramono RXA, Imtiaz SA, et al. Accuracy and usability of AcuPebble SA100 for automated diagnosis of obstructive sleep apnoea in the home environment setting: an evaluation study. *BMJ Open* 2021;11(12):e046803.
20. Sanchez Gomez J, Pramono RXA, Imtiaz SA, et al. Validation of a Wearable Medical Device for Automatic Diagnosis of OSA against Standard PSG. *J Clin Med* 2024;13(2) doi: 10.3390/jcm13020571 [published Online First: 2024/01/26]

21. Martinot J-B, Borel J-C, Cuthbert V, et al. Mandibular position and movements: Suitability for diagnosis of sleep apnoea. *Respirology* 2017;22(3):567-74. doi: https://doi.org/10.1111/resp.12929
22. Massie F, Mendes de Almeida D, Dreesen P, et al. An Evaluation of the NightOwl Home Sleep Apnea Testing System. *Journal of Clinical Sleep Medicine* 2018;14(10):1791-96.
23. Massie F, Van Pee B, Vits S, et al. Phenotyping REM OSA by means of peripheral arterial tone-based home sleep apnea testing and polysomnography: A critical assessment of the sensitivity and specificity of both methods. *J Sleep Res* 2022;31(2):e13481. doi: 10.1111/jsr.13481 [published Online First: 2021/09/13]
24. Van Pee B, Massie F, Vits S, et al. A multicentric validation study of a novel home sleep apnea test based on peripheral arterial tonometry. *Sleep* 2022;45(5) doi: 10.1093/sleep/zsac028 [published Online First: 2022/05/14]
25. Lyne CJ, Hamilton GS, Turton ARE, et al. Validation of a single-use and reusable home sleep apnea test based on peripheral arterial tonometry compared to laboratory polysomnography for the diagnosis of obstructive sleep apnea. *J Clin Sleep Med* 2023;19(8):1429-35. doi: 10.5664/jcsm.10568 [published Online First: 2023/04/20]
26. Pepin JL, Letesson C, Le-Dong NN, et al. Assessment of Mandibular Movement Monitoring With Machine Learning Analysis for the Diagnosis of Obstructive Sleep Apnea. *JAMA Network Open* 2020;3(1):e1919657.
27. Kelly JL, Ben Messaoud R, Joyeux-Faure M, et al. Diagnosis of Sleep Apnoea Using a Mandibular Monitor and Machine Learning Analysis: One-Night Agreement Compared to in-Home Polysomnography. *Frontiers in Neuroscience* 2022;16 doi: 10.3389/fnins.2022.726880
28. Storey MJ GJ, Flavell S, Royston R. WatchPAT ONE Vs NOX T3. ARTP (Association for Respiratory Technology & Physiology). Jury's Inn Hinckley Island Hotel, 2022.
29. Mueller CE, Li H, Begasse SM, et al. Sleep position, patient comfort, and technical performance with two established procedures for home sleep testing. *Sleep Breath* 2022;26(4):1673-81. doi: 10.1007/s11325-021-02530-w [published Online First: 2022/01/01]
30. Tauman R, Berall M, Berry R, et al. Watch-PAT is Useful in the Diagnosis of Sleep Apnea in Patients with Atrial Fibrillation. *Nat Sci Sleep* 2020;12:1115-21. doi: 10.2147/nss.S278752 [published Online First: 2020/12/11]
31. Pillar G, Berall M, Berry R, et al. Detecting central sleep apnea in adult patients using WatchPAT-a multicenter validation study. *Sleep Breath* 2020;24(1):387-98. doi: 10.1007/s11325-019-01904-5 [published Online First: 2019/08/14]
32. Nct. Clinical Evaluation of a Paediatric Sleep Diagnosis Technology. 2019
33. Martinot JB, Senny F, Denison S, et al. Mandibular movements identify respiratory effort in pediatric obstructive sleep apnea. *J Clin Sleep Med* 2015;11(5):567-74. doi: 10.5664/jcsm.4706 [published Online First: 2015/03/15]
34. Martinot JB, Cuthbert V, Le-Dong NN, et al. Clinical validation of a mandibular movement signal based system for the diagnosis of pediatric sleep apnea. *Pediatr Pulmonol* 2022;57(8):1904-13. doi: 10.1002/ppul.25320 [published Online First: 2021/03/02]
35. Martinot JB, Cuthbert V, Le‐Dong NN, et al. Clinical validation of a mandibular movement signal based system for the diagnosis of pediatric sleep apnea. *Pediatric pulmonology* 2022;57(8):1904-13.
36. Acurable. Co-morbidities Macarena Trial, 2023.
37. [redacted]
38. Nct. Clinical Evaluation of a Wearable Technology for the Diagnosis of Sleep Apnoea. 2019
39. Nct. Validation of an Integrated Digital Solution (SUNrise®) Versus Polysomnography for Obstructive Sleep Apnea Diagnosis. *https://clinicaltrialsgov/show/NCT05057975* 2021

40. Scotland SEEN. Evaluation of the Sunrise device for screening of adults with suspected Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS) and the Sunrise Patient Data Management Application.
41. Trust SEPCUHNF. Evaluation of the Sunrise device for screening of children with suspected Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS) and the Sunrise Patient Data Management Application.
42. Martinot J-B, Senny F, Denison S, et al. Mandibular movements identify respiratory effort in pediatric obstructive sleep apnea. *Journal of Clinical Sleep Medicine* 2015;11(5):567-74.
43. Khor YH, Khung S-W, Ruehland WR, et al. Portable evaluation of obstructive sleep apnea in adults: A systematic review. *Sleep Medicine Reviews* 2023;68:101743. doi: https://doi.org/10.1016/j.smrv.2022.101743
44. Xu L, Han F, Keenan BT, et al. Validation of the Nox-T3 Portable Monitor for Diagnosis of Obstructive Sleep Apnea in Chinese Adults. *J Clin Sleep Med* 2017;13(5):675-83. doi: 10.5664/jcsm.6582
10.5664/jcsm.6582. [published Online First: 2017/03/31]
45. Jaime Caro J, Eddy DM, Kan H, et al. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health* 2014;17(2):174-82. doi: 10.1016/j.jval.2014.01.003
10.1016/j.jval.2014.01.003. [published Online First: 2014/03/19]
46. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;8(36):iii-iv, ix-xi, 1-158. doi: 10.3310/hta8360
10.3310/hta8360. [published Online First: 2004/09/14]
47. Di Pumpo M, Nurchis MC, Moffa A, et al. Multiple-access versus telemedicine home-based sleep apnea testing for obstructive sleep apnea (OSA) diagnosis: a cost-minimization study. *Sleep & Breathing* 2022;26(4):1641-47.
48. Geessinck FAJ, Pleijhuis RG, Mentink RJ, et al. Cost-Effectiveness Analysis of the DiagnOSAS Screening Tool Compared With Polysomnography Diagnosis in Dutch Primary Care. *Journal of Clinical Sleep Medicine* 2018;14(6):1005-15.
49. Phua CQ, Jang IJ, Tan KB, et al. Reducing cost and time to diagnosis and treatment of obstructive sleep apnea using ambulatory sleep study: a Singapore sleep centre experience. *Sleep & Breathing* 2021;25(1):281-88.
50. Guest JF, Helter MT, Morga A, Stradling JR. Cost-effectiveness of using continuous positive airway pressure in the treatment of severe obstructive sleep apnoea/hypopnoea syndrome in the UK. *Thorax* 2008;63(10):860-65.
51. McMillan A, Bratton DJ, Faria R, et al. A multicentre randomised controlled trial and economic evaluation of continuous positive airway pressure for the treatment of obstructive sleep apnoea syndrome in older people: PREDICT. *Health Technol Assess* 2015;19(40):1-188. doi: 10.3310/hta19400
10.3310/hta19400. [published Online First: 2015/06/13]
52. National Guideline Centre. Obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s: Economic analyses NICE guideline NG202. Economic analysis report. London, 2021.
53. American Academy of Sleep Medicine. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. *Sleep* 1999;22(5):667-89. [published Online First: 1999/08/18]
54. McDaid C, Griffin S, Weatherly H, et al. Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea–hypopnoea syndrome: a systematic review and economic analysis. *Health Technol Assess* 2009;13:4. doi: 10.3310/hta13040

55. Skirko JR, James KT, Garrison LP, Weaver EM. Development of a Sleep Apnea-Specific Health State Utility Algorithm. *JAMA Otolaryngol Head Neck Surg* 2020;146(3):270-77. doi: 10.1001/jamaoto.2019.4469
10.1001/jamaoto.2019.4469. [published Online First: 2020/01/31]
56. Vogler K, Daboul A, Obst A, et al. Quality of life in patients with obstructive sleep apnea: Results from the study of health in Pomerania. *J Sleep Res* 2023;32(1):e13702. doi: 10.1111/jsr.13702
10.1111/jsr.13702. Epub 2022 Aug 21. [published Online First: 2022/09/03]
57. Sakki AJ, Roine RP, Mäkinen LK, et al. Impact of tonsillotomy versus tonsillectomy on health-related quality of life and healthcare costs in children with sleep-disordered breathing. *J Laryngol Otol* 2021:1-22. doi: 10.1017/s0022215121003832
10.1017/S0022215121003832. [published Online First: 2021/11/26]
58. Apajasalo M, Rautonen J, Holmberg C, et al. Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). *Qual Life Res* 1996;5(6):532-8. doi: 10.1007/bf00439227
10.1007/BF00439227. [published Online First: 1996/12/01]
59. Kataoka H, Miyatake N, Ichikawa H, et al. Sub-analysis of the prevalence of locomotive syndrome and its relationship with health-related quality of life in patients with obstructive sleep apnea syndrome as classified by age and sex. *Sleep Biol Rhythms* 2019;17(1):149-53. doi: 10.1007/s41105-018-0194-2
60. Lugo VM, Garmendia O, Suarez-Girón M, et al. Comprehensive management of obstructive sleep apnea by telemedicine: Clinical improvement and cost-effectiveness of a Virtual Sleep Unit. A randomized controlled trial. *PLoS One* 2019;14(10):e0224069. doi: 10.1371/journal.pone.0224069
10.1371/journal.pone.0224069. eCollection 2019. [published Online First: 2019/10/28]
61. Walia HK, Thompson NR, Katzan I, et al. Impact of Sleep-Disordered Breathing Treatment on Quality of Life Measures in a Large Clinic-Based Cohort. *J Clin Sleep Med* 2017;13(11):1255-63. doi: 10.5664/jcsm.6792
10.5664/jcsm.6792. [published Online First: 2017/10/11]
62. Cambron-Mellott MJ, Mettam S, Li VW, et al. Examining the impact of excessive daytime sleepiness on utility scores in patients with obstructive sleep apnoea and/or narcolepsy in five European countries. *BMC Neurol* 2022;22(1):317. doi: 10.1186/s12883-022-02827-7
10.1186/s12883-022-02827-7. [published Online First: 2022/08/26]
63. Sánchez-de-la-Torre M, Nadal N, Cortijo A, et al. Role of primary care in the follow-up of patients with obstructive sleep apnoea undergoing CPAP treatment: a randomised controlled trial. *Thorax* 2015;70(4):346-52. doi: 10.1136/thoraxjnl-2014-206287
10.1136/thoraxjnl-2014-206287. [published Online First: 2015/03/15]
64. Huber FL, Furian M, Kohler M, et al. Health Preference Measures in Patients with Obstructive Sleep Apnea Syndrome Undergoing Continuous Positive Airway Pressure Therapy: Data from a Randomized Trial. *Respiration* 2021;100(4):328-38. doi: 10.1159/000513306
10.1159/000513306. Epub 2021 Feb 4. [published Online First: 2021/02/05]
65. Wimms AJ, Kelly JL, Turnbull CD, et al. Continuous positive airway pressure versus standard care for the treatment of people with mild obstructive sleep apnoea (MERGE): a multicentre, randomised controlled trial. *Lancet Respir Med* 2020;8(4):349-58. doi: 10.1016/s2213-2600(19)30402-3
10.1016/S2213-2600(19)30402-3. Epub 2019 Dec 2. [published Online First: 2019/12/07]
66. Sharples L, Glover M, Clutterbuck-James A, et al. Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea–hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive airway pressure. *Health Technol Assess* 2014;18(67) doi: 10.3310/hta18670

67. Rizzi CF, Ferraz MB, Poyares D, Tufik S. Quality-adjusted life-years gain and health status in patients with OSAS after one year of continuous positive airway pressure use. *Sleep* 2014;37(12):1963-8. doi: 10.5665/sleep.4250
10.5665/sleep.4250. [published Online First: 2014/10/18]
68. Ylitalo-Heikkilä M, Virkkula P, Sintonen H, et al. Different rhinologic diseases cause a similar multidimensional decrease in generic health-related quality of life. *Clin Otolaryngol* 2018;43(6):1487-93. doi: 10.1111/coa.13190
10.1111/coa.13190. Epub 2018 Sep 2. [published Online First: 2018/07/12]
69. Kuik K, Tan ML, Ho JTF, et al. Subjective Efficiency Evaluation after Maxillomandibular Advancement Surgery in Obstructive Sleep Apnea Patients. *J Clin Med* 2023;12(12) doi: 10.3390/jcm12124023
10.3390/jcm12124023. [published Online First: 2023/06/28]
70. Pinczel AJ, Woods CM, Catcheside PG, et al. Sleep Apnea Multi-level Surgery (SAMS) Trial: Long-term Observational Outcomes. *Sleep* 2023 doi: 10.1093/sleep/zsad218
10.1093/sleep/zsad218. [published Online First: 2023/08/22]
71. [redacted]
72. [redacted]
73. National Institute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual (PMG36). Process and Methods, 2022.
74. Pepin JL, Le-Dong NN, Cuthbert V, et al. Mandibular Movements are a Reliable Noninvasive Alternative to Esophageal Pressure for Measuring Respiratory Effort in Patients with Sleep Apnea Syndrome. *Nat Sci Sleep* 2022;14:635-44. doi: 10.2147/nss.S346229 [published Online First: 2022/04/22]
75. National Guideline Centre. NICE Evidence Reviews Collection. Diagnostic tests for OSAHS, OHS and COPD–OSAHS overlap syndrome: Obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s: Evidence review D. London: National Institute for Health and Care Excellence (NICE) Copyright © NICE 2021. 2021.
76. Pereira EJ, Driver HS, Stewart SC, Fitzpatrick MF. Comparing a combination of validated questionnaires and level III portable monitor with polysomnography to diagnose and exclude sleep apnea. *J Clin Sleep Med* 2013;9(12):1259-66. doi: 10.5664/jcsm.3264
10.5664/jcsm.3264. [published Online First: 2013/12/18]
77. Ioachimescu OC, Allam JS, Samarghandi A, et al. Performance of peripheral arterial tonometry-based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort. *J Clin Sleep Med* 2020;16(10):1663-74. doi: 10.5664/jcsm.8620
10.5664/jcsm.8620. [published Online First: 2020/06/10]
78. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099. doi: 10.1136/bmj.j2099
10.1136/bmj.j2099. [published Online First: 2017/05/26]
79. Ward S, Lloyd Jones M, Pandor A, et al. A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technol Assess* 2007;11(14):1-160, iii-iv. doi: 10.3310/hta11140
10.3310/hta11140. [published Online First: 2007/04/06]
80. Davies AR, Smeeth L, Grundy EM. Contribution of changes in incidence and mortality to trends in the prevalence of coronary heart disease in the UK: 1996 2005. *Eur Heart J* 2007;28(17):2142-7. doi: 10.1093/eurheartj/ehm272
10.1093/eurheartj/ehm272. Epub 2007 Jul 18. [published Online First: 2007/07/20]
81. Department for T. Reported road collisions, vehicles and casualties tables for Great Britain [accessed 15th September 2023.
82. Office for National Statistics (ONS). National life tables, England and Wales, 2021.

83. Office for National Statistics (ONS). Death registration summary statistics, England and Wales: 2022, 2022.
84. Dennis MS, Burn JP, Sandercock PA, et al. Long-term survival after first-ever stroke: the Oxfordshire Community Stroke Project. *Stroke* 1993;24(6):796-800. doi: 10.1161/01.str.24.6.796
10.1161/01.str.24.6.796. [published Online First: 1993/06/01]
85. Clark TG, Murphy MF, Rothwell PM. Long term risks of stroke, myocardial infarction, and vascular death in "low risk" patients with a non-recent transient ischaemic attack. *J Neurol Neurosurg Psychiatry* 2003;74(5):577-80. doi: 10.1136/jnnp.74.5.577
10.1136/jnnp.74.5.577. [published Online First: 2003/04/18]
86. Rutten-Jacobs LC, Arntz RM, Maaijwee NA, et al. Long-term mortality after stroke among adults aged 18 to 50 years. *JAMA* 2013;309(11):1136-44. doi: 10.1001/jama.2013.842
10.1001/jama.2013.842. [published Online First: 2013/03/21]
87. Ellis CJ, Gamble GD, Williams MJA, et al. All-Cause Mortality Following an Acute Coronary Syndrome: 12-Year Follow-Up of the Comprehensive 2002 New Zealand Acute Coronary Syndrome Audit. *Heart Lung Circ* 2019;28(2):245-56. doi: 10.1016/j.hlc.2017.10.015
10.1016/j.hlc.2017.10.015. Epub 2017 Nov 7. [published Online First: 2017/11/19]
88. Plakht Y, Gilutz H, Shiyovich A. Excess long-term mortality among hospital survivors of acute myocardial infarction. Soroka Acute Myocardial Infarction (SAMI) project. *Public Health* 2017;143:25-36. doi: 10.1016/j.puhe.2016.09.032
10.1016/j.puhe.2016.09.032. Epub 2016 Nov 24. [published Online First: 2017/02/06]
89. Rosengren A, Wilhelmsen L, Hagman M, Wedel H. Natural history of myocardial infarction and angina pectoris in a general population sample of middle-aged men: a 16-year follow-up of the Primary Prevention Study, Göteborg, Sweden. *J Intern Med* 1998;244(6):495-505. doi: 10.1111/j.1365-2796.1998.00394.x
10.1111/j.1365-2796.1998.00394.x. [published Online First: 1999/01/20]
90. National Clinical Guideline C. Unstable angina and NSTEMI: the early management of unstable angina and non-ST-segment-elevation myocardial infarction. NICE clinical guideline 94. London, 2009.
91. Brønnum-Hansen H, Jørgensen T, Davidsen M, et al. Survival and cause of death after myocardial infarction: the Danish MONICA study. *J Clin Epidemiol* 2001;54(12):1244-50. doi: 10.1016/s0895-4356(01)00405-x
10.1016/s0895-4356(01)00405-x. [published Online First: 2001/12/26]
92. Dennis M, Bamford J, Sandercock P, Warlow C. Prognosis of transient ischemic attacks in the Oxfordshire Community Stroke Project. *Stroke* 1990;21(6):848-53. doi: 10.1161/01.str.21.6.848
10.1161/01.str.21.6.848. [published Online First: 1990/06/01]
93. Brønnum-Hansen H, Davidsen M, Thorvaldsen P. Long-term survival and causes of death after stroke. *Stroke* 2001;32(9):2131-6. doi: 10.1161/hs0901.094253
10.1161/hs0901.094253. [published Online First: 2001/09/08]
94. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;13(5):509-18. doi: 10.1111/j.1524-4733.2010.00700.x
10.1111/j.1524-4733.2010.00700.x. Epub 2010 Mar 10. [published Online First: 2010/03/17]
95. Pockett RD, McEwan P, Ray J, et al. Prospective utility study of patients with multiple cardiovascular events. *J Med Econ* 2018;21(6):616-21. doi: 10.1080/13696998.2018.1454453
10.1080/13696998.2018.1454453. Epub 2018 Apr 3. [published Online First: 2018/03/21]
96. Luengo-Fernandez R, Gray AM, Bull L, et al. Quality of life after TIA and stroke: ten-year results of the Oxford Vascular Study. *Neurology* 2013;81(18):1588-95. doi: 10.1212/WNL.0b013e3182a9f45f
10.1212/WNL.0b013e3182a9f45f. Epub 2013 Oct 9. [published Online First: 2013/10/11]

97. Rieckmann N, Neumann K, Feger S, et al. Health-related qualify of life, angina type and coronary artery disease in patients with stable chest pain. *Health Qual Life Outcomes* 2020;18(1):140. doi: 10.1186/s12955-020-01312-4 [published Online First: 2020/05/16]
98. Currie CJ, McEwan P, Peters JR, et al. The routine collation of health outcomes data from hospital treated subjects in the Health Outcomes Data Repository (HODaR): descriptive analysis from the first 20,000 subjects. *Value Health* 2005;8(5):581-90. doi: 10.1111/j.1524-4733.2005.00046.x
10.1111/j.1524-4733.2005.00046.x. [published Online First: 2005/09/24]
99. Pink J, Petrou S, Williamson E, et al. Properties of patient-reported outcome measures in individuals following acute whiplash injury. *Health Qual Life Outcomes* 2014;12:38. doi: 10.1186/1477-7525-12-38
10.1186/1477-7525-12-38. [published Online First: 2014/03/15]
100. Kohler M, Smith D, Tippett V, Stradling JR. Predictors of long-term compliance with continuous positive airway pressure. *Thorax* 2010;65(9):829-32. doi: 10.1136/thx.2010.135848
10.1136/thx.2010.135848. [published Online First: 2010/09/02]
101. Feltner C, Wallace IF, Aymes S, et al. Screening for Obstructive Sleep Apnea in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2022;328(19):1951-71. doi: 10.1001/jama.2022.18357
10.1001/jama.2022.18357. [published Online First: 2022/11/16]
102. Labarca G, Schmidt A, Dreyse J, et al. Efficacy of continuous positive airway pressure (CPAP) in patients with obstructive sleep apnea (OSA) and resistant hypertension (RH): Systematic review and meta-analysis. *Sleep Med Rev* 2021;58:101446. doi: 10.1016/j.smrv.2021.101446
10.1016/j.smrv.2021.101446. Epub 2021 Jan 28. [published Online First: 2021/02/20]
103. Patil SP, Ayappa IA, Caples SM, et al. Treatment of Adult Obstructive Sleep Apnea With Positive Airway Pressure: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and GRADE Assessment. *J Clin Sleep Med* 2019;15(2):301-34. doi: 10.5664/jcsm.7638
10.5664/jcsm.7638. [published Online First: 2019/02/10]
104. Zhang D, Luo J, Qiao Y, Xiao Y. Continuous positive airway pressure therapy in non-sleepy patients with obstructive sleep apnea: results of a meta-analysis. *J Thorac Dis* 2016;8(10):2738-47. doi: 10.21037/jtd.2016.09.40
10.21037/jtd.2016.09.40. [published Online First: 2016/11/22]
105. de Vries GE, Wijkstra PJ, Houwerzijl EJ, et al. Cardiovascular effects of oral appliance therapy in obstructive sleep apnea: A systematic review and meta-analysis. *Sleep Med Rev* 2018;40:55-68. doi: 10.1016/j.smrv.2017.10.004
10.1016/j.smrv.2017.10.004. Epub 2017 Oct 26. [published Online First: 2017/12/03]
106. Lewis EF, Wang R, Punjabi N, et al. Impact of continuous positive airway pressure and oxygen on health status in patients with coronary heart disease, cardiovascular risk factors, and obstructive sleep apnea: A Heart Biomarker Evaluation in Apnea Treatment (HEARTBEAT) analysis. *Am Heart J* 2017;189:59-67. doi: 10.1016/j.ahj.2017.03.001
10.1016/j.ahj.2017.03.001. Epub 2017 Mar 14. [published Online First: 2017/06/20]
107. Kushida CA, Nichols DA, Holmes TH, et al. Effects of continuous positive airway pressure on neurocognitive function in obstructive sleep apnea patients: The Apnea Positive Pressure Long-term Efficacy Study (APPLES). *Sleep* 2012;35(12):1593-602. doi: 10.5665/sleep.2226
10.5665/sleep.2226. [published Online First: 2012/12/04]
108. Jones K WHBSCACMDAFJGMHSMSODPSRD. Unit Costs of Health and Social Care, 2022.
109. Bloch KE, Huber F, Furian M, et al. Autoadjusted versus fixed CPAP for obstructive sleep apnoea: a multicentre, randomised equivalence trial. *Thorax* 2018;73(2):174-84. doi: 10.1136/thoraxjnl-2016-209699
10.1136/thoraxjnl-2016-209699. Epub 2017 Oct 5. [published Online First: 2017/10/07]

110. Danese MD, Gleeson M, Kutikova L, et al. Estimating the economic burden of cardiovascular events in patients receiving lipid-modifying therapy in the UK. *BMJ Open* 2016;6(8):e011805. doi: 10.1136/bmjopen-2016-011805
10.1136/bmjopen-2016-011805. [published Online First: 2016/08/09]
111. Vemer P, Corro Ramos I, Van Voorn G, et al. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics* 2016;34:349-61.
112. Paulden M. Calculating and interpreting ICERs and net benefit. *Pharmacoeconomics* 2020;38(8):785-807.
113. Marcus CL, Moore RH, Rosen CL, et al. A randomized trial of adenotonsillectomy for childhood sleep apnea. *New England Journal of Medicine* 2013;368(25):2366-76.
114. Witmans M, Shafazand S. Adenotonsillectomy or watchful waiting in the management of childhood obstructive sleep apnea. *Journal of Clinical Sleep Medicine* 2013;9(11):1225-27.
115. Bergeron M, Duggins AL, Cohen AP, Ishman SL. Comparison of patient-and parent-reported quality of life for patients treated for persistent obstructive sleep apnea. *Otolaryngology–Head and Neck Surgery* 2018;159(4):789-95.
116. Katz SL, MacLean JE, Barrowman N, et al. Long-term impact of sleep-disordered breathing on quality of life in children with obesity. *Journal of Clinical Sleep Medicine* 2018;14(3):451-58.
117. Türkoğlu S, Tahsin Somuk B, Sapmaz E, Bilgiç A. Effect of adenotonsillectomy on sleep problems, attention deficit hyperactivity disorder symptoms, and quality of life of children with adenotonsillar hypertrophy and sleep-disordered breathing. *The International Journal of Psychiatry in Medicine* 2019;54(3):231-41.
118. Lambe T, Frew E, Ives NJ, et al. Mapping the Paediatric Quality of Life Inventory (PedsQL™) Generic Core Scales onto the Child Health Utility Index–9 Dimension (CHU-9D) Score for Economic Evaluation in Children. *PharmacoEconomics* 2018;36(4):451-65. doi: 10.1007/s40273-017-0600-7
119. Mpundu-Kaambwa C, Chen G, Huynh E, et al. Mapping the PedsQL™ onto the CHU9D: An Assessment of External Validity in a Large Community-Based Sample. *PharmacoEconomics* 2019;37(9):1139-53. doi: 10.1007/s40273-019-00808-2
120. Yu PK, Cook K, Liu J, et al. Comparison of Caregiver‐and Child‐Reported Quality of Life in Children With Sleep‐Disordered Breathing. *Otolaryngology–Head and Neck Surgery* 2023;168(1):74-81.
121. Martinot JB L-DN, Malhotra A, Pépin JL. . Respiratory Effort During Sleep and Prevalent Hypertension in Obstructive Sleep Apnoea. *European Respiratory Journal* 2022;61(3) doi: 10.1183/13993003.01486-2022
122. ENT UK, Royal College of Surgeons. Commissioning guide: tonsillectomy, 2021.
123. Venekamp RP, Hearne BJ, Chandrasekharan D, et al. Tonsillectomy or adenotonsillectomy versus non-surgical management for obstructive sleep-disordered breathing in children. *Cochrane Database of Systematic Reviews* 2015(10) doi: 10.1002/14651858.CD011165.pub2
124. Cottrell J, Zahr SK, Propst EJ, et al. Morbidity and mortality from adenotonsillectomy in children with trisomy 21. *International Journal of Pediatric Otorhinolaryngology* 2020;138:110377. doi: https://doi.org/10.1016/j.ijporl.2020.110377
125. Yang B, Mallett S, Takwoingi Y, et al. QUADAS-C: A Tool for Assessing Risk of Bias in Comparative Diagnostic Accuracy Studies. *Ann Intern Med* 2021;174(11):1592-99. doi: 10.7326/m21-2234 [published Online First: 2021/10/27]
126. Economic Evaluations & Models - Embase. CADTH Search Filters Database. Ottawa: CADTH, 2023.
127. Economic Evaluations & Models - MEDLINE. CADTH Search Filters Database. Ottawa: CADTH, 2023.

128. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health* 2012;15(6):796-803. doi: 10.1016/j.jval.2012.06.012 [published Online First: 2012/09/25]
129. Arber M, Garcia S, Veale T, et al. PERFORMANCE OF OVID MEDLINE SEARCH FILTERS TO IDENTIFY HEALTH STATE UTILITY STUDIES. *International Journal of Technology Assessment in Health Care* 2017;33(4):472-80. doi: 10.1017/S0266462317000897 [published Online First: 2017/10/25]
130. Abraham WT, Trupp RJ, Phillilps B, et al. Validation and clinical utility of a simple in-home testing tool for sleep-disordered breathing and arrhythmias in heart failure: results of the Sleep Events, Arrhythmias, and Respiratory Analysis in Congestive Heart Failure (SEARCH) study. *Congestive Heart Failure* 2006;12(5):241-7; quiz 48-9.
131. Agency for Care E. Ambulatory sleep study for diagnosing obstructive sleep apnoea. Singapore: Agency for Care Effectiveness (ACE), 2019.
132. Aldabayan YS ARA, Van Der Leer J, Mansell SK, Hurst JR, Mandal S. A comparison of peripheral arterial tone technology with limited polygraphy to diagnose obstructive sleep apnoea in UK. ERS International Congress 2017 2017.
133. Alonso Alvarez MD, Santos JT, Guevara JC, et al. Reliability of home respiratory polygraphy for the diagnosis of sleep apnea-hypopnea syndrome: analysis of costs. *Archivos de Bronconeumologia* 2008;44(1):22-28.
134. Alonso-Alvarez ML, Teran-Santos J, Ordax Carbajo E, et al. Reliability of home respiratory polygraphy for the diagnosis of sleep apnea in children. *Chest* 2015;147(4):1020-28.
135. Andreu AL, Chiner E, Sancho-Chust JN, et al. Effect of an ambulatory diagnostic and treatment programme in patients with sleep apnoea. *European Respiratory Journal* 2012;39(2):305-12.
136. Aurora RN, Patil SP, Punjabi NM. Portable Sleep Monitoring for Diagnosing Sleep Apnea in Hospitalized Patients With Heart Failure. *Chest* 2018;154(1):91-98.
137. Ayas NT, Pittman S, MacDonald M, White DP. Assessment of a wrist-worn device in the detection of obstructive sleep apnea. *Sleep Medicine* 2003;4(5):435-42.
138. Badami V, Naqvi SF, Easterling J, et al. High Resolution Pulse Oximetry (HRPO) to Determine Risk for Obstructive Sleep Apnea (OSA) in Hospitalized Patients. *American Journal of Respiratory and Critical Care Medicine Conference: American Thoracic Society International Conference, ATS* 2021;203(9)
139. Badami V, Rojas E, Naqvi SF, et al. Correlation of Inpatient and Outpatient Sleep Studies to Determine Risk of Sleep Disordered Breathing. *American Journal of Respiratory and Critical Care Medicine Conference: International Conferenceof the American Thoracic Society, ATS* 2022;205(1)
140. Bar A, Pillar G, Dvir I, et al. Evaluation of a portable device based on peripheral arterial tone for unattended home sleep studies. *Chest* 2003;123(3):695-703.
141. Berry RB, Hill G, Thompson L, McLaurin V. Portable monitoring and autotitration versus polysomnography for the diagnosis and treatment of sleep apnea. *Sleep* 2008;31(10):1423-31.
142. Bhattacharjee R, Benjafield A, Blase A, et al. The accuracy of a portable sleep monitor to diagnose obstructive sleep apnea in adolescent patients. *Journal of Clinical Sleep Medicine* 2021;17(7):1379-87.
143. Boulos M, Kamra M, Colelli D, et al. Screening for obstructive sleep apnea in stroke and TIA patients using a home sleep apnea test vs. in-laboratory polysomnography: A randomized controlled trial. *Neurology Conference: 72nd Annual Meeting of the American Academy of Neurology, AAN* 2020;94(15 Supplement)
144. Boulos M, Kamra M, Kirollos N, et al. Screening for obstructive sleep apnea in stroke and tia patients using a portable sleep monitor: a randomized controlled trial. *Sleep medicine* 2019;64:S47. doi: 10.1016/j.sleep.2019.11.129

145. Boulos M, Kamra M, Kirolos N, et al. Screening for obstructive sleep apnea in stroke and tia patients using a home sleep apnea test versus in-laboratory polysomnography: a randomized controlled trial. *International journal of stroke* 2018;13(2):20. doi: 10.1177/1747493018789543
146. Boulos M, Kamra M, Patel N, et al. Sleep apnea screening using mobile ambulatory recorders after TIA/stroke (sleap smart): a randomized controlled trial. *International journal of stroke* 2017;12(4):23. doi: 10.1177/1747493017721569
147. Boulos MI, Kamra M, Colelli DR, et al. SLEAP SMART (Sleep Apnea Screening Using Mobile Ambulatory Recorders After TIA/Stroke): A Randomized Controlled Trial. *Stroke* 2022;53(3):710-18.
148. Braun F, Baty F, Van Zaen J, et al. Multi-parametric vital sign estimation for monitoring of sleep disordered breathing. *Respiration* 2021;100(7):671.
149. Influence of Measurement Location on Reflectance Pulse Oximetry in Sleep Apnea Patients: Wrist vs. Upper Arm. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021 1-5 Nov. 2021.
150. Bresler M, Sheffy K, Pillar G, et al. Differentiating between light and deep sleep stages using an ambulatory device based on peripheral arterial tonometry. *Physiological Measurement* 2008;29(5):571-84. doi: 10.1088/0967-3334/29/5/004
151. Bubu O, White P, Gross J, et al. Examining the Diagnostic Validity of the Watchpat in a Preliminary Sample of Cognitive Normal Black/African-American Older Adults. *Sleep* 2022;45(SUPPL 1):A320-A21.
152. Carey SA, van Zyl JS, Williams S, et al. The Utility of Home Sleep Apnea Testing in the Advanced Heart Failure Populations. *American Journal of Cardiology* 2023;191:8-13.
153. Cassibba JM, JB; Hullo, E; Aubertin,G; Beydon, N; Dupont-Athénor, A; Mortamet, G; Pépin, JL. Analysis of mandibular movements to improve ventilatory management of children with obstructive sleep apnea syndrome treated with continuous positive airway pressure or non-invasive ventilation. European Respiratory Society International Congress 2023. Milan, Italy, 2023.
154. Castillo-Escario Y, Ferrer-Lluis I, Montserrat JM, Jane R. Automatic Silence Events Detector from Smartphone Audio Signals: A Pilot mHealth System for Sleep Apnea Monitoring at Home. *Annual International Conference Of The IEEE Engineering In Medicine And Biology Society* 2019;2019:4982-85.
155. Cerina L, Overeem S, Papini GB, et al. A sleep stage estimation algorithm based on cardiorespiratory signals derived from a suprasternal pressure sensor. *Journal of Sleep Research* 2023:e14015.
156. Ceylan T, Firat H, Kuran G, et al. Quick Diagnosis in Obstructive Sleep Apnea Syndrome: WatchPAT-200. *Iranian Red Crescent Medical Journal* 2012;14(8):475-8.
157. Chai-Coetzer CL, Antic NA, Hamilton GS, et al. Physician Decision Making and Clinical Outcomes With Laboratory Polysomnography or Limited-Channel Sleep Studies for Obstructive Sleep Apnea: A Randomized Trial. *Annals of Internal Medicine* 2017;166(5):332-40.
158. Chakar B SF, Cambron L, Poirrier AL, Bruwier A, Baharloo F, Poirrier R. Validation of mandibular movements' analysis to measure sleep in adults with sleep complaints by comparison with actigraphy and polysomnography. *Sleep Science* 2022;15(3):318–25. doi: 10.5935/1984-0063.20220057
159. Chen Y, Wang W, Guo Y, et al. A Single-Center Validation of the Accuracy of a Photoplethysmography-Based Smartwatch for Screening Obstructive Sleep Apnea. *Nature & Science of Sleep* 2021;13:1533-44.
160. Chernyshev OY, McCarty DE, Moul DE, et al. A pilot study: portable out-of-center sleep testing as an early sleep apnea screening tool in acute ischemic stroke. *Nature & Science of Sleep* 2015;7:127-38.

161. Cheung TW, Lam DS, Chan PC, et al. Comparing respiratory polygraphy with pulse transit time analysis versus overnight polysomnography in the diagnosis of obstructive sleep apnoea in children. *Sleep Medicine* 2021;81:457-62.
162. Chiner E, Canovas C, Molina V, et al. Home respiratory polygraphy is useful in the diagnosis of childhood obstructive sleep apnea syndrome. *Journal of Clinical Medicine* 2020;9(7):1-13.
163. Y C, Stafford PL, Ferguson JD, et al. Difference in Obstructive Sleep Apnea Measures Between In-lab Vs. Home in Patients With Atrial Fibrillation. A73 DISPATCH FROM THE SLEEP LAB: UNCOVERING OSA PATHOPHYSIOLOGY:A2383-A83.
164. Choi JH, Kim EJ, Kim YS, et al. Validation study of portable device for the diagnosis of obstructive sleep apnea according to the new AASM scoring criteria: Watch-PAT 100. *Acta Oto-Laryngologica* 2010;130(7):838-43.
165. Choi JH, Lee B, Lee JY, Kim HJ. Validating the Watch-PAT for diagnosing obstructive sleep apnea in adolescents. *Journal of Clinical Sleep Medicine* 2018;14(10):1741-47.
166. Corral J, Sanchez-Quiroga MA, Carmona-Bernal C, et al. Conventional polysomnography is not necessary for the management of most patients with suspected obstructive sleep apnea noninferiority, randomized controlled trial. *American Journal of Respiratory and Critical Care Medicine* 2017;196(9):1181-90.
167. Crupi R, Faetti T, Paradiso R. Preliminary evaluation of Wearable Wellness System for Obstructive Sleep Apnea detection. *Annual International Conference Of The IEEE Engineering In Medicine And Biology Society* 2015;2015:4142-5.
168. Damiani MF, Carratu P, Ventura VA, et al. Home-based evaluation in patients with high risk for moderate to severe OSA. *European respiratory journal* 2012;40
169. De Chazal P, Fox N, O'Hare E, et al. Sleep/wake measurement using a non-contact biomotion sensor. *Journal of Sleep Research* 2011;20(2):356-66.
170. de Chazal P, Heneghan C, McNicholas WT. Multimodal detection of sleep apnoea using electrocardiogram and oximetry signals. *Philosophical Transactions of the Royal Society London, Series a (Mathematical, Physical & Engineering Sciences)* 2009;367(1887):369-89.
171. de Oliveira ACT, Martinez D, Vasconcelos LFT, et al. Diagnosis of obstructive sleep apnea syndrome and its outcomes with home portable monitoring. *Chest* 2009;135(2):330-36.
172. Ding Q, Liu J, Wu J, et al. Validation of a portable monitor compared with polysomnography for screening of obstructive sleep apnea in polio survivors. *Frontiers in Neurology* 2023;14 (no pagination)
173. Domingo C, Vigil L. Effectiveness of unattended ambulatory sleep studies for the diagnosis and treatment of OSAS. *Journal of Evaluation in Clinical Practice* 2011;17(1):26-31.
174. Drks. Follow-up of PAP treated OSA patients in their home environment – a comparative study. *https://trialsearchwhoint/Trial2aspx?TrialID=DRKS00007705* 2015
175. Duran-Cantolla J, Corral Penafiel J, Alonso ML, et al. Validity and cost-efectiviness analysis of pediatric home respiratory poligraphy for the diagnosis of obstructive sleep apnea: rationale, design, and methodology. *Sleep medicine* 2017;40:e84 - e85.
176. Duran-Cantolla J, Oceja E, Alvarez A, et al. Validity and cost-efectiviness of pediatric home respiratory poligraphy for the diagnosis of obstructive sleep apnea: rationale, designv the diagnosis of obstructive sleep apnea: rationale, design and methodology. *American journal of respiratory and critical care medicine* 2018;197(MeetingAbstracts)
177. Escobar A, Reyes S, Majid R. FACTORS ASSOCIATED WITH DISCORDANT INITIAL HOME SLEEP APNEA TESTING AND POLYSOMNOGRAPHY. *Sleep* 2023;46:A209-A09. doi: https://doi.org/10.1093/sleep/zsad077.0470
178. Ferguson KA, Heighway K, George CF. Home studies versus polysomnography in the diagnosis of suspected obstructive sleep apnea (OSA). *American journal of respiratory and critical care medicine* 2002;165(Suppl 8):A407.

179. Fischer J, Raschke F. Cost-benefit analysis in patients with sleep-related breathing disorders - diagnosis and ncpap therapy during medical rehabilitation. *Biomedizinische Technik Biomedical engineering* 2003;48(9):245 - 51. doi: 10.1515/bmte.2003.48.9.245
180. Fishman H, Massicotte C, Li R, et al. The accuracy of an ambulatory level III sleep study compared to a level I sleep study for the diagnosis of sleep-disordered breathing in children with neuromuscular disease. *Sleep medicine* 2019;64:S116. doi: 10.1016/j.sleep.2019.11.319
181. Fishman H, Massicotte C, Li R, et al. The Accuracy of an Ambulatory Level III Sleep Study Compared to a Level I Sleep Study for the Diagnosis of Sleep-Disordered Breathing in Children With Neuromuscular Disease. *Journal of Clinical Sleep Medicine* 2018;14(12):2013-20.
182. Folmer RL, Boudreau EA, Atwood CW, et al. Study protocol to assess de-implementation of the initial provider encounter for diagnosis and treatment of obstructive sleep apnea: the DREAM (Direct Referral for Apnea Monitoring) Project. *BMC Pulmonary Medicine* 2022;22(1):123.
183. Foo JY, Lim CS. Development of a home screening system for pediatric respiratory sleep studies. *Telemedicine Journal & E-Health* 2006;12(6):698-701.
184. Fynn TA, Alnemri A, Grewal RG, et al. Diagnostic Accuracy of Home Sleep Apnea Testing Based on Recording Duration. *American Journal of Respiratory and Critical Care Medicine* 2020;201
185. Gan YJ, Lim L, Chong YK. Validation study of WatchPat 200 for diagnosis of OSA in an Asian cohort. *European Archives of Oto-Rhino-Laryngology* 2017;274(3):1741-45.
186. Ganglberger W, Bucklin AA, Tesh RA, et al. Sleep apnea and respiratory anomaly detection from a wearable band and oxygen saturation. *Sleep & Breathing* 2022;26(3):1033-44.
187. Garg N, Rolle AJ, Lee TA, Prasad B. Home-based diagnosis of obstructive sleep apnea in an urban population. *Journal of Clinical Sleep Medicine* 2014;10(8):879-85.
188. Ghandeharioun H, Rezaeitalab F, Lotfi R. Accurate Estimation of Obstructive Sleep Apnea Severity Using Non-Polysomnographic Features For Home-Based Screening. *Iranian Journal of Public Health* 2015;44(10):1433-5.
189. Ghahjaverestan NM, Fan W, Aguiar C, et al. Respiratory Motion and Airflow Estimation During Sleep Using Tracheal Movement and Sound. *Nature and Science of Sleep* 2022;14:1213-23. doi: 10.2147/nss.S360970
190. Goldstein CA, Karnib H, Williams K, et al. The utility of home sleep apnea tests in patients with low versus high pre-test probability for moderate to severe OSA. *Sleep & Breathing* 2018;22(3):641-51.
191. Golpe R, Jimenez A, Carpizo R. Home sleep studies in the assessment of sleep apnea/hypopnea syndrome. *Chest* 2002;122(4):1156-61.
192. Green A, Nagel N, Kemer L, Dagan Y. Comparing in-lab full polysomnography for diagnosing sleep apnea in children to home sleep apnea tests (HSAT) with an online video attending technician. *Sleep & Biological Rhythms* 2022;20(3):397-401.
193. Gros P, Mery VP, Lafontaine AL, et al. Diagnosis of Obstructive Sleep Apnea in Parkinson's Disease Patients: Is Unattended Portable Monitoring a Suitable Tool? *Parkinsons Disease* 2015;2015:258418.
194. Guerrero A, Embid C, Isetta V, et al. Management of sleep apnea without high pretest probability or with comorbidities by three nights of portable sleep monitoring. *Sleep* 2014;37(8):1363-73.
195. Guerrero A, Masa JF, Embid C, Montserrat JM. Diagnosis, cost, and therapeutic decision-making of home respiratory poligraphy for patients without high suspicion of OSA or with comorbidity: hospital polysomography in comparison with three nights of home respiratory polygraphy. *Chest* 2014;145(3 MEETING ABSTRACT)
196. Gupta M, Ish P, Chakrabarti S, et al. Diagnostic accuracy and feasibility of portable sleep monitoring in patients with obstructive sleep apnea: Re-exploring the utility in the current COVID-19 pandemic. *Monaldi Archives for Chest Disease* 2021;92(1):14.

197. Gursoy TR, Eyuboglu TS, Aslan AT, et al. The reliability of polygraphy for the diagnosis of sleep disordered breathing in children. *European Respiratory Journal Conference: International Congress of the European Respiratory Society, ERS* 2021;58(SUPPL 65)
198. Gurubhagavatula I, Fields BG, Morales CR, et al. Screening for severe obstructive sleep apnea syndrome in hypertensive outpatients. *Journal of Clinical Hypertension* 2013;15(4):279-88.
199. Hafezi M, Montazeri N, Saha S, et al. Sleep Apnea Severity Estimation From Tracheal Movements Using a Deep Learning Model. *Ieee Access* 2020;8:22641-49. doi: 10.1109/access.2020.2969227
200. Hafezi M, Montazeri N, Zhu K, et al. Sleep Apnea Severity Estimation from Respiratory Related Movements Using Deep Learning. *Annual International Conference Of The IEEE Engineering In Medicine And Biology Society* 2019;2019:1601-04.
201. Hansen C, Sonnesen L, Markstrom A. Signal quality of home polygraphy in children and adolescents. *Acta Paediatrica* 2023;04:04.
202. Harris L, Swinson K. Home Sleep Testing for Diagnosing Obstructive Sleep Apnea. *American Family Physician* 2017;96(10):Online.
203. Hayano J, Yamamoto H, Nonaka I, et al. Quantitative detection of sleep apnea with wearable watch device. *PLoS ONE [Electronic Resource]* 2020;15(11):e0237279.
204. He K, Mendez M, Atwood CW. Comparison of sleep specialist vs primary care driven home sleep apnea testing in routine practice. *Sleep* 2020;43(SUPPL 1):A231.
205. He K, Kim R, Kapur VK. Home- vs. Laboratory-Based Management Of OSA: An Economic Review. *Current Sleep Medicine Reports* 2016;2(2):107-13.
206. Hedner J, Pillar G, Pittman SD, et al. A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep* 2004;27(8):1560-6.
207. Hilmisson H, Berman S, Magnusdottir S. Sleep apnea diagnosis in children using software-generated apnea-hypopnea index (AHI) derived from data recorded with a single photoplethysmogram sensor (PPG) Results from the Childhood Adenotonsillectomy Study (CHAT) based on cardiopulmonary coupling analysis. *Sleep and Breathing* 2020;24(4):1739-49. doi: 10.1007/s11325-020-02049-6
208. Holmedahl NH, Fjeldstad OM, Engan H, et al. Validation of peripheral arterial tonometry as tool for sleep assessment in chronic obstructive pulmonary disease. *Scientific Reports* 2019;9(1):19392.
209. Hung CJ, Kang BH, Lin YS, Su HH. Comparison of a home sleep test with in-laboratory polysomnography in the diagnosis of obstructive sleep apnea syndrome. *Journal of the Chinese Medical Association: JCMA* 2022;85(7):788-92.
210. Ikizoglu NB, Kiyan E, Polat B, et al. Are home sleep studies useful in diagnosing obstructive sleep apnea in children with down syndrome? *Pediatric Pulmonology* 2019;54(10):1541-46.
211. Ioachimescu OC, Dholakia SA, Venkateshiah SB, et al. Improving the performance of peripheral arterial tonometry-based testing for the diagnosis of obstructive sleep apnea. *Journal of Investigative Medicine* 2020;68(8):1370-78.
212. Ioan I, Renard E, Da Mota S, et al. Unattended home sleep studies for the diagnosis of obstructive sleep apnea in a population of French children. *Sleep Medicine* 2023;102:117-22.
213. Isrctn. A randomised controlled trial to identify Obstructive Sleep Apnoea (OSA) in primary care. *https://trialsearchwhoint/Trial2aspx?TrialID=ISRCTN16982033* 2023
214. Ito K, Ikeda T. Accuracy of Type III Portable Monitors for Diagnosing Obstructive Sleep Apnea. *Biomedicine Hub* 2018;3(2):1-10.
215. Jen R, Orr JE, Li Y, et al. Accuracy of WatchPAT for the Diagnosis of Obstructive Sleep Apnea in Patients with Chronic Obstructive Pulmonary Disease. *Copd: Journal of Chronic Obstructive Pulmonary Disease* 2020;17(1):34-39.
216. Jiang P, Zhao S, Zhu R. Smart Sensing Strip Using Monolithically Integrated Flexible Flow Sensor for Noninvasively Monitoring Respiratory Flow. *Sensors* 2015;15(12):31738-50. doi: 10.3390/s151229881

217. Jurado Gamez B, Redel Montero J, Munoz Cabrera L, et al. Cost-effectiveness and degree of satisfaction with home sleep monitoring in patients with symptoms of sleep apnea. *Archivos de Bronconeumologia* 2007;43(11):605-10.
218. Kalkbrenner C, Brucher R, Kesztyus T, et al. Automated sleep stage classification based on tracheal body sound and actigraphy. *German Medical Science* 2019;17:Doc02.
219. Kalkbrenner C, Eichenlaub M, Rudiger S, et al. Validation of a New System Using Tracheal Body Sound and Movement Data for Automated Apnea-Hypopnea Index Estimation. *Journal of Clinical Sleep Medicine* 2017;13(10):1123-30.
220. Kapur VK, Johnston JC, Rueschman M, et al. Patient satisfaction with sleep study experience: findings from the Sleep Apnea Patient-Centered Outcomes Network. *Sleep* 2018;41(8):01.
221. Kasai T, Takata Y, Yoshihisa A, et al. Comparison of the Apnea-Hypopnea Index Determined by a Peripheral Arterial Tonometry-Based Device With That Determined by Polysomnography - Results From a Multicenter Study. *Circulation Reports* 2020;2(11):674-81.
222. Khan KA, Akke V, Amoia M, et al. Precision and utility of an ambulatory sleep diagnostic system based on peripheral arterial tonometry (PAT) compared to simultaneous polysomnography in patients with OSAS. *Irish journal of medical science* 2009;178(Suppl 11):S465.
223. Kim RD, Kapur VK, Redline-Bruch J, et al. An Economic Evaluation of Home Versus Laboratory-Based Diagnosis of Obstructive Sleep Apnea. *Sleep* 2015;38(7):1027-37.
224. Kinoshita T, Yahaba M, Terada J, et al. Impact of Arterial Stiffness on WatchPAT Variables in Patients With Obstructive Sleep Apnea. *Journal of Clinical Sleep Medicine* 2018;14(3):319-25.
225. Kissow Lildal T, Boudewyns A, Kamperis K, et al. Validity of in-lab and home respiratory polygraphy for detecting obstructive sleep apnea in children. *Sleep Medicine* 2023;103:195-203.
226. Körkuyu E, Düzlü M, Karamert R, et al. The efficacy of Watch PAT in obstructive sleep apnea syndrome diagnosis. *European Archives of Oto-Rhino-Laryngology* 2015;272(1):111-16. doi: 10.1007/s00405-014-3097-0
227. Kotzian ST, Schwarzinger A, Haider S, et al. Home polygraphic recording with telemedicine monitoring for diagnosis and treatment of sleep apnoea in stroke (HOPES Study): Study protocol for a single-blind, randomised controlled trial. *BMJ Open* 2018;8(1) (no pagination)
228. Kukwa W, Laba J, Lis T, et al. Supine sleep patterns as a part of phenotyping patients with sleep apnea-a pilot study. *Sleep & Breathing* 2022;26(4):1771-78.
229. Kukwa W, Migacz E, Lis T, Ishman SL. The effect of in-lab polysomnography and home sleep polygraphy on sleep position. *Sleep & Breathing* 2021;25(1):251-55.
230. Kwon S, Kim HS, Kwon K, et al. At-home wireless sleep monitoring patches for the clinical assessment of sleep quality and sleep apnea. *Science Advances* 2023;9(21):eadg9671.
231. Lachapelle P, Khadadah S, Taseen R, et al. Scoring heart rate increases as a surrogate arousal marker on portable monitor studies for obstructive sleep apnea: Impact on diagnostic accuracy and clinical decision-making. *Journal of Sleep Research* 2022;31(5):e13594.
232. Le-Dong NN MJ, Coumans N, Tamisier R, Bailly S, Pépin JL. Machine Learning-based Sleep Staging in Patients With Sleep Apnea Using a Single Mandibular Movement Signal. *American Journal of Respiratory and Critical Care Medicine* 2021;204(10):1227-31. doi: 10.1164/rccm.202103-0680LE
233. Levendowski DJ, Veljkovic B, Seagraves S, Westbrook PR. Capability of a neck worn device to measure sleep/wake, airway position, and differentiate benign snoring from obstructive sleep apnea. *Journal of Clinical Monitoring & Computing* 2015;29(1):53-64.
234. Levendowski D, Steward D, Woodson BT, et al. The impact of obstructive sleep apnea variability measured in-lab versus in-home on sample size calculations. *International Archives of Medicine* 2009;2(1) (no pagination)
235. Li S, Xu L, Dong X, et al. Home sleep apnea testing of adults with chronic heart failure. *Journal of Clinical Sleep Medicine* 2021;17(7):1453-63.

236. Manoni A, Loreti F, Radicioni V, et al. A New Wearable System for Home Sleep Apnea Testing, Screening, and Classification. *Sensors* 2020;20(24):08.
237. Martinot J-B, Le-Dong N-N, Tamisier R, et al. Determinants of apnea-hypopnea index variability during home sleep testing. *Sleep Medicine* 2023;111:86-93. doi: https://doi.org/10.1016/j.sleep.2023.09.002
238. Martinot JB L-DN, Cutheberg V, Denison S, Silkoff P, Guénard H, Gozal D, Pepin J-L and Borel JC,. Mandibular Movements As Accurate Reporters Of Respiratory Effort During Sleep : Validation Against Diaphragmatic Electromyography. *Front Neurol* 2017;8 doi: 10.3389/fneur.2017.00353
239. Martinot J-B, Pépin J-L. Mandibular jaw movements as a non-invasive measure of respiratory effort during sleep: application in clinical practice. *Frontiers in Sleep* 2023;2 doi: 10.3389/frsle.2023.1145620
240. Martinot JB L-DN, Borel AL, Tamisier, Malhotra A, Pépin JL. Respiratory effort during sleep and the rate of prevalent type 2 diabetes in obstructive sleep apnea.
241. Martinot J-B BJ-C, Le-Dong NN, Guénard H, Cuthbert V, Silkoff P, Gozal D and Pepin J-L, . Monitoring Mandibular Movements To Detect Cheyne-Stokes Breathing. *Respiratory Research* 2017;18 doi: https://doi.org/10.1186/s12931-017-0551-8
242. Martinot JP, Le-Dong N, Cuthbert V, Pepin J. Mandibular Movements Are a Reliable Noninvasive Alternative to Esophageal Pressure for Identifying Respiratory Effort in Sleep Apnea. *American Journal of Respiratory and Critical Care Medicine Conference: American Thoracic Society International Conference, ATS* 2021;203(9)
243. Martinot JB, Le Dong NN, Cuthbert V, et al. Mandibular Movements Are a Reliable Noninvasive Alternative to Esophageal Pressure for Measuring Respiratory Effort in Patients with Sleep Apnea Syndrome. *Sleep* 2022;45(SUPPL 1):A321-A22.
244. Martinot JB, Le-Dong NN, Cuthbert V, et al. Mandibular movements are a reliable noninvasive alternative to esophageal pressure for measuring respiratory effort in patients with sleep apnea syndrome. *Sleep Medicine* 2022;100(Supplement 1):S258.
245. Martinot JB L-DN, Cuthbert V, Denison S, Gozal D, Lavigne G, Pépin JL. Artificial Intelligence Analysis of Mandibular Movements Enables Accurate Detection of Phasic Sleep Bruxism in OSA Patients: A Pilot Study *Nature and Science of Sleep* 2021;13:1449–59. doi: 10.2147/NSS.S320664
246. Martinot J-B L-DN, Denison S, Guénard H, Borel J-C, Silkoff P, Pepin JL, Gozal D. Persistent Respiratory Effort After Adenotonsillectomy In Children With Sleep-Disorded Breathing. *Laryngoscope* 2017;128:1230-37.
247. Martinot JB, Bailly S, Hostaux L, et al. Development of a digital medicine solution for homebased obstructive sleep apnea phenotyping. *European Respiratory Journal Conference: International Congress of the European Respiratory Society, ERS* 2021;58(SUPPL 65)
248. Martinot JB, Bailly S, Tamisier R, et al. Phenotypes in obstructive sleep apnea: new promises with a home-based digital medicine solution. *Sleep Medicine* 2022;100(Supplement 1):S265.
249. Martinot JB, Le Dong N, Denison S, et al. Validation of mandibular movements as a surrogate marker for sleep respiratory effort. *European respiratory journal* 2017;50 doi: 10.1183/1393003.congress-2017.PA2972
250. Martinot JB, Le-Dong NN, Nannan F, et al. Mandibular movements as a surrogate marker for sleep-related respiratory effort: a validation study against crural diaphragmatic electromyography. *American journal of respiratory and critical care medicine* 2017;195 doi: 10.1164/ajrccmconference.2017.C109
251. Masa JF, Corral J, De Cos JS, et al. Effectiveness of three sleep apnea management alternatives. *Sleep* 2013;36(12):1799-807.
252. Masa JF, Corral J, De Terreros JG, et al. Significance of including a surrogate arousal for sleep apnea-hypopnea syndrome diagnosis by respiratory polygraphy. *Sleep* 2013;36(2):249-57.

253. Masa JF, Corral J, Pereira R, et al. Therapeutic Decision For Sleep Apnea And Hypopnea Syndrome By Home Respiratory Polygraphy. *American journal of respiratory and critical care medicine* 2011;183(Meeting Abstracts):A6325.
254. Masa JF, Corral J, Pereira R, et al. Significance of including a subrogated arousal for sleep apnea-hypopnea syndrome diagnosis by respiratory polygrafy. *American journal of respiratory and critical care medicine* 2012;185
255. Masa JF, Corral J, Pereira R, et al. Effectiveness of home respiratory polygraphy for the diagnosis of sleep apnoea and hypopnoea syndrome. *Thorax* 2011;66(7):567-73.
256. Masa JF, Corral J, Pereira R, et al. Therapeutic decision-making for sleep apnea and hypopnea syndrome using home respiratory polygraphy: A large multicentric study. *American Journal of Respiratory and Critical Care Medicine* 2011;184(8):964-71.
257. Masa JF, Corral J, Pereira R, et al. Effectiveness of sequential automaticmanual home respiratory polygraphy scoring. *European Respiratory Journal* 2013;41(4):879-87.
258. Masa JF, Corral J, Pereira R, et al. Cost-Effectiveness of the teletransmission of home respiratory polygraphy for the diagnosis of sleep apnoea and hypopnoea syndrome. *Journal of sleep research* 2012;21(Suppl 1):49. doi: 10.1111/j.1365-2869.2012.01044.x
259. Masa JF, Corral J, Pereira R, et al. Therapeutic decision-making for sleep apnea and hypopnea syndrome using home respiratory polygraphy. *Sleep Biol Rhythms* 2011;9(4):416. doi: 10.1111/j.1479-8425.2011.00518.x
260. Masa JF, Corral J, Sanchez De Cos J, et al. Cost and effectiveness of three sleep apnea management alternatives. *American journal of respiratory and critical care medicine* 2013;187
261. Maury G, Cambron L, Jamart J, et al. Added value of a mandible movement automated analysis in the screening of obstructive sleep apnea. *J Sleep Res* 2013;22(1):96-103. doi: 10.1111/j.1365-2869.2012.01035.x [published Online First: 2012/07/28]
262. Mello AAF, G DA, Santos RB, et al. Influence of the device used for obstructive sleep apnea diagnosis on body position: a comparison between polysomnography and portable monitor. *Sleep and Breathing* 2023;27(3):887-91.
263. Michelet M, Blanchon S, Guinand S, et al. Successful home respiratory polygraphy to investigate sleep-disordered breathing in children. *Sleep Medicine* 2020;68:146-52.
264. Milici S, Lazaro A, Villarino R, et al. Wireless Wearable Magnetometer-Based Sensor for Sleep Quality Monitoring. *Ieee Sensors Journal* 2018;18(5):2145-52. doi: 10.1109/jsen.2018.2791400
265. Miller JN, Kupzyk KA, Zimmerman L, et al. Comparisons of measures used to screen for obstructive sleep apnea in patients referred to a sleep clinic. *Sleep Medicine* 2018;51:15-21.
266. Mohammadieh AM, Sutherland K, Kanagaratnam LB, et al. Clinical screening tools for obstructive sleep apnea in a population with atrial fibrillation: a diagnostic accuracy trial. *Journal of Clinical Sleep Medicine* 2021;17(5):1015-24.
267. Montazeri Ghahjaverestan N, Kabir M, Saha S, et al. Automatic Respiratory Phase Identification Using Tracheal Sounds and Movements During Sleep. *Annals of Biomedical Engineering* 2021;49(6):1521-33.
268. Montazeri Ghahjaverestan N, Saha S, Kabir MM, et al. Estimating Sleep Apnea Severity from Tracheal Signals Using Snore Related Features. *American Journal of Respiratory and Critical Care Medicine Conference: American Thoracic Society International Conference, ATS* 2021;203(9)
269. Morillo D, Rojas Ojeda JL, Crespo Foix LF, Jimenez A. An accelerometer-based device for sleep apnea screening. *IEEE Transactions on Information Technology in Biomedicine* 2010;14(2):491-9.
270. Morillo DS, Rojas Ojeda JL, Crespo Foix LF, Jimenez AL. An accelerometer-based device for sleep apnea screening. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society* 2010;14(2):491-99.

271. Mulgrew AT, Fox N, Ayas NT, Ryan CF. Ambulatory diagnosis and CPAP titration in obstructive sleep apnea patients: - can we take the pressure off the sleep lab. *European respiratory journal* 2005;26(Suppl 49):Abstract No. 805.
272. Munoz-Ferrer A, Cervantes MA, Garcia-Olive I, et al. In-Home Diagnosis of Obstructive Sleep Apnea Using Automatic Video Analysis. *Archivos de Bronconeumologia* 2020;56(11):704-09.
273. Nagubadi S, Mehta R, Abdoh M, et al. The Accuracy of Portable Monitoring in Diagnosing Significant Sleep Disordered Breathing in Hospitalized Patients. *PLoS ONE [Electronic Resource]* 2016;11(12):e0168073.
274. Nakano H, Hayashi M, Ohshima E, et al. Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome. *Sleep* 2004;27(5):951-7.
275. Nakase-Richardson R, Hoffman JM, Magalang U, et al. Cost-Benefit Analysis From the Payor's Perspective for Screening and Diagnosing Obstructive Sleep Apnea During Inpatient Rehabilitation for Moderate to Severe TBI. *Archives of Physical Medicine & Rehabilitation* 2020;101(9):1497-508.
276. Nakase-Richardson R, Schwartz DJ, Ketchum JM, et al. Comparison of Diagnostic Sleep Studies in Hospitalized Neurorehabilitation Patients With Moderate to Severe Traumatic Brain Injury. *Chest* 2020;158(4):1689-700.
277. Nct. Comparative Assessment of Compumedics Somfit Sleep Monitoring Device. 2022
278. Nct. Evaluation of WP200 With the Unified Probe (WP200U). 2013
279. Nct. Home Diagnosis and Treatment of Obstructive Sleep Apnea. *https://clinicaltrialsgov/show/NCT00139022* 2005
280. Nct. Evaluation of New Diagnostic and Treatment Algorithm for Obstructive Sleep Apnea. *https://clinicaltrialsgov/show/NCT00425659* 2007
281. Nct. Portable Monitoring for Diagnosis and Management of Sleep Apnea. *https://clinicaltrialsgov/show/NCT00642486* 2008
282. Nct. Cost Effectiveness of Ambulatory Management for Veterans With Sleep Apnea. *https://clinicaltrialsgov/show/NCT00880165* 2009
283. Nct. Domiciliary Diagnosis and Follow up in Obstructive Apnoea Syndrome. *https://clinicaltrialsgov/show/NCT01001858* 2009
284. Nct. Cost-effectiveness of Home Respiratory Polygraphy. *https://clinicaltrialsgov/show/NCT01752556* 2012
285. Nct. Sustainable Methods, Algorithms, and Research Tools for Delivering Optimal Care Study. *https://clinicaltrialsgov/show/NCT02037438* 2014
286. Nct. SLEep APnea Screening Using Mobile Ambulatory Recorders After TIA/Stroke. *https://clinicaltrialsgov/show/NCT02454023* 2015
287. Nct. Sleep Apnoea Management by a Communication Based Technology (ICT). *https://clinicaltrialsgov/show/NCT02779894* 2016
288. Nct. Validity of a Home Respiratory Polygraphy for the Diagnosis of Obstructive Sleep Apnea in Children (DINISAS). *https://clinicaltrialsgov/show/NCT03449550* 2017
289. Nct. ENhancing Outcomes in Cognitive Impairment Through Use of Home Sleep ApNea Testing. *https://clinicaltrialsgov/show/NCT04335994* 2020
290. Nct. Home Apnea Testing in CHildren Trial. *https://clinicaltrialsgov/show/NCT05382754* 2022
291. Nct. HSAT for OSA Management in Children. *https://clinicaltrialsgov/show/NCT05516524* 2022
292. Nct. The Use of Wearable Technology to Acquire Signals for COPD Research. 2020
293. Nct. Comparative Effectiveness Research to Enhance Outcomes in African-Americans With Obstructive Sleep Apnea. *https://clinicaltrialsgov/show/NCT01997723* 2013
294. Nct. Validation of the Peripheral Arterial Tone to Detect Sleep-disordered Breathing in Patients With Chronic Heart Failure. 2016
295. Nct. Screening Using Portable Electronic Recorders for Sleep Apnea in Hypertensive At-Risk Populations (SUPER-SHARP Trial). *https://clinicaltrialsgov/ct2/show/NCT05918120* 2023

296. O'Brien LM, Bullough AS, Shelgikar AV, et al. Validation of Watch-PAT-200 against polysomnography during pregnancy. *Journal of Clinical Sleep Medicine* 2012;8(3):287-94.
297. Oliveira MG, Treptow EC, Fukuda C, et al. Diagnostic Accuracy of Home-Based Monitoring System in Morbidly Obese Patients with High Risk for Sleep Apnea. *Obesity Surgery* 2014;25
298. Oliveira MG, Treptow EC, Fukuda C, et al. Diagnostic accuracy of home-based monitoring system in morbidly obese patients with high risk for sleep apnea. *Obesity Surgery* 2015;25(5):845-51.
299. Onder NS, Akpinar ME, Yigit O, Gor AP. Watch peripheral arterial tonometry in the diagnosis of obstructive sleep apnea: Influence of aging. *Laryngoscope* 2012;122(6):1409-14. doi: 10.1002/lary.23233
300. O'Reilly BM, Wang Q, Collen J, et al. Performance comparison of peripheral arterial tonometry-based testing and polysomnography to diagnose obstructive sleep apnea in military personnel. *Journal of Clinical Sleep Medicine* 2022;18(6):1523-30.
301. Ounhasuttiyanon A, Pimolsri C, Choeycheep P, et al. Diagnostic accuracy of Home sleep apnea testing (HSAT) in Thai population. *Journal of the Medical Association of Thailand* 2021;104(2):300-03.
302. Pallin M, O'Hare E, Zaffaroni A, et al. Comparison of a novel non-contact biomotion sensor with wrist actigraphy in estimating sleep quality in patients with obstructive sleep apnoea. *Journal of Sleep Research* 2014;23(4):475-84.
303. Parra O, Garcia-Esclasans N, Montserrat JM, et al. Should patients with sleep apnoea/hypopnoea syndrome be diagnosed and managed on the basis of home sleep studies? *European Respiratory Journal* 1997;10(8):1720-24.
304. Pei G, Ou Q, Lao M, et al. APAP Treatment Acceptance Rate and Cost-Effectiveness of Telemedicine in Patients with Obstructive Sleep Apnea: A Randomized Controlled Trial. *Nature & Science of Sleep* 2023;15:607-22.
305. Penzel T, Kesper K, Pinnow I, et al. Peripheral arterial tonometry, oximetry and actigraphy for ambulatory recording of sleep apnea. *Physiological Measurement* 2004;25(4):1025-36.
306. Penzel T, Kesper K, Ploch T, et al. Ambulatory recording of sleep apnea using peripheral arterial tonometry. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine & Biology Society* 2004;2004:3856-9.
307. Pillar G, Bar A, Betito M, et al. An automatic ambulatory device for detection of AASM defined arousals from sleep: the WP100. *Sleep Medicine* 2003;4(3):207-12.
308. Pinto JA, de Godoy LBM, Ribeiro RC, et al. Accuracy of peripheral arterial tonometry in the diagnosis of obstructive sleep apnea. *Brazilian Journal of Otorhinolaryngology* 2015;81(5):473-78.
309. Pittman SD, Ayas NT, MacDonald MM, et al. Using a wrist-worn device based on peripheral arterial tonometry to diagnose obstructive sleep apnea: in-laboratory and ambulatory validation. *Sleep* 2004;27(5):923-33.
310. Pittman SD, Tal N, Malhotra A, et al. Automated Detection of Obstructive Sleep-Disordered Breathing Events Using Peripheral Arterial Tonometry and Oximetry. *Computers in Cardiology* 2000;27:485-88.
311. Planes C, Leroy M, Bouach Khalil N, et al. Home diagnosis of obstructive sleep apnoea in coronary patients: validity of a simplified device automated analysis. *Sleep & Breathing* 2010;14(1):25-32.
312. Polese JF, Santos-Silva R, de Oliveira Ferrari PM, et al. Is portable monitoring for diagnosing obstructive sleep apnea syndrome suitable in elderly population? *Sleep & Breathing* 2013;17(2):679-86.
313. Prasad B, Carley D, Weaver TE, et al. Effectiveness of home portable monitoring in urban African-Americans. *American journal of respiratory and critical care medicine* 2012;185

314. Quintana-Gallego E, Villa-Gil M, Carmona-Bernal C, et al. Home respiratory polygraphy for diagnosis of sleep-disordered breathing in heart failure. *European Respiratory Journal* 2004;24(3):443-8.
315. Quiroga MAS, Jimenez JFM, Illa FB, et al. Efficacy and cost-effectiveness of home respiratory polygraphy. *European respiratory journal* 2016;48 doi: 10.1183/13993003.congress-2016.OA4797
316. Redline S, Tosteson T, Boucher MA, Millman RP. Measurement of sleep-related breathing disturbances in epidemiologic studies. Assessment of the validity and reproducibility of a portable monitoring device. *Chest* 1991;100(5):1281-6.
317. Revana A, Vecchio J, Guffey D, et al. Clinical application of home sleep apnea testing in children: a prospective pilot study. *Journal of Clinical Sleep Medicine* 2022;18(2):533-40.
318. Rocken J, Schumann DM, Herrmann MJ, et al. Peripheral arterial tonometry versus polysomnography in suspected obstructive sleep apnoea. *European Journal of Medical Research* 2023;28(1):251.
319. Rodriguez-Villegas E, Chen G, Radcliffe J, Duncan J. A pilot study of a wearable apnoea detection device. *BMJ Open* 2014;4(10):e005299.
320. Rosen CL, Auckley D, Benca R, et al. A multisite randomized trial of portable sleep studies and positive airway pressure autotitration versus laboratory-based polysomnography for the diagnosis and treatment of obstructive sleep apnea: the HomePAP study. *Sleep* 2012;35(6):757-67.
321. Roth R, Stafford P, Mehta N, et al. Accuracy of Watchpat Portable Sleep Monitoring and Sleep Assessment in Patients with Atrial Fibrillation. *Sleep* 2022;45(SUPPL 1):A315.
322. Device SEotS.
323. Scalzitti N, Hansen S, Maturo S, et al. Comparison of home sleep apnea testing versus laboratory polysomnography for the diagnosis of obstructive sleep apnea in children. *International Journal of Pediatric Otorhinolaryngology* 2017;100:44-51.
324. Schafer H, Ewig S, Hasper E, Luderitz B. Predictive diagnostic value of clinical assessment and nonlaboratory monitoring system recordings in patients with symptoms suggestive of obstructive sleep apnea syndrome. *Respiration* 1997;64(3):194-9.
325. Schindhelm F, Oldenburg O, Fox H, Bitter T. Measurement of peripheral arterial tone to detect sleep-disordered breathing in patients with heart failure. *Sleep and Breathing* 2023 doi: 10.1007/s11325-023-02923-z
326. Serra A, Cocuzza S, Maiolino L, et al. The watch-pat in pediatrics sleep disordered breathing: Pilot study on children with negative nocturnal pulse oximetry. *International Journal of Pediatric Otorhinolaryngology* 2017;97:245-50. doi: https://doi.org/10.1016/j.ijporl.2017.04.021
327. Skomro R, Urroz O, Reid J, et al. Outcomes of home diagnosis and treatment of obstructive sleep apnea - results of a randomised trial. *Sleep* 2006;29(Suppl):A158.
328. Stansbury R, Badami V, Rojas E, et al. Addressing rural health disparity with a novel hospital sleep apnea screening: Precision of a high-resolution pulse oximeter in screening for sleep-disordered breathing. *Sleep & Breathing* 2022;26(4):1821-28.
329. Stern J, Martinez-Kratz J, Jackson A. Experience with over 10,000 sleep telemedicine visits for diagnosis and treatment of OSAS. *Sleep Medicine* 2022;100(Supplement 1):S297.
330. Su M, Yu C, Zhang Y, et al. Clinical value of portable sleep testing in children with obstructive sleep apnea syndrome. *Zhonghua er ke za zhi Chinese journal of pediatrics* 2015;53(11):845 - 49.
331. MICRO-FABRICATED SKIN-LIKE ALN/POLYIMIDE/PDMS SENSOR FOR OBSTRUCTIVE SLEEP APNEA MONITORING. 33rd IEEE International Conference on Micro Electro Mechanical Systems (MEMS); 2020 Jan 18-22; Vancouver, CANADA.

332. Takama N, Kurabayashi M. Effectiveness of a portable device and the need for treatment of mild-to-moderate obstructive sleep-disordered breathing in patients with cardiovascular disease. *Journal of Cardiology* 2010;56(1):73-8.
333. Tanaka N, Okada M, Tanaka K, et al. Severity of sleep apnea in patients receiving catheter ablation of atrial fibrillation: A comparison of home sleep apnea test and polysomnography. *Circulation Conference: American Hearts Association's* 2021;144(SUPPL 1)
334. Tanphaichitr A, Thianboonsong A, Banhiran W, et al. Watch peripheral arterial tonometry in the diagnosis of pediatric obstructive sleep apnea. *Otolaryngology–Head and Neck Surgery* 2018;159(1):166-72.
335. Tedeschi E, Carratù P, Damiani MF, et al. Home unattended portable monitoring and automatic CPAP titration in patients with high risk for moderate to severe obstructive sleep apnea. *Respiratory care* 2013;58(7):1178 - 83. doi: 10.4187/respcare.01939
336. Tellez D, Uhles M, Alisic S, et al. Comparison of a Disposable Home Sleep Apnea Test to Polysomnography in Patients Referred for Obstructive Sleep Apnea Investigation. *Sleep* 2022;45(SUPPL 1):A305.
337. Ting J, Tan LL, Balakrishnan ID, et al. Watch-PAT versus level III device in diagnosing sleep disordered breathing in first myocardial infarction. *The clinical respiratory journal* 2018;12(8):2332-39.
338. Tondo P, Dell'Olio F, Lacedonia D, et al. A consumer wearable device for tracking sleep respiratory events. *Sleep and Breathing* 2023;27(4):1485-89.
339. Tondo P, Dell'Olio F, Lacedonia D, et al. A consumer wearable device for tracking sleep respiratory events. *Sleep & Breathing* 2022;15:15.
340. Tondo P, Drigo R, Scioscia G, et al. Usefulness of sleep events detection using a wrist worn peripheral arterial tone signal device (WatchPAT™) in a population at low risk of obstructive sleep apnea. *J Sleep Res* 2021;30(6):e13352. doi: 10.1111/jsr.13352
10.1111/jsr.13352. Epub 2021 Apr 12. [published Online First: 2021/04/13]
341. Tung PH, Hsieh MJ, Chuang LP, et al. Efficacy of portable sleep monitoring device in diagnosing central sleep apnea in patients with congestive heart failure. *Frontiers in Neurology* 2022;13 doi: 10.3389/fneur.2022.1043413
342. van Gilst MM, van Dijk JP, Krijn R, et al. Protocol of the SOMNIA project: an observational study to create a neurophysiological database for advanced clinical sleep monitoring. *BMJ Open* 2019;9(11):e030996.
343. van Veldhuisen SL, Nijland LMG, Ravesloot MJL, et al. Preoperative Assessment of Obstructive Sleep Apnea in Bariatric Patients Using Polysomnography or Polygraphy. *Obesity Surgery* 2022;32(6):1814-21.
344. Veitz S, Schumann D, Strobel W, et al. Peripherial arterial tonometry versus polysomnography in suspected obstructive sleep apnea. *European Respiratory Journal Conference: European Respiratory Society International Congress, ERS* 2020;56(Supplement 64)
345. Walter J, Lee JY, Blake S, et al. A new wearable diagnostic home sleep testing platform: comparison with available systems and benefits of multinight assessments. *Journal of Clinical Sleep Medicine* 2023;19(5):865-72.
346. Ward KL, McArdle N, James A, et al. A Comprehensive Evaluation of a Two-Channel Portable Monitor to "Rule in" Obstructive Sleep Apnea. *Journal of Clinical Sleep Medicine* 2015;11(4):433-44.
347. Watkins MR, Talmage JB, Thiese MS, et al. Correlation between screening for obstructive sleep apnea using a portable device versus polysomnography testing in a commercial driving population. *Journal of Occupational & Environmental Medicine* 2009;51(10):1145-50.
348. Weimin L, Rongguang W, Dongyan H, et al. Assessment of a portable monitoring device WatchPAT 200 in the diagnosis of obstructive sleep apnea. *European Archives of Oto-Rhino-Laryngology* 2013;270(12):3099-105.

349. Westenberg JN, Petrof BJ, Noel F, et al. Validation of home portable monitoring for the diagnosis of sleep-disordered breathing in adolescents and adults with neuromuscular disorders. *Journal of Clinical Sleep Medicine* 2021;17(8):1579-90.
350. Xian ZX, Wang X, Chen YC, Teng YS. Preliminary assessment of portable sleep monitoring for diagnosis of obstructive sleep apnea in children. *Sleep & Breathing* 2023;18:18.
351. Xu Y, Ou Q, Cheng Y, et al. Comparative study of a wearable intelligent sleep monitor and polysomnography monitor for the diagnosis of obstructive sleep apnea. *Sleep & Breathing* 2023;27(1):205-12.
352. Yadollahi A, Giannouli E, Moussavi Z. Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals. *Medical & Biological Engineering & Computing* 2010;48(11):1087-97.
353. Yadollahi A, Moussavi Z. Acoustic obstructive sleep apnea detection. *Annual International Conference Of The IEEE Engineering In Medicine And Biology Society* 2009;2009:7110-3.
354. Yamada Y, Kasagi S, Tamura H, et al. Clinical utility of a type 4 portable device for in-home screening of sleep disordered breathing. *Annals of Palliative Medicine* 2020;9(5):2895-902.
355. Yang S, De Goederen R, Bredero-Boelhouwer H, et al. Accuracy of detecting obstructive sleep apnea using ambulatory sleep studies in patients with syndromic craniosynostosis. *Cogent Medicine Conference: 13th Excellence in Pediatrics Conference, EiP* 2021;8(no pagination)
356. Yang S, de Goederen R, Bredero-Boelhouwer H, et al. Accuracy of Detecting Obstructive Sleep Apnea Using Ambulatory Sleep Studies in Patients With Syndromic Craniosynostosis. *Journal of Craniofacial Surgery* 2022;33(8):2538-42.
357. Yoon H, Hwang SH, Choi SH, et al. Wakefulness evaluation during sleep for healthy subjects and OSA patients using a patch-type device. *Computer Methods & Programs in Biomedicine* 2018;155:127-38.
358. Zandieh S, Kirschenbaum MA, Greenberg H, Ancoli-Israel S. Keep it simple: A novel technique for measuring airflow using a wireless patch. *Sleep Health* 2023;9(1):100-07.
359. Zhang Z, Sowho M, Otvos T, et al. A comparison of automated and manual sleep staging and respiratory event recognition in a portable sleep diagnostic device with in-lab sleep study. *Journal of Clinical Sleep Medicine* 2020;16(4):563-73.
360. Zou D, Grote L, Peker Y, et al. Validation a portable monitoring device for sleep apnea diagnosis in a population based cohort using synchronized home polysomnography. *Sleep* 2006;29(3):367-74.
361. Arguelles J CJ, Henry L, Villalpando C, Romero M, Montoya E, Becker K, Kim JB, Derose S, Hwang D. Comparing Three Home Sleep Apnea Testing Devices to Polysomnography: Evaluating Diagnostic Rates for Obstructive Sleep Apnea. *Sleep* 2019;42:A181. doi: https://doi.org/10.1093/sleep/zsz067.450
362. Campbell CD, Cronin E, Kane T, Sulaiman I. Comparison of the Watchpat Device With Home Limited Sleep Studies for the Diagnosis of Obstructive Sleep Apnoea. *American Journal of Respiratory and Critical Care Medicine* 2023;207
363. Campbell CD, Cronin E, Kane T, Sulaiman I. Comparison of the WatchPAT device with Home Limited Sleep Studies for the Diagnosis of Obstructive Sleep Apnoea. *Irish Journal of Medical Science* 2021;190(SUPPL 5):195-95.
364. Fildes A KA. WatchPat: Comparison to in-lab overnight polysomnography in patients with suspected obstructive sleep apnoea. BSS (British Sleep Society) Scientific Conference. Birmingham, England, 2019.
365. Groshaeny L, Gomez E, Oranger M, et al. Type III sleep test at home versus Type I polysomnography in the laboratory before bariatric surgery. *European Respiratory Journal Conference: European Respiratory Society International Congress, ERS* 2022;60(Supplement 66)

366. Jen R, Li Y, DeYoung P, et al. Comparison of peripheral arterial tonometry and polysomnography for the diagnosis of OSA in patients with chronic obstructive pulmonary disease. *Sleep* 2017;40:A177.
367. Jenad H, Junna M, Selim B, Park J. PERIPHERAL ARTERIAL TONOMETRY-BASED HOME SLEEP APNEA TEST VS. POLYSOMNOGRAPHY IN DETECTION OF CENTRAL DISORDERED BREATHING EVENTS. *Sleep* 2023;46
368. Kelly B, Moua T, Escalante P. Diagnosis of Sleep Disordered Breathing in Patients with Interstitial Lung Disease: A Retrospective Evaluation of Polysomnogram and Home Sleep Apnea Testing Using Peripheral Arterial Tonometry. *Sleep* 2022;45(SUPPL 1):A313.
369. Martinot J, Le-Dong N, Letesson C, et al. Design and validation of a new technology for home-based screening of sleep apnea syndrome. *American Journal of Respiratory and Critical Care Medicine Conference: American Thoracic Society International Conference, ATS* 2020;201(1)
370. Maury G CL, Jamart J, Marchand E, Senny F, Poirrier R. Added Value Of A Mandible Movement Automated Analysis In The Screening Of Obstructive Sleep Apnea. *Journal of Sleep Research* 2013;22:96-103. doi: https://doi.org/10.1111/j.1365-2869.2012.01035.x
371. Morse CD, Meissner S, Kodali L. Convenience vs accuracy: Negative predictability of the home sleep test in clinical practice. *Sleep* 2020;43(SUPPL 1):A225.
372. Nct. WatchPAT Device Validation Study Compared to Polysomnography (WPAT). 2017
373. Nct. WatchPAT Compared to PSG at 8500'. 2020
374. Nct. Validation of Home Sleep Testing (WP) Compared to an Overnight Sleep Testing in the Sleep Laboratory. 2012
375. Orbea CP, McIntire C, Wang L, Mehra R. Influence of sex-specific differences in inpatient sleep testing approach for diagnosis of obstructive sleep apnea. *Sleep* 2021;44(SUPPL 2):A167.
376. Pang KP, Gourin CG, Terris DJ. A comparison of polysomnography and the WatchPAT in the diagnosis of obstructive sleep apnea. *Otolaryngology - Head & Neck Surgery* 2007;137(4):665-8.
377. Safadi A, Etzioni T, Fliss D, et al. The Effect of the Transition to Home Monitoring for the Diagnosis of OSAS on Test Availability, Waiting Time, Patients' Satisfaction, and Outcome in a Large Health Provider System. *Sleep Disorders Print* 2014;2014:418246.
378. Senny F, Destine J, Poirrier R. Midsagittal jaw movement analysis for the scoring of sleep apneas and hypopneas. *IEEE Transactions on Biomedical Engineering* 2008;55(1):87-95.
379. Senny F, Destine J, Poirrier R. Midsagittal jaw movements as a sleep/wake marker. *IEEE Transactions on Biomedical Engineering* 2009;56(2):303-9.
380. Townsend D SA, Brauer E, Scattarelli D, McEiver J, Eiken T. Assessing efficacy, outcomes, and cost savings for patients with obstructive sleep apnea using two diagnostic and treatment strategies. *Sleep Diagn Therapy* 2007(7):1-8.
381. Waicus S, Moloney E, Byrne P, et al. Diagnosis of Obstructive Sleep Apnoea Using Home Sleep Testing Compared to Polysomnography: A Service Evaluation. *Irish Journal of Medical Science* 2022;191(Supplement 5):S154.
382. Nct. The Sunrise OSA Trial. *https://clinicaltrialsgov/show/NCT05204004* 2022
383. Nct. Clinical Evaluation of a Wearable Sleep Diagnosis Technology. 2018
384. Nct. Sunrise a New Medical Device to Diagnose Sleep Apnea Syndrome (SENSAPNEA). 2020
385. Lyne CJ, Hamilton GS, Turton ARE, et al. Validation of a single-use and reusable home sleep apnea test based on peripheral arterial tonometry compared to laboratory polysomnography for the diagnosis of obstructive sleep apnea. *Journal of Clinical Sleep Medicine* 2023;20:20.
386. Martinot JB PJ, Malhotra A, Le-Dong N. Near-boundary Double-labelling Based Classification: The New Standard When Evaluating Performances of New Sleep Apnoea Diagnosis Solution Against Polysomnography? *Sleep* 2022;45(10) doi: https://doi.org/10.1093/sleep/zsac188
387. Nct. Validation of Ambulatory Sleep Test (WP200/U) Compared In-lab Sleep Testing in Cardiac Subjects. 2015

388. Nct. A Validation Study of the NightOwl PAT-based Home Sleep Apnea Test. 2019
389. Mills EW, Cassidy M, Sofer T, et al. Evaluation of obstructive sleep apnea among consecutive patients with all patterns of atrial fibrillation using WatchPAT home sleep testing. *American heart journal* 2023;261:95‐103. doi: 10.1016/j.ahj.2023.03.014
390. Ssegonja R, Ljunggren M, Sampaio F, et al. Economic evaluation of telemonitoring as a follow-up approach for patients with obstructive sleep apnea syndrome starting treatment with continuous positive airway pressure. *Journal of Sleep Research* 2023:e13968.
391. Perleth M, Leyen Uvd, Schmitt H, et al. Diagnosis and treatment of sleep apnea - systematic review of diagnostics, therapy, and cost-effectiveness. Germany: German Agency for Health Technology Assessment at the German Institute for Medical Documentation and Information (DAHTA DIMDI), 2003.
392. Ayas NT, Fox J, Epstein L, et al. Initial use of portable monitoring versus polysomnography to confirm obstructive sleep apnea in symptomatic patients: an economic decision model. *Sleep Med* 2010;11(3):320-4. doi: 10.1016/j.sleep.2009.08.015
10.1016/j.sleep.2009.08.015. Epub 2010 Feb 4. [published Online First: 2010/02/06]
393. Ayas NT, Jen R, Baumann B. Revisiting level II sleep studies in the era of COVID-19: a theoretical economic decision model in patients with suspected obstructive sleep apnea. *Sleep Sci Pract* 2021;5(1):11. doi: 10.1186/s41606-021-00063-5
10.1186/s41606-021-00063-5. Epub 2021 Jul 15. [published Online First: 2021/07/27]
394. Barriuso B, Martin L, Sevilla C, et al. Self-setup of home respiratory polygraphy for the diagnosis of sleep apnea syndrome: cost-efficiency study. *Sleep Breath* 2020;24(3):1067-74. doi: 10.1007/s11325-019-01959-4
10.1007/s11325-019-01959-4. Epub 2019 Nov 30. [published Online First: 2019/12/02]
395. Bravata DM, Lightner N, Yaggi HK, Miech EJ. Economic Assessment of 4 Approaches to the Diagnosis and Initial Treatment of Sleep Apnea. *Respir Care* 2018;63(1):50-61. doi: 10.4187/respcare.05355
10.4187/respcare.05355. Epub 2017 Oct 24. [published Online First: 2017/10/27]
396. Deutsch PA, Simmons MS, Wallace JM. Cost-effectiveness of split-night polysomnography and home studies in the evaluation of obstructive sleep apnea syndrome. *J Clin Sleep Med* 2006;2(2):145-53. [published Online First: 2007/06/15]
397. Fletcher EC, Stich J, Yang KL. Unattended home diagnosis and treatment of obstructive sleep apnea without polysomnography. *Arch Fam Med* 2000;9(2):168-74. doi: 10.1001/archfami.9.2.168
10.1001/archfami.9.2.168. [published Online First: 2000/02/29]
398. Pietzsch JB, Garner A, Cipriano LE, Linehan JH. An integrated health-economic analysis of diagnostic and therapeutic strategies in the treatment of moderate-to-severe obstructive sleep apnea. *Sleep* 2011;34(6):695-709. doi: 10.5665/sleep.1030
10.5665/SLEEP.1030. [published Online First: 2011/06/02]
399. Veloso IL, Corrêa CC, Tagliarini JV, Weber SAT. Unsupervised type III polygraphy in children undergoing adenotonsillectomy: a technical and economic report. *Sleep Sci* 2021;14(4):370-74. doi: 10.5935/1984-0063.20200094
10.5935/1984-0063.20200094. [published Online First: 2022/01/29]
400. Natsky AN, Vakulin A, Coetzer CLC, et al. Economic evaluation of diagnostic sleep studies for obstructive sleep apnoea: a systematic review protocol. *Syst Rev* 2021;10(1):104. doi: 10.1186/s13643-021-01651-3
10.1186/s13643-021-01651-3. [published Online First: 2021/04/11]
401. Natsky AN, Chai-Coetzer CL, Vakulin A, et al. Economic evaluation of diagnostic sleep studies for obstructive sleep apnoea in the adult population: a systematic review. *Sleep Med Rev* 2022;62:101608. doi: 10.1016/j.smrv.2022.101608
10.1016/j.smrv.2022.101608. Epub 2022 Feb 10. [published Online First: 2022/03/08]

402. Masa JF, Corral J, Pereira R, et al. Effectiveness of home respiratory polygraphy for the diagnosis of sleep apnoea and hypopnoea syndrome. *Thorax* 2011;66(7):567-73. doi: 10.1136/thx.2010.152272
10.1136/thx.2010.152272. Epub 2011 May 20. [published Online First: 2011/05/24]
403. Nct. Validity of a Home Respiratory Polygraphy for the Diagnosis of Obstructive Sleep Apnea. *https://clinicaltrialsgov/show/NCT03415633* 2016
404. Atwood CW, Kuna ST, Little KC, et al. Cost-effectiveness of home management of obstructive sleep apnea: the veterans sleep apnea treatment trial (Abstract). *American journal of respiratory and critical care medicine* 2014;189:A3672.
405. Krishnan S, Chai-Coetzer CL, Grivell N, et al. Comorbidities and quality of life in Australian men and women with diagnosed and undiagnosed high-risk obstructive sleep apnea. *Journal of Clinical Sleep Medicine* 2022;18(7):1757-67.
406. Wireklint S, Ericsson E. Health-related quality of life after tonsillotomy versus tonsillectomy in young adults: 6 years postsurgery follow-up. *European archives of oto-rhino-laryngology* 2012;269(8):1951‐58. doi: 10.1007/s00405-012-1990-y
407. Jennum P, Coaquira Castro J, Mettam S, et al. Socioeconomic and humanistic burden of illness of excessive daytime sleepiness severity associated with obstructive sleep apnoea in the European Union 5. *Sleep Medicine* 2021;84:46-55.
408. McMillan A, Bratton DJ, Faria R, et al. A 12 month multicenter, parallel, randomized trial of continuous positive airway pressure in older people with obstructive sleep apnea syndrome. *American journal of respiratory and critical care medicine* 2014;189(no pagination)
409. de Vries GE, Hoekema A, Vermeulen KM, et al. Clinical- and Cost-Effectiveness of a Mandibular Advancement Device Versus Continuous Positive Airway Pressure in Moderate Obstructive Sleep Apnea. *Journal of Clinical Sleep Medicine* 2019;15(10):1477-85.
410. Tarraubella N, Sanchez-de-la-Torre M, Nadal N, et al. Management of obstructive sleep apnoea in a primary care vs sleep unit setting: a randomised controlled trial. *Thorax* 2018;73(12):1152-60.
411. Batool-Anwar S, Quan S. The effect of continuous positive airway pressure (CPAP) on health related quality of life (HRQOL) as measured by quality of well being self administered questionnaire (QWB-SA). *Sleep* 2018;41:A218‐A19.
412. Amin N, Bhargava E, Prentice JG, et al. Coblation intracapsular tonsillectomy in children: A prospective study of 1257 consecutive cases with long-term follow-up. *Clinical Otolaryngology* 2021;46(6):1184-92.
413. Betavani VMP, Davey MJ, Nixon GM, et al. Effects of Treatment of Sleep Disordered Breathing on Sleep Macro- and Micro-Architecture in Children with Down Syndrome. *Children* 2022;9(7):30.
414. Carvalho FR, Lentini-Oliveira DA, Prado LBF, et al. Oral appliances and functional orthopaedic appliances for obstructive sleep apnoea in children. *Cochrane Database of Systematic Reviews* 2016(10) doi: 10.1002/14651858.CD005520.pub3
415. Clements AC, Dai X, Walsh JM, et al. Outcomes of Adenotonsillectomy for Obstructive Sleep Apnea in Prader-Willi Syndrome: Systematic Review and Meta-analysis. *Laryngoscope* 2021;131(4):898-906.
416. Decuzzi J, Redline S, Isaiah A. Secondary Analyses of the Childhood Adenotonsillectomy Trial A Narrative Review. *Jama Otolaryngology-Head & Neck Surgery* 2022;148(8):779-84. doi: 10.1001/jamaoto.2022.1330
417. End C, Propst EJ, Cushing SL, et al. Risks and Benefits of Adenotonsillectomy in Children With Cerebral Palsy With Obstructive Sleep Apnea: A Systematic Review. *Laryngoscope* 2022;132(3):687-94.
418. Hamid AAAE, Askoura AM, Hamed DMA, et al. Surgical versus non-surgical management of obstructive sleep-disordered breathing in children: A meta-analysis. *Open Respiratory Medicine Journal* 2020;14(1):47-52.

419. Idris G, Galland B, Robertson CJ, et al. Mandibular advancement appliances for sleep-disordered breathing in children: A randomized crossover clinical trial. *Journal of Dentistry* 2018;71:9-17.
420. Jackman AR, Biggs SN, Walter LM, et al. Sleep disordered breathing in early childhood: quality of life for children and families. *Sleep* 2013;36(11):1639-46.
421. Kaditis AG, Gozal D. Adenotonsillectomy: the good, the bad and the unknown. *Current Opinion in Pulmonary Medicine* 2022;28(6):537-42.
422. Kanona H, Sharma S, Chaidas K, Kotecha B. Pulse oximetry in paediatric obstructive sleep apnoea: is it used appropriately? *Journal of Laryngology and Otology* 2015;129(9):874-81. doi: 10.1017/s0022215115001814
423. Kao SS, Peters MDJ, Dharmawardana N, et al. Scoping review of pediatric tonsillectomy quality of life assessment instruments. *Laryngoscope* 2017;127(10):2399-406.
424. Nazir Othman M, Goh BS, Hasniah AL. Impact of adenotonsillectomy on the quality of life in children with sleep disordered breathing. *International Journal of Pediatric Otorhinolaryngology* 2016;91:105-07.
425. Roy R, Banger S, Singh SK, et al. Post Surgical Outcomes in Paediatric Adenotonsillar Hypertrophy with Obstructive Sleep Apnea: Subjective and Objective Evaluation. *Indian Journal of Otolaryngology & Head & Neck Surgery* 2023;75(Suppl 1):822-27.
426. Saju S, Bharathi MB, Prakash BG, Rakesh BS. Assessment of Quality of Life and Symptoms in Non Syndromic Paediatric Obstructive Sleep Apnoea Patients: Before and After Adenotonsillectomy. *Indian Journal of Otolaryngology & Head & Neck Surgery* 2022;74(Suppl 2):1453-58.
427. Todd CA, Bareiss AK, McCoul ED, Rodriguez KH. Adenotonsillectomy for Obstructive Sleep Apnea and Quality of Life: Systematic Review and Meta-analysis. *Otolaryngology - Head & Neck Surgery* 2017;157(5):767-73.
428. Torretta S, Rosazza C, Pace ME, et al. Impact of adenotonsillectomy on pediatric quality of life: review of the literature. *Italian Journal of Pediatrics* 2017;43(1):107.
429. Zhang LY, Zhong L, David M, Cervin A. Tonsillectomy or tonsillotomy? A systematic review for paediatric sleep-disordered breathing. *International Journal of Pediatric Otorhinolaryngology* 2017;103:41-50.
430. Bhushan B, Zee PC, Grandner MA, et al. Associations of deep sleep and obstructive sleep apnea with family relationships, life satisfaction and physical stress experience in children: a caregiver perspective. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine* 2023;14
431. Bironneau V, Ingrand P, Pontier S, et al. Auto-adjusted versus fixed positive airway pressure in patients with severe OSA: A large randomized controlled trial. *Respirology* 2023 doi: 10.1111/resp.14569
432. Chen M, Shi F, Wu H, et al. Impact of Obstructive Sleep Apnea on Health-Related Quality of Life in Patients With Partial Laryngectomy for Laryngeal Cancer. *Ear, Nose, & Throat Journal* 2023:1455613231178955.
433. De Weerdt S, Schotte C, Barbe K, et al. The DS-14 questionnaire: psychometric characteristics and utility in patients with obstructive sleep apnea. *Sleep & Breathing* 2023;07:07.
434. Dundervill C, Al-Asadi Z, Behnke J, et al. Gender Differences in Quality of Life of Adolescent Patients With Chronic Rhinosinusitis. *Annals of Otology, Rhinology and Laryngology* 2023
435. Horne RSC, Shetty M, Davey MJ, et al. Follow-up of children with Down syndrome and sleep disordered breathing and the effects of treatment on actigraphically recorded sleep, quality of life, behaviour, and daytime functioning. *Journal of Sleep Research* 2023:e13970.
436. Jensen S, Abeler K, Friborg O, et al. Sleep disorders in FKRP-related limb-girdle muscular dystrophy R9. *Neuromuscular Disorders* 2023;33(Supplement 1):S164-S65.

437. Jensen S, Abeler K, Friborg O, et al. Insomnia and sleep-disordered breathing in FKRP-related limb-girdle muscular dystrophy R9. The Norwegian LGMDR9 cohort study (2020). *Journal of Neurology* 2023
438. Lee MR, Jung SM. Obstructive sleep apnea related to mental health, health-related quality of life and multimorbidity: A nationwide survey of a representative sample in Republic of Korea. *PLoS ONE [Electronic Resource]* 2023;18(6):e0287182.
439. Luz GP, Badke L, Nery LE, et al. Effect of CPAP vs. mandibular advancement device for excessive daytime sleepiness, fatigue, mood, sustained attention, and quality of life in patients with mild OSA. *Sleep and Breathing* 2023;27(3):991-1003.
440. Samota P, Singh A, Aggarwal A, Malhotra R. Sleep Disorders and Quality of Life in Children with Cerebral Palsy. *Indian Journal of Pediatrics* 2023;90(6):560-65.
441. Schwartz AR, Jacobowitz O, Mickelson SA, et al. Sustained Therapeutic Benefits for at Least 3 Years in the THN3 Randomized, Controlled Trial of Targeted Hypoglossal Nerve Stimulation for Obstructive Sleep Apnea. *American journal of respiratory and critical care medicine* 2023;207(1) doi: 10.1164/ajrccm-conference.2023.A18
442. Shen C, Ou Y, Ouyang R, Zong D. Prevalence and characteristics of pain in moderate-to-severe obstructive sleep apnea patients and effect of CPAP treatment. *Scientific Reports* 2023;13(1):15758.
443. Sudarsan SS, Lohith Kumar MK, Bhumireddy Chalamaiahgari S, Kandaswamy S. Covid-19 and Obstructive Sleep Apnea: The Theatricals and Quality-of-Life Impact. *Indian Journal of Otolaryngology & Head & Neck Surgery* 2023:1-11.
444. Vilarrasa N, Miralles L, Monasterio C, et al. Effectiveness of an integrated hospital-primary care lifestyle change program in the treatment of obstructive sleep apnea syndrome and obesity. *Obesity Facts* 2023;16(Supplement 1):245.
445. Vilma Fidelis-da Silva L, Henrique Kiemle-Trindade S, Noel Marzano-Rodrigues M, et al. Obstructive Sleep Apnea in Adults with Treacher Collins Syndrome is Related with Altered Anthropometric Measurements, Increased Blood Pressure and Impaired Quality of Life. *Sleep Science* 2023;16(2):139-47.
446. Xiao L, Voutsas G, Ryan CM, et al. The association between sleep quality and obstructive sleep apnea with health-related quality of life in children with obesity. *Journal of Clinical Sleep Medicine* 2023;06:06.

9 APPENDICES

Appendix 1 Literature search strategies for the systematic reviews of clinical effectiveness, cost effectiveness and HRQoL

All the database search strategies for the clinical effectiveness, cost-effectiveness and HRQoL searches are reported below. Each strategy was first developed in MEDLINE (Ovid) and then adapted for the other databases. Reference management and deduplication of search results were carried out in EndNote™ (Clarivate™).

1a Searches for diagnostic test evaluation and clinical effectiveness studies

The searches for diagnostic test evaluation and clinical effectiveness had no date limits, the databases were searched from inception, and we applied an English language limit. In order to be sensitive and retrieve all relevant studies, we did not use any study design search filters. Table 52 below details the search strategies for the databases. See also section 3.1 of this report.

Table 52 Search strategies for diagnostic accuracy/efficacy and clinical effectiveness

Database, Host, Years searched, Date searched	Literature search strategy	Results
Ovid MEDLINE(R) ALL 1946 to May 22, 2023 Date of original search: 23/05/2023 Date of update search: 25/09/2023	1 sleep apnea syndromes/ or sleep apnea, obstructive/ 2 (sleep* adj4 hypopn?ea*).ti,ab,kf. 3 ("obstructive sleep*" adj apn?ea*).ti,ab,kf. 4 (sleep* adj4 disorder* adj4 breath*).ti,ab,kf. 5 (OSA or SDB or OSAS or OSAHS).ti,ab,kf. 6 1 or 2 or 3 or 4 or 5 7 Actigraphy/ 8 (actigraph* or "sleep monitor*" or accelerometer).ti,ab,kf. 9 exp Oximetry/ 10 (oxymet* or oximet*).ti,ab,kf. 11 "oxygen desaturation".ti,ab,kf. 12 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*).ti,ab,kf. 13 Capnography/ 14 (capnogra* or ((CO2 or "carbon dioxide") adj1 monitor*)).ti,ab,kf. 15 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT).ti,ab,kf. 16 Mobile Applications/ 17 ("limited channel*" or limited-channel* or multichannel or multi-channel).ti,ab,kf.	Original search: 1790 Update Search:49

	18 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 19 Monitoring, Ambulatory/ and (test* or device* or detect* or identif* or diagnos* or screen*).ti,ab,kf. 20 (home or at-home or home-based or unattended or portable or ambulatory).ti,ab,kf. 21 19 or 20 22 18 and 21 23 Monitoring, Ambulatory/ and (test* or device* or detect* or identif* or diagnos* or screen*).ti,ab,kf. 24 (((home or at-home or home-based or unattended or portable or ambulatory) adj3 (test* or device* or detect* or identif* or diagnos* or screen*)) or HSAT).ti,ab,kf. 25 Wearable Electronic Devices/ and (test* or device* or detect* or identif* or diagnos* or screen*).ti,ab,kf. 26 (((wearable* or nearable*) adj3 (test* or device* or detect* or identif* or diagnos* or screen*)) or WADD).ti,ab,kf. 27 (Acupebble or Acurable).ti,ab,kf. 28 (Brizzy or JAWAC or Nomics).ti,ab,kf. 29 (NightOwl or Ectosense or ResMed).ti,ab,kf. 30 Sunrise.ti,ab,kf. 31 (WatchPAT or Itamar or Zoll).ti,ab,kf. 32 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 33 6 and 32 34 (CPAP or "continuous positive airway pressure").ti. 35 33 not 34 36 letter/ 37 editorial/ 38 news/ 39 exp historical article/ 40 Anecdotes as Topic/ 41 comment/ 42 case reports/ 43 (letter or comment*).ti. 44 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 45 randomized controlled trial/ or random*.ti,ab. 46 44 not 45 47 animals/ not humans/ 48 exp Animals, Laboratory/ 49 exp Animal Experimentation/ 50 exp Models, Animal/ 51 exp Rodentia/ 52 (rat or rats or mouse or mice).ti. 53 46 or 47 or 48 or 49 or 50 or 51 or 52 54 35 not 53 55 limit 54 to english language 56 remove duplicates from 55	
Ovid Embase Classic+Embase 1947 to 2023 May 22	1 sleep disordered breathing/ 2 (sleep* adj4 hypopn?ea*).ti,ab,kf. 3 ("obstructive sleep*" adj apn?ea*).ti,ab,kf.	Original search: 2520

Date of original search: 23/05/2023 Date of update search: 25/09/2023	4 (sleep* adj4 disorder* adj4 breath*).ti,ab,kf. 5 (OSA or SDB or OSAS or OSAHS).ti,ab,kf. 6 1 or 2 or 3 or 4 or 5 7 actimetry/ 8 (actigraph* or "sleep monitor*" or accelerometer).ti,ab,kf. 9 oximetry/ or transcutaneous oxygen monitoring/ 10 (oxymet* or oximet*).ti,ab,kf. 11 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*).ti,ab,kf. 12 capnometry/ 13 (capnogra* or ((CO2 or "carbon dioxide") adj1 monitor*)).ti,ab,kf. 14 apnea monitor/ 15 polysomnograph/ 16 exp pulse oximeter/ 17 peripheral arterial tonometry device/ 18 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT).ti,ab,kf. 19 mobile application/ or mobile health application/ 20 ("limited channel*" or limited-channel* or multichannel or multi-channel).ti,ab,kf. 21 (home* or at-home or home-based or unattended or portable or ambulatory).ti,ab,kf. 22 portable equipment/ 23 portable patient monitor/ 24 (7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20) and (21 or 22 or 23) 25 (((home or at-home or home-based or unattended or portable or ambulatory) adj3 (test* or device* or detect* or identif* or diagnos* or screen*)) or HSAT).ti,ab,kf. 26 portable patient monitor/ 27 wearable sensor/ and (test* or device* or detect* or identif* or diagnos* or screen*).ti,ab,kf. 28 (((wearable* or nearable*) adj3 (test* or device* or detect* or identif* or diagnos* or screen*)) or WADD).ti,ab,kf. 29 (((home or at-home or home-based or unattended) adj3 (polygraph* or polysomnograph*)) or HRP).ti,ab,kf. 30 home monitoring/ and (test* or device* or detect* or identif* or diagnos* or screen*).ti,ab,kf. 31 home sleep apnea testing/ 32 "home use apnea monitor"/ 33 (Acupebble or Acurable).ti,ab,kf. 34 (Brizzy or JAWAC or Nomics).ti,ab,kf. 35 (NightOwl or Ectosense or ResMed).ti,ab,kf. 36 Sunrise.ti,ab,kf. 37 (WatchPAT or Itamar or Zoll).ti,ab,kf. 38 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 39 6 and 38 40 (CPAP or "continuous positive airway pressure").ti.	Update search: 120

	41 39 not 40 42 letter.pt. or letter/ 43 note.pt. 44 editorial.pt. 45 case report/ or case study/ 46 (letter or comment*).ti. 47 or/42-46 48 randomized controlled trial/ or random*.ti,ab. 49 47 not 48 50 animal/ not human/ 51 nonhuman/ 52 exp Animal Experiment/ 53 exp Experimental Animal/ 54 animal model/ 55 exp Rodent/ 56 (rat or rats or mouse or mice).ti. 57 or/49-56 58 41 not 57 59 (conference abstract* or conference review or conference paper or conference proceeding).db,pt,su. 60 limit 59 to yr="2020 -Current" 61 59 not 60 62 58 not 61 63 limit 62 to english language 64 remove duplicates from 63	
Cochrane Library (Wiley) for **CENTRAL** and **CDSR** Date of original search: 23/05/2023 Date of update search: 25/09/2023	#1 MeSH descriptor: [Sleep Apnea Syndromes] this term only #2 MeSH descriptor: [Sleep Apnea, Obstructive] this term only #3 (sleep* near/4 hypopn?ea*):ti,ab,kw #4 ("obstructive sleep*" next apn?ea*):ti,ab,kw #5 (sleep* near/4 disorder* near/4 breath*):ti,ab,kw #6 (OSA or SDB or OSAS or OSAHS):ti,ab,kw #7 4-#6 #8 MeSH descriptor: [Actigraphy] this term only #9 (actigraph* or "sleep monitor*" or accelerometer):ti,ab,kw #10 MeSH descriptor: [Oximetry] explode all trees #11 (oxymet* or oximet*):ti,ab,kw #12 ("oxygen desaturation"):ti,ab,kw #13 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*):ti,ab,kw #14 MeSH descriptor: [Capnography] this term only #15 (capnogra* or ((CO2 or "carbon dioxide") near/1 monitor*)):ti,ab,kw #16 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT):ti,ab,kw #17 MeSH descriptor: [Mobile Applications] this term only #18 ("limited channel*" or limited-channel* or multichannel or multi-channel):ti,ab,kw #19 125-#18	Original search: 539 (0 reviews; 539 trials) Update search: 18 (0 reviews; 18 trials)

	#20 MeSH descriptor: [Monitoring, Ambulatory] this term only #21 (home or at-home or home-based or unattended or portable or ambulatory):ti,ab,kw #22 #20 or #21 #23 #19 and #22 #24 MeSH descriptor: [Monitoring, Ambulatory] this term only #25 (((home or at-home or home-based or unattended or portable or ambulatory) near/3 (test* or device* or detect* or identif* or diagnos* or screen*)) or HSAT):ti,ab,kw #26 MeSH descriptor: [Wearable Electronic Devices] this term only #27 (((wearable* or nearable*) near/3 (test* or device* or detect* or identif* or diagnos* or screen*)) or WADD):ti,ab,kw #28 (((home or at-home or home-based or unattended) near/3 (polygraph* or polysomnograph*)) or HRP):ti,ab,kw #29 (Acupebble or Acurable):ti,ab,kw #30 (Brizzy or JAWAC or Nomics):ti,ab,kw #31 (NightOwl or Ectosense or ResMed):ti,ab,kw #32 (Sunrise):ti,ab,kw #33 (WatchPAT or Itamar or Zoll):ti,ab,kw #34 #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 #35 #7 and #34	
Web of Science for **Science Citation Index Expanded** (SCI-EXPANDED) and **Conference Proceedings Citation Index – Science** (CPCI-S) Date of original search: 23/05/2023 Date of update search: 25/09/2023	1 TS=((sleep* NEAR/4 (apn$ea* OR hypopn$ea*)) OR (sleep* NEAR/4 disorder* NEAR/4 breath*) OR OSA OR SDB OR OSAS OR OSAHS) 2 TS=(actimet* OR actigraph* OR "sleep monitor*" OR accelerometer OR oximet* OR oxymet* OR "oxygen monitor*" OR oxi-capnogra* OR oxicapnogra* OR oxy-capnogra* OR oxycapnogra* OR capnogra* OR ((CO2 OR "carbon dioxide") NEAR/1 monitor*) OR polysomnogra*) 3 TS=("peripheral arterial ton*" OR PAT) 4 TS=("limited channel*" OR limited-channel* OR multichannel OR multi-channel) 5 TS=((home* OR at-home OR home-based OR unattended OR portable OR ambulatory) NEAR/3 (test* OR device* OR detect* OR identif* OR diagnos* OR screen*)) 6 #5 AND (#2 OR #3 OR #4) 7 TS=(((home OR at-home OR home-based OR portable OR unattended OR ambulatory) NEAR/3 (test* OR device* OR detect* OR identif* OR diagnos* OR screen*)) OR HSAT) 8 TS=(((wearable* OR nearable*) NEAR/3 (test* OR device* OR detect* OR identif* OR diagnos* OR screen*)) OR WADD) 9 TS=(Acupebble OR Acurable) 10 TS=(Brizzy OR JAWAC OR Nomics) 11 TS=(NightOwl OR Ectosense OR ResMed)	Original search: 1334 Update search: 109

	12 TS=(Sunrise) 13 TS=(WatchPAT OR Itamar OR Zoll) 14 #13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 15 #14 AND #1 16 TI=(CPAP or "continuous positive airway pressure") 17 #15 NOT #16 18 #15 NOT #16 and Editorial Material or Letter or Note (Exclude – Document Types) 19 #15 NOT #16 and Editorial Material or Letter or Note (Exclude – Document Types) and Proceeding Paper or Meeting Abstract (Document Types) 20 #15 NOT #16 and Editorial Material or Letter or Note (Exclude – Document Types) and Proceeding Paper or Meeting Abstract (Document Types) and 2022 or 2021 or 2020 (Exclude – Publication Years) 21 #18 NOT #20 22 #18 NOT #20 and English (Languages)	
International HTA Database (database.inahta.org) Date of original search: 22/05/2023 Date of update search: 25/09/2023	1 (Sleep Apnea, Obstructive)[mh] OR (Sleep Apnea Syndromes)[mh] OR ((sleep* AND (apnea* OR hypopnea*)) OR (sleep* AND (apnoea* OR hypopnoea*)) OR (OSAHS OR OSA OR OSAS))) 2 (Monitoring, ambulatory)[mh] OR ((home OR at-home OR home-based OR unattended OR portable OR ambulatory) AND (test* OR device* OR detect* OR identif* OR diagnos* OR screen* OR polygraph* OR oximet* OR oxymet* OR capnograph* OR oxicapnograph* OR oxycapnograph* OR actigraph* OR HSAT OR HRP OR “peripheral arterial ton*” OR PAT) OR (Acupebble OR Acurable OR NightOwl OR Ectosense OR ResMed OR Sunrise OR WatchPAT OR Itamar OR Zoll)) 3 1 AND 2 English language limit	Original search: 16 Update search: 0
CRD Database for **Database of Abstracts of Reviews of Effects** (DARE) and **NHS Economic Evaluations Database** (NHS EED) Date of original search: 23/05/2023 Date of update search: not applicable	1 MeSH DESCRIPTOR Sleep Apnea Syndromes IN DARE, NHSEED 2 MeSH DESCRIPTOR Sleep Apnea, Obstructive IN DARE, NHSEED 3 (sleep* near4 (apnea* or hypopnea*)) IN DARE, NHSEED 4 (sleep* near4 (apnoea* or hypopnoea*)) IN DARE, NHSEED 5 (sleep* near4 disorder* near4 breath*) IN DARE, NHSEED 6 (OSA or SDB or OSAS or OSAHS) IN DARE, NHSEED 7 (#1 OR #2 OR #3 OR #4 OR #5 OR #6) 8 MeSH DESCRIPTOR Monitoring, Physiologic IN DARE, NHSEED 9 MeSH DESCRIPTOR Actigraphy IN DARE, NHSEED 10 (actigraph* or "sleep monitor*" or accelerometer) IN DARE, NHSEED 11 MeSH DESCRIPTOR Oximetry IN DARE, NHSEED	Original search: 22 Update search: not applicable

	12 (oxymet* or oximet*) IN DARE, NHSEED 13 ("oxygen desaturation") IN DARE, NHSEED 14 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*) IN DARE, NHSEED 15 MeSH DESCRIPTOR Capnography IN DARE, NHSEED 16 (capnogra* or ((CO2 or "carbon dioxide") near1 monitor)) IN DARE, NHSEED 17 (#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16) 18 MeSH DESCRIPTOR Monitoring, Ambulatory IN DARE, NHSEED 19 (home or at-home or home-based or unattended or portable) IN DARE, NHSEED 20 (#18 OR #19) 21 (#17 AND #20) 22 MeSH DESCRIPTOR Monitoring, Ambulatory IN DARE, NHSEED 23 (((home or at-home or home-based) near3 (test* or device* or detect* or identif* or diagnos* or screen*)) or HSAT) IN DARE, NHSEED 24 MeSH DESCRIPTOR Wearable Electronic Devices IN DARE, NHSEED 25 MeSH DESCRIPTOR Mobile Applications IN DARE, NHSEED 26 (((wearable* or nearable* or portable or bed-mounted or ambulatory or unattended) near3 (device* or technolog* or test* or detect* or diagnos* or identif* or sensor* or biosensor* or tracker* or tracking)) or WADD) IN DARE, NHSEED 27 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT) IN DARE, NHSEED 28 ("limited channel*" or limited-channel* or multichannel or multi-channel) IN DARE, NHSEED 29 (((home or at-home or home-based or unattended) near3 (polygraph* or polysomnograph*)) or HRP) IN DARE, NHSEED 30 (Acupebble or Acurable) IN DARE, NHSEED 31 (Brizzy or JAWAC or Nomics) IN DARE, NHSEED 32 (NightOwl or Ectosense or ResMed) IN DARE, NHSEED 33 (Sunrise) IN DARE, NHSEED 34 (WatchPAT or Itamar or Zoll) IN DARE, NHSEED 35 (#21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 or #34) 36 #7 and #35	
Epistemonikos (epistemonikos.org) Date of original search: 22/05/2023 Date of update search: 25/09/2023	title:(((sleep* AND (apnea* OR hypopnea*)) OR (sleep* AND (apnoea* OR hypopnoea*)) OR OSAHS OR OSA OR OSAS) AND title:(((home OR at-home OR home-based OR unattended OR portable OR ambulatory) AND (test* OR device* OR detect* OR identif* OR diagnos* OR screen* OR polygraph* OR oximet* OR oxymet* OR capnograph* OR oxicapnograph* OR oxycapnograph* OR actigraph* OR	Original search: 157 Update search: 20

	HSAT OR HRP OR "peripheral arterial ton*" OR PAT) OR (Acupebble OR Acurable OR Brizzy OR JAWAC OR Nomics OR NightOwl OR Ectosense OR ResMed OR Sunrise OR WatchPAT OR Itamar OR Zoll))))	
ClinicalTrials.gov (beta.clinicaltrials.gov) Date of original search: 24/05/2023	(Sleep Apnea, Obstructive \(Diagnosis\) OR Sleep Hypopnea \(Diagnosis\)) AND ((home OR at-home OR home-based OR unattended OR portable OR ambulatory) AND (test* OR detect* OR identif* OR diagnos* OR screen*))	Original search: 451
BePartOfResearch (formerly the UK Clinical Trials Gateway) Date of original search: 24/05/2023	I'm looking for research about: Obstructive sleep apnoea 9 studies Obstructive sleep apnea 4 studies Sleep hypopnoea 1 study Sleep hypopnea 0 matches	Original search: 14
NIHR Clinical Research Network Public Search (public-odp.nihr.ac.uk/) Date of original search: 24/05/2023	All subspecialities: Sleep medicine	Original search: 19
OpenGrey via DANS EASY Archive (Data Archiving and Networked Services) Date of original search: 24/05/2023	sleep AND (apnea* OR apnoea* OR hypopnea* OR hypopnoea* OR "disordered breathing") AND (diagnos* OR detect*)	Original search: 0
PROSPERO register of systematic reviews Date of original search: 24/05/2023	#1 MeSH DESCRIPTOR sleep apnea syndromes #2 MeSH DESCRIPTOR sleep apnea, obstructive #3 sleep* adj4 (hypopnea* OR hypopnoea*) #4 "obstructive sleep*" adj (apnea* OR apnoea*) #5 sleep* adj4 disorder* adj4 breath* #6 OSA or SDB or OSAS or OSAHS #7 #1 OR #2 OR #3 OR #4 OR #5 OR #6 #8 MeSH DESCRIPTOR Actigraphy #9 actigraph* or "sleep monitor*" or accelerometer #10 MeSH DESCRIPTOR Oximetry EXPLODE ALL TREES #11 oxymet* or oximet* #12 "oxygen desaturation" #13 oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra* #14 MeSH DESCRIPTOR Capnography	Original search: 96

	#15 capnogra* or ((CO2 or "carbon dioxide") adj1 monitor*) #16 "peripheral arterial tone" or "peripheral arterial tonometry" or PAT #17 MeSH DESCRIPTOR Mobile Applications #18 "limited channel*" or limited-channel* or multichannel or multi-channel #19 MeSH DESCRIPTOR Wearable Electronic Devices #20 ((wearable* or nearable*) adj3 (test* or device* or detect* or identif* or diagnos* or screen*)) or WADD #21 #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 #22 MeSH DESCRIPTOR Monitoring, Ambulatory #23 test* or device* or detect* or identif* or diagnos* or screen* #24 #22 AND #23 #25 home or at-home or home-based or unattended or portable or ambulatoryhome or at-home or home-based or unattended or portable or ambulatory #26 #24 OR #25 #27 #21 AND #26 #28 ((home or at-home or home-based or unattended or portable or ambulatory) adj3 (test* or device* or detect* or identif* or diagnos* or screen*)) or HSAT #29 Acupebble or Acurable #30 Brizzy or JAWAC or Nomics #31 NightOwl or Ectosense or ResMed #32 Sunrise #33 WatchPAT or Itamar or Zoll #34 #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 #35 #7 AND #34 #36 (CPAP or "continuous positive airway pressure"):TI #37 #35 NOT #36	

1b Searches for cost-effectiveness studies

The database search strategies for the cost effectiveness searches aimed to identify economic evaluations and cost-related studies associated with OSAHS. The overall strategy includes slightly broader population and intervention terms than for the clinical effectiveness searches, and economic evaluation and cost terms. We used the CADTH economic evaluations filters for the MEDLINE and Embase database searches.[126 127] We applied search limits for English language and for publication in the last ten years. Table 53 below details the search strategies for the databases. See also section 5.1.1 of this report.

Table 53 Search strategies for cost-effectiveness

Database, Host, Years searched, Date searched	Literature search strategy	Results
Ovid MEDLINE(R) ALL 1946 to May 23, 2023 Date of original search: 24/05/2023 Date of update search: 25/09/2023	1 sleep apnea syndromes/ or sleep apnea, obstructive/ 2 (sleep* adj4 (apn?ea* or hypopn?ea*)).ti,ab,kf. 3 (sleep* adj4 disorder* adj4 breath*).ti,ab,kf. 4 (OSA or SDB or OSAS or OSAHS).ti,ab,kf. 5 1 or 2 or 3 or 4 6 Monitoring, physiologic/ 7 Actigraphy/ 8 (actigraph* or "sleep monitor*" or accelerometer).ti,ab,kf. 9 exp Oximetry/ 10 (oxymet* or oximet*).ti,ab,kf. 11 "oxygen desaturation".ti,ab,kf. 12 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*).ti,ab,kf. 13 Capnography/ 14 (capnogra* or ((CO2 or "carbon dioxide") adj1 monitor*)).ti,ab,kf. 15 Monitoring, Ambulatory/ 16 (home or at-home or home-based or unattended or portable).ti,ab,kf. 17 (((home or at-home or home-based) adj3 (test* or device* or monitor* or detect* or identif* or diagnos* or screen*)) or HSAT).ti,ab,kf. 18 Wearable Electronic Devices/ 19 Mobile Applications/ 20 (((wearable* or nearable* or portable or bed-mounted or ambulatory or unattended) adj3 (device* or technolog* or monitor* or test* or detect* or diagnos* or identif* or sensor* or biosensor* or tracker* or tracking)) or WADD).ti,ab,kf. 21 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT).ti,ab,kf. 22 ("limited channel*" or limited-channel* or multichannel or multi-channel).ti,ab,kf. 23 (((home or at-home or home-based or unattended) adj3 (polygraph* or polysomnograph*)) or HRP).ti,ab,kf. 24 (Acupebble or Acurable).ti,ab,kf. 25 (NightOwl or Ectosense or ResMed).ti,ab,kf. 26 Sunrise.ti,ab,kf. 27 (WatchPAT or Itamar or Zoll).ti,ab,kf. 28 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 29 5 and 28 30 letter/ 31 editorial/ 32 news/	Original search: 192 Update search: 7

	33 exp historical article/ 34 Anecdotes as Topic/ 35 comment/ 36 case reports/ 37 (letter or comment*).ti. 38 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 39 randomized controlled trial/ or random*.ti,ab. 40 38 not 39 41 animals/ not humans/ 42 exp Animals, Laboratory/ 43 exp Animal Experimentation/ 44 exp Models, Animal/ 45 exp Rodentia/ 46 (rat or rats or mouse or mice).ti. 47 40 or 41 or 42 or 43 or 44 or 45 or 46 48 29 not 47 49 limit 48 to english language 50 Economics/ 51 exp "Costs and Cost Analysis"/ 52 Economics, Nursing/ 53 Economics, Medical/ 54 Economics, Pharmaceutical/ 55 exp Economics, Hospital/ 56 Economics, Dental/ 57 exp "Fees and Charges"/ 58 exp Budgets/ 59 budget*.ti,ab,kf. 60 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ti,kf. 61 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ab. /freq=2 62 (cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kf. 63 (value adj2 (money or monetary)).ti,ab,kf. 64 exp models, economic/ 65 economic model*.ab,kf. 66 markov chains/ 67 markov.ti,ab,kf. 68 monte carlo method/ 69 monte carlo.ti,ab,kf. 70 exp Decision Theory/ 71 (decision* adj2 (tree* or analy* or model*)).ti,ab,kf. 72 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 73 49 and 72	

	74 limit 73 to yr="2013 -Current" 75 remove duplicates from 74	
Ovid Embase Classic+Embase 1947 to 2023 May 23 Date of original search: 24/05/2023 Date of update search: 25/09/2023	1 sleep disordered breathing/ 2 (sleep* adj4 (apn?ea* or hypopn?ea*)).ti,ab,kf. 3 (sleep* adj4 disorder* adj4 breath*).ti,ab,kf. 4 (OSA or SDB or OSAS or OSAHS).ti,ab,kf. 5 1 or 2 or 3 or 4 6 actimetry/ 7 (actigraph* or "sleep monitor*" or accelerometer).ti,ab,kf. 8 oximetry/ or transcutaneous oxygen monitoring/ 9 (oxymet* or oximet*).ti,ab,kf. 10 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*).ti,ab,kf. 11 capnometry/ 12 (capnogra* or ((CO2 or "carbon dioxide") adj1 monitor*)).ti,ab,kf. 13 apnea monitor/ 14 polysomnograph/ 15 exp pulse oximeter/ 16 (home* or at-home or home-based or unattended or portable).ti,ab,kf. 17 portable equipment/ 18 portable patient monitor/ 19 (6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15) and (16 or 17 or 18) 20 (((home or at-home or home-based) adj3 (test* or device* or monitor* or detect* or identif* or diagnos* or screen*)) or HSAT).ti,ab,kf. 21 portable patient monitor/ 22 peripheral arterial tonometry device/ 23 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT).ti,ab,kf. 24 wearable sensor/ 25 (((wearable* or nearable* or portable or bed-mounted or ambulatory or unattended) adj3 (device* or technolog* or monitor* or test* or detect* or diagnos* or identif* or sensor* or biosensor* or tracker* or tracking)) or WADD).ti,ab,kf. 26 mobile application/ or mobile health application/ 27 ("limited channel*" or limited-channel* or multichannel or multi-channel).ti,ab,kf. 28 (((home or at-home or home-based or unattended) adj3 (polygraph* or polysomnograph*)) or HRP).ti,ab,kf. 29 home monitoring/ 30 home sleep apnea testing/ 31 "home use apnea monitor"/ 32 (Acupebble or Acurable).ti,ab,kf. 3 33 (NightOwl or Ectosense or ResMed).ti,ab,kf. 34 Sunrise.ti,ab,kf. 35 (WatchPAT or Itamar or Zoll).ti,ab,kf.	Original search: 238 Update search: 18

	36 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 37 5 and 36 38 letter.pt. or letter/ 39 note.pt. 40 editorial.pt. 41 case report/ or case study/ 42 (letter or comment*).ti. 43 38 or 39 or 40 or 41 or 42 44 randomized controlled trial/ or random*.ti,ab. 45 43 not 44 46 animal/ not human/ 47 nonhuman/ 48 exp Animal Experiment/ 49 exp Experimental Animal/ 50 animal model/ 51 exp Rodent/ 52 (rat or rats or mouse or mice).ti. 53 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 54 37 not 53 55 (conference abstract* or conference review or conference paper or conference proceeding).db,pt,su. 56 limit 55 to yr="2020 -Current" 57 55 not 56 58 54 not 57 59 Economics/ 60 Cost/ 61 exp Health Economics/ 62 Budget/ 63 budget*.ti,ab,kf. 64 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ti,kf. 65 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ab. /freq=2 66 (cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kf. 67 (value adj2 (money or monetary)).ti,ab,kf. 68 Statistical Model/ 69 economic model*.ab,kf. 70 Probability/ 71 markov.ti,ab,kf. 72 monte carlo method/ 73 monte carlo.ti,ab,kf. 74 Decision Theory/ 75 Decision Tree/ 76 (decision* adj2 (tree* or analy* or model*)).ti,ab,kf.	

	77 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 78 58 and 77 79 limit 78 to english language 80 limit 79 to yr="2013 -Current" 81 remove duplicates from 80	
Cochrane Library (Wiley) for **CENTRAL** and **CDSR** Date of original search: 24/05/2023 Date of update search: 25/09/2023	#1 MeSH descriptor: [Sleep Apnea Syndromes] this term only #2 MeSH descriptor: [Sleep Apnea, Obstructive] this term only #3 (sleep* near/4 (apn?ea* or hypopn?ea*)):ti,ab,kw #4 (sleep* near/4 disorder* near/4 breath*):ti,ab,kw #5 (OSA or SDB or OSAS or OSAHS):ti,ab,kw #6 #1 or #2 or #3 or #4 or #5 #7 MeSH descriptor: [Monitoring, Physiologic] this term only #8 MeSH descriptor: [Actigraphy] this term only #9 (actigraph* or (sleep next monitor*) or accelerometer):ti,ab,kw #10 MeSH descriptor: [Oximetry] explode all trees #11 (oxymet* or oximet*):ti,ab,kw #12 ("oxygen desaturation"):ti,ab,kw #13 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*):ti,ab,kw #14 MeSH descriptor: [Capnography] this term only #15 (capnogra* or ((CO2 or "carbon dioxide") near/1 monitor*)):ti,ab,kw #16 MeSH descriptor: [Monitoring, Ambulatory] this term only #17 (home or at-home or home-based or unattended or portable):ti,ab,kw #18 (((home or at-home or home-based) near/3 (test* or device* or monitor* or detect* or identif* or diagnos* or screen*)) or HSAT):ti,ab,kw #19 MeSH descriptor: [Wearable Electronic Devices] this term only #20 MeSH descriptor: [Mobile Applications] this term only #21 (((wearable* or nearable* or portable or bed-mounted or ambulatory or unattended) near/3 (device* or technolog* or monitor* or test* or detect* or diagnos* or identif* or sensor* or biosensor* or tracker* or tracking)) or WADD):ti,ab,kw #22 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT):ti,ab,kw #23 ((limited next channel*) or limited-channel* or multichannel or multi-channel):ti,ab,kw #24 (((home or at-home or home-based or unattended) near/3 (polygraph* or polysomnograph*)) or HRP):ti,ab,kw #25 (Acupebble or Acurable):ti,ab,kw #26 (NightOwl or Ectosense or ResMed):ti,ab,kw #27 (Sunrise):ti,ab,kw	Original search: 185 (1 review; 184 trials) Update search: 6 (0 reviews; 6 trials)

	#28 (WatchPAT or Itamar or Zoll):ti,ab,kw #29 128-#28 #30 #6 and #29 #31 MeSH descriptor: [Economics] this term only #32 MeSH descriptor: [Costs and Cost Analysis] explode all trees #33 MeSH descriptor: [Economics, Nursing] this term only #34 MeSH descriptor: [Economics, Medical] this term only #35 MeSH descriptor: [Economics, Pharmaceutical] this term only #36 MeSH descriptor: [Economics, Hospital] explode all trees #37 MeSH descriptor: [Economics, Dental] this term only #38 MeSH descriptor: [Fees and Charges] explode all trees #39 MeSH descriptor: [Budgets] explode all trees #40 (budget*):ti,ab,kw #41 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed):ti,ab,kw #42 (cost* near/2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)):ti,ab,kw #43 (value near/2 (money or monetary)):ti,ab,kw #44 MeSH descriptor: [Models, Economic] explode all trees #45 (economic next model*):ti,ab,kw #46 MeSH descriptor: [Markov Chains] this term only #47 (markov):ti,ab,kw #48 MeSH descriptor: [Monte Carlo Method] this term only #49 ("monte carlo"):ti,ab,kw #50 MeSH descriptor: [Decision Theory] explode all trees #51 (decision* near/2 (tree* or analy* or model*)):ti,ab,kw #52 84-#51 #53 #30 and #52 with Cochrane Library publication date Between Jan 2013 and May 2023	
Web of Science for **Science Citation Index Expanded** (SCI-EXPANDED) and **Conference Proceedings Citation Index – Science** (CPCI-S)	1 TS=((sleep* NEAR/4 (apn$ea* OR hypopn$ea*)) OR (sleep* NEAR/4 disorder* NEAR/4 breath*) OR OSA OR SDB OR OSAS OR OSAHS) 2 TS=(actimet* OR actigraph* OR "sleep monitor*" OR accelerometer OR oximet* OR oxymet* OR "oxygen monitor*" OR oxi-capnogra* OR oxicapnogra* OR oxy-capnogra* OR oxycapnogra* OR capnogra* OR ((CO2 OR "carbon dioxide") NEAR/1 monitor*)) 3 TS=("peripheral arterial ton*" OR PAT)	Original search: 324 Update search: 30

Date of original search: 24/05/2023 Date of update search: 25/09/2023	4 TS=(((wearable* OR nearable*) NEAR/3 (test* OR device* OR detect* OR identif* OR diagnos* OR screen*)) OR WADD) 5 TS=(((home OR at-home OR home-based OR unattended) NEAR/3 (polygraph* OR polysomnograph*)) OR HRP) 6 TS=((home* OR at-home OR home-based OR unattended OR portable OR ambulatory) NEAR/3 (test* OR device* OR detect* OR identif* OR diagnos* OR screen*)) 7 TS=("limited channel*" OR limited-channel* OR multichannel OR multi-channel) 8 TS=(Acupebble OR Acurable) 9 TS=(Brizzy OR JAWAC OR Nomics) 10 TS=(NightOwl OR Ectosense OR ResMed) 11 TS=(Sunrise) 12 TS=(WatchPAT OR Itamar OR Zoll) 13 #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 14 #13 AND #1 15 TS=(economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed) 16 TS=(budget*) 17 TS=(cost* near/2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)) 18 TS=(value near/2 (money or monetary)) 19 TS=(markov or “monte carlo”) 20 TS=(decision near/2 (tree* or analy* or model*)) 21 #20 OR #19 OR #18 OR #17 OR #16 OR #15 22 #21 AND #14 23 #21 AND #14 and Editorial Material (Exclude – Document Types) 24 #21 AND #14 and Editorial Material (Exclude – Document Types) and Meeting Abstract or Proceeding Paper (Document Types) 25 #21 AND #14 and Editorial Material (Exclude – Document Types) and Meeting Abstract or Proceeding Paper (Document Types) and 2022 or 2021 or 2020 (Exclude – Publication Years) 26 #23 NOT #25 27 #23 NOT #25 and 2023 or 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 (Publication Years) 28 #23 NOT #25 and 2023 or 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 (Publication Years) and English (Languages)	
CRD Database for **DARE** and **NHS EED**	1 MeSH DESCRIPTOR Sleep Apnea Syndromes IN DARE, NHSEED 2 MeSH DESCRIPTOR Sleep Apnea, Obstructive IN DARE, NHSEED	Original search: 23

Date of original search: 24/05/2023 Date of update search: Not applicable	3 (sleep* near4 (apnea* or hypopnea*)) IN DARE, NHSEED 4 (sleep* near4 (apnoea* or hypopnoea*)) IN DARE, NHSEED 5 (sleep* near4 disorder* near4 breath*) IN DARE, NHSEED 6 (OSA or SDB or OSAS or OSAHS) IN DARE, NHSEED 7 (#1 OR #2 OR #3 OR #4 OR #5 OR #6) 231 8 MeSH DESCRIPTOR Monitoring, Physiologic IN DARE, NHSEED 9 MeSH DESCRIPTOR Actigraphy IN DARE, NHSEED 10 (actigraph* or "sleep monitor*" or accelerometer) IN DARE, NHSEED 11 MeSH DESCRIPTOR Oximetry IN DARE, NHSEED 12 (oxymet* or oximet*) IN DARE, NHSEED 13 ("oxygen desaturation") IN DARE, NHSEED 14 (oxi-capnogra* or oxicapnogra* or oxy-capnogra* or oxycapnogra*) IN DARE, NHSEED 15 MeSH DESCRIPTOR Capnography IN DARE, NHSEED 16 (capnogra* or ((CO2 or "carbon dioxide") near1 monitor)) IN DARE, NHSEED 17 (#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16) 18 MeSH DESCRIPTOR Monitoring, Ambulatory IN DARE, NHSEED 19 (home or at-home or home-based or unattended or portable) IN DARE, NHSEED 20 (#18 OR #19) 21 (#17 AND #20) 22 MeSH DESCRIPTOR Monitoring, Ambulatory IN DARE, NHSEED 23 (((home or at-home or home-based) near3 (test* or device* or monitor* or detect* or identif* or diagnos* or screen*)) or HSAT) IN DARE, NHSEED 24 MeSH DESCRIPTOR Wearable Electronic Devices IN DARE, NHSEED 25 MeSH DESCRIPTOR Mobile Applications IN DARE, NHSEED 26 (((wearable* or nearable* or portable or bed-mounted or ambulatory or unattended) near3 (device* or technolog* or monitor* or test* or detect* or diagnos* or identif* or sensor* or biosensor* or tracker* or tracking)) or WADD) IN DARE, NHSEED 27 ("peripheral arterial tone" or "peripheral arterial tonometry" or PAT) IN DARE, NHSEED 28 ("limited channel*" or limited-channel* or multichannel or multi-channel) IN DARE, NHSEED	Update search: not applicable

	29 (((home or at-home or home-based or unattended) near3 (polygraph* or polysomnograph*)) or HRP) IN DARE, NHSEED 30 (Acupebble or Acurable) IN DARE, NHSEED 31 (Brizzy or JAWAC or Nomics) IN DARE, NHSEED 32 (NightOwl or Ectosense or ResMed) IN DARE, NHSEED 33 (Sunrise) IN DARE, NHSEED 34 (WatchPAT or Itamar or Zoll) IN DARE, NHSEED 35 (#21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 or #34) 36 #7 and #35	
International HTA Database (database.inahta.org) Date of original search: 24/05/2023 Date of update search: 25/09/2023	((Monitoring, ambulatory)[mh] OR ((home OR at-home OR home-based OR unattended OR portable OR ambulatory) AND (test* OR device* OR monitor* OR detect* OR identif* OR diagnos* OR screen* OR polygraph* OR oximet* OR oxymet* OR capnograph* OR oxicapnograph* OR oxycapnograph* OR actigraph* OR HSAT OR HRP OR "peripheral arterial ton*" OR PAT) OR (Acupebble OR Acurable OR NightOwl OR Ectosense OR ResMed OR Sunrise OR WatchPAT OR Itamar OR Zoll))) AND ((Sleep Apnea, Obstructive)[mh] OR (Sleep Apnea Syndromes)[mh] OR ((sleep* AND (apnea* OR hypopnea*)) OR (sleep* AND (apnoea* OR hypopnoea*)) OR (OSAHS OR OSA OR OSAS)))) Limited to English language	Original search: 17 Update search: 0
EconLit (EBSCO) Date of original search: 24/05/2023 Date of update search: 25/09/2023	S1 TI (sleep* N4 (apn#ea* or hypopn#ea*)) OR AB (sleep* N4 (apn#ea* or hypopn#ea*)) OR SU (sleep* N4 (apn#ea* or hypopn#ea*)) S2 TI sleep* N4 disorder* N4 breath* OR AB sleep* N4 disorder* N4 breath* OR SU sleep* N4 disorder* N4 breath* S3 TI ((OSA or SDB or OSAS or OSAHS) and sleep*) OR AB ((OSA or SDB or OSAS or OSAHS) and sleep*) OR SU ((OSA or SDB or OSAS or OSAHS) and sleep*)3 S4 S1 OR S2 OR S3 S5 S1 OR S2 OR S3 Narrow by Language: - english	Original search: 11 Update search: 0

1c Searches for health-related quality of life studies

The overall strategy to identify health-related quality of life studies relevant to OSAHS included terms for the condition and for health-related quality of life. We used the YHEC FSF1 sensitivity maximising filter to identify health state utility studies in the MEDLINE search.[129] We applied search limits for English language and for publication in the last ten years. Table 54 below details the search strategies for the databases. See also section 5.2.1 of this report.

Table 54 Search strategies for health-related quality of life

Database, Host, Years searched, Date searched	Literature search strategy	Results
Ovid MEDLINE(R) ALL 1946 to May 24, 2023 Date of original search: 25/05/2023 Date of update search: 27/09/2023	1 sleep apnea syndromes/ or sleep apnea, obstructive/ 2 (sleep* adj4 hypopn?ea*).ti,ab,kf. 3 ("obstructive sleep*" adj apn?ea*).ti,ab,kf. 4 (sleep* adj4 disorder* adj4 breath*).ti,ab,kf. 5 ((OSA or SDB or OSAS or OSAHS) and sleep).ti,ab,kf. 6 1 or 2 or 3 or 4 or 5 7 Quality-Adjusted Life Years/ 8 (quality adjusted or adjusted life year$).ti,ab,kf. 9 (qaly$ or qald$ or qale$ or qtime$).ti,ab,kf. 10 (illness state$1 or health state$1).ti,ab,kf. 11 (hui or hui1 or hui2 or hui3).ti,ab,kf. 12 (multiattribute$ or multi attribute$).ti,ab,kf. 13 (utility adj3 (score$1 or valu$ or health$ or cost$ or measur$ or disease$ or mean or gain or gains or index$)).ti,ab,kf. 14 utilities.ti,ab,kf. 15 (eq-5d or eq5d or eq-5 or eq5 or euro qual or euroqual or euro qual5d or euroqual5d or euro qol or euroqol or euro qol5d or euroqol5d or euro quol or euroquol or euro quol5d or euroquol5d or eur qolor eurqol or eur qol5d or eur qol5d or eur?qul or eur?qul5d or euro$ quality of life or european qol).ti,ab,kf. 16 (euro$ adj3 (5 d or 5d or 5 dimension$ or 5dimension$ or 5 domain$ or 5domain$)).ti,ab,kf. 17 (sf36$ or sf 36$ or sf thirtysix or sf thirty six).ti,ab,kf. 18 (time trade off$1 or time tradeoff$1 or tto or timetradeoff$1).ti,ab,kf. 19 quality of life/ and ((quality of life or qol) adj (score$1 or measure$1)).ti,ab,kf. 20 quality of life/ and ec.fs. 21 quality of life/ and (health adj3 status).ti,ab,kf. 22 (quality of life or qol).ti,ab,kf. and Cost-Benefit Analysis/ 23 ((qol or hrqol or quality of life).ti,kf. or *quality of life/) and ((qol or hrqol$ or quality of life) adj2 (increas$ or decrease$ or improv$ or declin$ or reduc$ or high$ or low$ or effect or effects or worse or score or scores or change$1 or impact$1 or impacted or deteriorat$)).ab. 24 Cost-Benefit Analysis/ and (cost-effectiveness ratio$ and (perspective$ or life expectanc$)).ti,ab,kf. 25 *quality of life/ and (quality of life or qol).ti. 26 quality of life/ and ((quality of life or qol) adj3 (improv$ or chang$)).ti,ab,kf. 27 quality of life/ and health-related quality of life.ti,ab,kf.	Original search: 619 Update search: 22

	28 models, economic/ 29 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 30 6 and 29 31 letter/ 32 editorial/ 33 news/ 34 exp historical article/ 35 Anecdotes as Topic/ 36 comment/ 37 case reports/ 38 (letter or comment*).ti. 39 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 40 randomized controlled trial/ or random*.ti,ab. 41 39 not 40 42 animals/ not humans/ 43 exp Animals, Laboratory/ 44 exp Animal Experimentation/ 45 exp Models, Animal/ 46 exp Rodentia/ 47 (rat or rats or mouse or mice).ti. 48 41 or 42 or 43 or 44 or 45 or 46 or 47 49 30 not 48 50 limit 49 to english language 51 limit 50 to yr="2013 -Current" 52 remove duplicates from 51	
Ovid Embase Classic+Embase 1947 to 2023 May 24 Date of original search: 25/05/2023 Date of update search: 27/09/2023	1 sleep disordered breathing/ 2 (sleep* adj4 hypopn?ea*).ti,ab,kf. 3 ("obstructive sleep*" adj apn?ea*).ti,ab,kf. 4 (sleep* adj4 disorder* adj4 breath*).ti,ab,kf. 5 (OSA or SDB or OSAS or OSAHS).ti,ab,kf. 6 1 or 2 or 3 or 4 or 5 7 quality adjusted life year/ 8 (quality adjusted or adjusted life year$).ti,ab,kf. 9 (qaly$ or qald$ or qale$ or qtime$).ti,ab,kf. 10 (illness state$1 or health state$1).ti,ab,kf. 11 (hui or hui1 or hui2 or hui3).ti,ab,kf. 12 (multiattribute$ or multi attribute$).ti,ab,kf. 13 (utility adj3 (score$1 or valu$ or health$ or cost$ or measur$ or disease$ or mean or gain or gains or index$)).ti,ab,kf. 14 utilities.ti,ab,kf. 15 (eq-5d or eq5d or eq-5 or eq5 or euro qual or euroqual or euro qual5d or euroqual5d or euro qol or euroqol or euro qol5d or euroqol5d or euro quol or euroquol or euro quol5d or euroquol5d or eur qol or eurqol or eur qol5d or eur qol5d or eur?qul or eur?qul5d or euro$ quality of life or european qol).ti,ab,kf. 16 (euro$ adj3 (5 d or 5d or 5 dimension$ or 5dimension$ or 5 domain$ or 5domain$)).ti,ab,kf.	Original search: 1151 Update search: 49

	17 (sf36$ or sf 36$ or sf thirtysix or sf thirty six).ti,ab,kf. 18 (time trade off$1 or time tradeoff$1 or tto or timetradeoff$1).ti,ab,kf. 19 "quality of life"/ and de.fs. 20 "quality of life"/ and (health adj3 status).ti,ab,kf. 21 (quality of life or qol).ti,ab,kf. and "cost benefit analysis"/ 22 ((qol or hrqol or quality of life).ti,kf. or *"quality of life"/) and ((qol or hrqol$ or quality of life) adj2 (increas$ or decrease$ or improv$ or declin$ or reduc$ or high$ or low$ or effect or effects or worse or score or scores or change$1 or impact$1 or impacted or deteriorat$)).ab. 23 "cost benefit analysis"/ and (cost-effectiveness ratio$ and (perspective$ or life expectanc$)).ti,ab,kf. 24 *"quality of life"/ and (quality of life or qol).ti. 25 "quality of life"/ and ((quality of life or qol) adj3 (improv$ or chang$)).ti,ab,kf. 26 "quality of life"/ and health-related quality of life.ti,ab,kf. 27 economic model/ 28 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 29 6 and 28 30 letter.pt. or letter/ 31 note.pt. 32 editorial.pt. 33 case report/ or case study/ 34 (letter or comment*).ti. 35 30 or 31 or 32 or 33 or 34 36 randomized controlled trial/ or random*.ti,ab. 37 35 not 36 38 animal/ not human/ 39 nonhuman/ 40 exp Animal Experiment/ 41 exp Experimental Animal/ 42 animal model/ 43 exp Rodent/ 44 (rat or rats or mouse or mice).ti. 45 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 46 29 not 45 47 (conference abstract* or conference review or conference paper or conference proceeding).db,pt,su. 48 limit 47 to yr="2020 -Current" 49 47 not 48 50 46 not 49 51 limit 50 to english language 52 limit 51 to yr="2013 -Current" 53 remove duplicates from 52	

Cochrane Library (Wiley) for **CENTRAL** and **CDSR** Date of original search: 25/05/2023 Date of update search: 27/09/2023	#1 MeSH descriptor: [Sleep Apnea Syndromes] this term only #2 MeSH descriptor: [Sleep Apnea, Obstructive] this term only #3 (sleep* near/4 hypopn?ea*):ti,ab,kw #4 ("obstructive sleep*" next apn?ea*):ti,ab,kw #5 (sleep* near/4 disorder* near/4 breath*):ti,ab,kw #6 (OSA or SDB or OSAS or OSAHS):ti,ab,kw #7 4-#6 #8 MeSH descriptor: [Quality-Adjusted Life Years] this term only #9 ("quality adjusted" or ("adjusted life" next year*)):ti,ab,kw #10 (qaly* or qald* or qale* or qtime*):ti,ab,kw #11 ((illness next state?) or (health next state?)):ti,ab,kw #12 (hui or hui1 or hui2 or hui3):ti,ab,kw #13 (multiattribute* or (multi next attribute*)):ti,ab,kw #14 (utility near/3 (score? or valu* or health* or cost* or measur* or disease* or mean or gain or gains or index*)):ti,ab,kw #15 (utilities):ti,ab,kw #16 (eq-5d or eq5d or eq-5 or eq5 or "euro qual" or euroqual or "euro qual5d" or euroqual5d or "euro qol" or euroqol or "euro qol5d" or euroqol5d or "euro quol" or euroquol or "euro quol5d" or euroquol5d or "eur qolor eurqol" or "eur qol5d" or "eur qol5d" or eur?qul or eur?qul5d or (euro* next "quality of life") or "european qol"):ti,ab,kw #17 (euro* near/3 ("5 d" or 5d or (5 next dimension*) or 5dimension* or (5 next domain*) or 5domain*)):ti,ab,kw #18 (sf36* or (sf next 36*) or "sf thirtysix" or "sf thirty six"):ti,ab,kw #19 (("time trade" next off?) or (time next tradeoff?) or tto or timetradeoff?):ti,ab,kw #20 MeSH descriptor: [Quality of Life] this term only #21 (("quality of life" or qol) near/1 (score? or measure?)):ti,ab,kw #22 (health near/3 status):ti,ab,kw #23 ("quality of life" or qol):ti #24 (("quality of life" or qol) near/3 (improv* or chang*)):ti,ab,kw #25 ("health-related quality of life"):ti,ab,kw #26 #21 or #22 or #23 or #24 or #25 #27 #20 and #26 #28 MeSH descriptor: [Cost-Benefit Analysis] this term only #29 ("quality of life" or qol):ti,ab,kw #30 ((cost-effectiveness next ratio*) and (perspective* or (life next expectanc*))):ti,ab,kw #31 #29 or #30 #32 #28 and #31	Original search: 447 Update search: 8

	#33 (qol or hrqol or "quality of life"):ti #34 (qol or hrqol or "quality of life"):kw #35 ((qol or hrqol* or "quality of life") near/2 (increas* or decrease* or improv* or declin* or reduc* or high* or low* or effect or effects or worse or score or scores or change? or impact? or impacted or deteriorat*)):ab #36 (#20 or #33 or #34) and #35 #37 MeSH descriptor: [Models, Economic] this term only #38 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #27 or #32 or #36 or #37 #39 #7 and #38 with Cochrane Library publication date Between Jan 2013 and May 2023	
ScHARR Health Utilities Database, Sheffield University (ScHARRHUD) (scharrhud.org/) Date of original search: 25/05/2023 Date of update search: 27/09/2023	Any field: “sleep apnea” or “sleep apnoea” or hypopnea or hypopnoea **1** result from 2010 (out of past ten years date range) **Utility indices in patients with the obstructive sleep apnea syndrome** Schmidlin,M., Fritsch,K., Matthews,F., Thurnheer,R., Senn,O., Bloch,K.E. *Respiration* 2010 79 200 - 208	Original search: 0 Update search: 0
Web of Science for **Science Citation Index Expanded** (SCI-EXPANDED) and **Conference Proceedings Citation Index – Science** (CPCI-S) Date of original search: 25/05/2023 Date of update search: 27/09/2023	1 TS=((sleep* NEAR/4 (apn$ea* OR hypopn$ea*)) OR (sleep* NEAR/4 disorder* NEAR/4 breath*) OR OSA OR SDB OR OSAS OR OSAHS) 2 TS=("quality adjusted" OR "adjusted life year*" OR QALY* OR QALD* OR QALE* OR QTIME*) 3 TS=("illness state$" OR "health state$") 4 TS=(hui or hui1 or hui2 or hui3) 5 TS=(multiattribute* or multi attribute*) 6 TS=((utilit* OR disutilit*) NEAR/3 (score$ or valu* or health* or cost* or measur* or disease* or mean or gain or gains or index*)) 7 TS=(eq-5d or eq5d or eq-5 or eq5 or "euro qual" or euroqual or "euro qual5d" or euroqual5d or "euro qol" or euroqol or "euro qol5d" or euroqol5d or "euro quol" or euroquol or "euro quol5d" or euroquol5d or "eur qol" or eurqol or "eur qol5d" or "eur qol5d" or eur$qul or eur$qul5d or "euro* quality of life" or "european qol") 8 TS=(euro* NEAR/3 ("5 d" or 5d or "5 dimension*" or 5dimension* or "5 domain*" or 5domain*)) 9 TS=(sf36* or "sf 36*" or "sf thirtysix" or "sf thirty six") 10 TS=("time trade off$" or "time tradeoff$" or tto or timetradeoff$) 11 TI=(qol or hrqol or "quality of life") 12 AB=((qol or hrqol* or "quality of life") NEAR/2 (increas* or decrease* or improv* or declin* or reduc* or	Original search:1145 Update search: 87

	high* or low* or effect or effects or worse or score or scores or change? or impact? or impacted or deteriorat*)) 13 #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 14 #13 AND #1 15 #13 AND #1 and Editorial Material or Letter or Correction (Exclude – Document Types) 16 #13 AND #1 and Editorial Material or Letter or Correction (Exclude – Document Types) and Meeting Abstract or Proceeding Paper (Document Types) 17 #13 AND #1 and Editorial Material or Letter or Correction (Exclude – Document Types) and Meeting Abstract or Proceeding Paper (Document Types) and 2022 or 2021 or 2020 (Exclude – Publication Years) 18 #15 NOT #17 19 #15 NOT #17 and 2023 or 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 (Publication Years) 20 #15 NOT #17 and 2023 or 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 (Publication Years) and English (Languages)	

Appendix 2 Further detail on inclusion/exclusion of studies in the systematic review of clinical effectiveness

2a. Inclusion/exclusion criteria

Table 55 Inclusion / exclusion screening worksheet

POPULATION Eligible: People aged ≥ 2 years[1] with suspected OSAHS (this can include patients with: • suspected OSA or HS • suspected OSA and HS • Mixed population of people with a suspected diagnosis (OSA/HS/ OSA & HS) and healthy people or people unsuspected of having OSA/HS/OSA & HS	Yes ↓ next question	Unclear ↓ next question	No → EXCLUDE
INTERVENTION Eligible: Any of the following devices: • AcuPebble SA100 (Acurable) • Brizzy (NomicsCare) • NightOwl (ResMed)[2] • Sunrise (Sunrise) • WatchPAT 300 (Zoll/Itamar) • WatchPAT ONE (Zoll/Itamar) • Unnamed novel devices for home testing that are less intrusive than comparator devices widely used in the NHS, are easier to put on and operate and potentially more comfortable to wear. • Unnamed devices described as "home respiratory polygraphy", "home sleep apnoea test", "type III test" or similar whose channels do **not include**: a nasal cannula and/or chest or abdominal respiratory inductive plethysmography (RIP) belts For children and young people (2-16 years), use of the interventions may be alongside CO2 monitoring. Ineligible: Home respiratory polysomnography, "type II test" or similar	Yes ↓ next question	Unclear ↓ next question	No → EXCLUDE
COMPARATOR Eligible: • **For people over 16:** Home respiratory polygraphy[3] or home oximetry (can include home test devices currently used in clinical practice but cannot include any of the named interventions). **For people over 16 with COPD**: Home oximetry alone is not recommended and will therefore not be considered a suitable comparator for this subgroup. • **For children and young people aged 2 to 16 years:** Home respiratory polygraphy[3] or home oximetry. CO2 monitoring maybe used alongside these technologies. • The **reference standard** can include:	Yes ↓ next question	Unclear ↓ next question	No → EXCLUDE

in-hospital polysomnography, polysomnography done outside hospital or respiratory polygraphy done in a healthcare setting (rather than at home).			
OUTCOMES Eligible: one or more from any of the following: • **Intermediate outcomes:** o Measures of performance to detect OSAHS and assess severity o Measures of concordance or agreement between intervention technologies, or between intervention technologies and comparators o Impact on clinical decision-making o Time to interpret device outputs and reach a diagnosis o Time to diagnosis or starting treatment o Number of repeat studies done (at home or in hospital) o Use of healthcare resources (such as number and length of hospital admissions, use of pharmacological and non-pharmacological interventions for management of OSAHS) o Test failure rate (including incidences where data recorded can't be analysed or a person doesn't sleep long enough to generate enough data for assessment) • **Clinical outcomes:** o Morbidity o Mortality • **Patient- and carer-reported outcomes** o Health-related quality of life o Ease of use and acceptability for patients and carers o Patient and carer experience	Yes ↓ next question	Unclear ↓ next question	No → EXCLUDE
DESIGN Eligible: Any study design[4]	Yes ↓ next question	Unclear ↓ next question	No → EXCLUDE
PUBLICATION STATUS Eligible: Journal articles; study reports; conference abstracts / proceedings if • published in the last three years (2019 onwards) • sufficient details are presented to allow appraisal of the methodology and the assessment of results to be undertaken.	Yes ↓ next question	Unclear ↓ next question	No → EXCLUDE
Final Decision	**INCLUDE**	**UNCLEAR (Discuss)**	**EXCLUDE (if full text state reason)**

[1] Sunrise is for people aged ≥ 3 years (AcuPebble is for adults only, Night Owl is for people aged ≥ 13 years WatchPAT is for ≥ 12 years). We have not received a dossier for Brizzy, therefore the lower age range for this device is unknown.
[2]Ectosense, which is part of ResMed, is also listed as the manufacturer on some material in the company submission.
[3] NG202 defines home respiratory polygraphy as including at least 4 channels. The final scope gives an example of oximetry, breathing rate, apnoeas and hypopnoeas, snoring and body position, while NICE NG 202 Review D gives an example of oximetry, pulse rate, air flow and chest or abdomen effort band).
[4] If both trial-based and observational evidence is available for any of the comparisons relevant to this review, priority will be given to analysis of the trial-based evidence.

2b List of studies excluded from the systematic review of clinical effectiveness

(Please note, this section will be completed for the final version of this report).

Full text publications not included in this systematic review were classified as either 'excluded' or 'unclear', as follows:

- Excluded: publications which did not meet the inclusion criteria, as listed in Table 56 below. Where more than one exclusion criterion is applicable to a publication we recorded only the first exclusion criterion.
- Unclear: publications whose eligibility for inclusion remain unclear after contacting the authors for further information. These are listed in Table 57 below.

Table 56 Full text publications excluded from systematic review of clinical effectiveness

Publication	Publication type	Exclusion reason
Abraham 2006[130]	Journal article	Intervention
Agency for Care, Effectiveness 2019[131]	Other	Design
Aldabayan 2017[132]	Poster	Intervention (WatchPAT 200)
Alonso Alvarez 2008[133]	Journal article	Intervention
Alonso Alvarez 2015[134]	Journal article	Intervention
Andreu 2012[135]	Journal article	Intervention
Aurora 2018[136]	Journal article	Intervention
Ayas 2003[137]	Journal article	Intervention (WatchPAT 100)
Badami 2021[138]	Conference abstract	Intervention
Badami 2022[139]	Conference abstract	Intervention
Bar 2003[140]	Journal article	Intervention (WatchPAT 100)
Berry 2008[141]	Journal article	Intervention (WatchPAT 100)
Bhattacharjee 2021[142]	Journal article	Intervention
Boulos 2020[143]	Conference abstract	Intervention
Boulos 2019[144]	Conference abstract	Intervention
Boulos 2018[145]	Conference abstract	Intervention
Boulos 2017[146]	Conference abstract	Intervention
Boulos 2022[147]	Journal article	Intervention
Braun 2021[148]	Conference abstract	Intervention
Braun 2021[149]	Journal article	Intervention
Bresler 2008[150]	Journal article	Population
Bubu 2022[151]	Conference abstract	Population
Carey 2023[152]	Journal article	Comparator
Cassiba 2023[153]	Powerpoint presentation	Population
Castillo-Escario 2019[154]	Journal article	Intervention

Cerina 2023[155]	Journal article	Intervention
Ceylan 2012[156]	Journal article	Intervention (WatchPAT 200)
Chai-Coetzer 2017[157]	Journal article	Intervention
Chakar 2022[158]	Journal article	Population
Chen 2021[159]	Journal article	Intervention
Chernyshev 2015[160]	Journal article	Intervention
Cheung 2021[161]	Journal article	Intervention
Chiner 2020[162]	Journal article	Intervention
Cho 2023[163]	Conference abstract	Intervention (WatchPAT 200)
Choi 2010[164]	Journal article	Intervention (WatchPAT 100)
Choi 2018[165]	Journal article	Intervention (WatchPAT 200)
Corral 2017[166]	Journal article	Intervention
Crupi 2015[167]	Journal article	Intervention
Damiani 2012[168]	Conference abstract	Design
De Chazal 2011[169]	Journal article	Intervention
De Chazal 2009[170]	Journal article	Intervention
De Oliveira 2009[171]	Journal article	Intervention
Ding 2023[172]	Journal article	Intervention
Domingo 2011[173]	Journal article	Intervention
Drks 2015[174]	Trial register record	Population
Duran-Cantolla 2017[175]	Conference abstract	Intervention
Duran-Cantolla 2018[176]	Conference abstract	Intervention
Escobar 2023[177]	Conference abstract	Population
Ferguson 2002[178]	Conference abstract	Intervention
Fischer 2003[179]	Journal article	Language
Fishman 2019[180]	Conference abstract	Intervention
Fishman 2018[181]	Journal article	Intervention
Folmer 2022[182]	Journal article	Intervention
Foo 2006[183]	Journal article	Population
Fynn 2020[184]	Conference abstract	Intervention
Gan 2017[185]	Journal article	Intervention (WatchPAT 200)
Ganglberger 2022[186]	Journal article	Intervention
Garg 2014[187]	Journal article	Intervention (WatchPAT 200)
Ghandeharioun 2015[188]	Letter	Intervention
Ghahjaverestan 2022[189]	Journal article	Intervention
Goldstein 2018[190]	Journal article	Intervention
Golpe 2022[191]	Journal article	Intervention
Green 2022[192]	Journal article	Intervention
Gros 2015[193]	Journal article	Intervention
Guerrero 2014[194]	Journal article	Intervention
Guerrero 2014[195]	Conference abstract	Intervention
Gupta 2021[196]	Journal article	Intervention
Gursoy 2021[197]	Journal article	Intervention

Gurubhagavatula 2013[198]	Journal article	Intervention
Hafezi 2020[199]	Journal article	Intervention
Hafezi 2019[200]	Journal article	Intervention
Hansen 2023[201]	Journal article	Intervention
Harris 2017[202]	Other	Design
Hayano 2020[203]	Journal article	Intervention
He 2020[204]	Conference abstract	Comparator
He 2016[205]	Journal article	Design
Hedner 2004[206]	Journal article	Intervention (WatchPAT 100)
Hilmisson 2020[207]	Journal article	Intervention
Holmedahl 2019[208]	Journal article	Intervention (WatchPAT 200)
Hung 2022[209]	Journal article	Intervention
Ikizoglu 2019[210]	Journal article	Intervention
Ioachimescu 2020[77]	Journal article	Intervention (WatchPAT 200)
Ioachimescu 2020[211]	Journal article	Intervention (WatchPAT 200)
Ioan 2023[212]	Journal article	Intervention
Isaka 2022[212]	Journal article	Intervention
ISRCT16982033 2023[213]	Trial register record	Comparator
Ito 2018[214]	Journal article	Intervention
Jen 2020[215]	Journal article	Intervention (WatchPAT 200)
Jiang 2015[216]	Journal article	Intervention
Jurado Gamez 2007[217]	Journal article	Intervention
Kalkbrenner 2019[218]	Journal article	Intervention
Kalkbrenner 2017[219]	Journal article	Intervention
Kapur 2018[220]	Journal article	Intervention
Kasai 2020[221]	Journal article	Intervention (WatchPAT 200)
Khan2009[222]	Conference abstract	Intervention (WatchPAT 100
Kim 2015[223]	Journal article	Intervention
Kinoshita 2018[224]	Journal article	Intervention (WatchPAT 200)
Kissow Lildal 2023[225]	Journal article	Intervention
Körkuyu 2015[226]	Journal article	Intervention (WatchPAT 200)
Kotzian 2018[227]	Journal article	Intervention
Kukwa 2022[228]	Journal article	Intervention
Kukwa 2021[229]	Journal article	Intervention (WatchPAT 200)
Kwon 2023[230]	Journal article	Intervention
Lachapelle 2022[231]	Journal article	Intervention
Le-Dong 2021[232]	Letter	Outcomes
Levendowski 2015[233]	Journal article	Intervention
Levendowski 2009[234]	Journal article	Population
Li 2021[235]	Journal article	Intervention
Manoni 2020[236]	Journal article	Intervention

Martinot 2023[237]	Journal article	Comparator
Martinot 2017[238]	Journal article	Comparator
Martinot 2023[239]	Journal article	Design
Martinot 2023[121]	Journal article	Outcomes
Martinot unknown date[240]	Draft manuscript	Outcomes
Martinot 2017[241]	Journal article	Outcomes
Martinot 2021[242]	Conference abstract	Population
Martinot 2022[243]	Conference abstract	Comparator
Martinot 2022[244]	Conference abstract	Comparator
Martinot 2021[245]	Journal article	Population
Martinot 2017[246]	Journal article	Population
Martinot 2021[247]	Conference abstract	Population
Martinot 2022[248]	Conference abstract	Population
Martinot 2017[249]	Conference abstract	Population
Martinot 2017[250]	Conference abstract	Population
Masa 2013[251]	Journal article	Intervention
Masa 2013[252]	Journal article	Intervention
Masa 2011[253]	Conference abstract	Intervention
Masa 2012[254]	Conference abstract	Intervention
Masa 2011[255]	Journal article	Intervention
Masa 2011[256]	Journal article	Intervention
Masa 2013[257]	Journal article	Intervention
Masa 2012[258]	Conference abstract	Intervention
Masa 2011[259]	Conference abstract	Intervention
Masa 2013[260]	Conference abstract	Intervention
Maury 2013[261]	Journal article	Intervention
Mello 2023[262]	Journal article	Intervention
Michelet 2020[263]	Journal article	Intervention
Milici 2018[264]	Journal article	Intervention
Miller 2018[265]	Journal article	Intervention
Mohammadieh 2021 [266]	Journal article	Intervention
Montazeri Ghahjaverestan 2021[267]	Journal article	Intervention
Montazeri Ghahjaverestan 2021[268]	Conference abstract	Intervention
Morillo 2010[269]	Journal article	Duplicate
Morillo 2010[270]	Journal article	Intervention
Mulgrew 2005[271]	Conference abstract	Intervention
Munoz-Ferrer 2020[272]	Journal article	Intervention
Nagubadi 2016[273]	Journal article	Intervention
Nakano 2004[274]	Journal article	Intervention
Nakase-Richardson 2020[275]	Journal article	Intervention
Nakase-Richardson 2020[276]	Journal article	Intervention
NCT05647746 2022[277]	Trial register record	Comparator
NCT01929447 2013[278]	Trial register record	Comparator
NCT00139022 2005[279]	Trial register record	Intervention
NCT00425659 2007[280]	Trial register record	Intervention
NCT00642486 2008[281]	Trial register record	Intervention
NCT00880165 2009[282]	Trial register record	Intervention

NCT01001858 2009[283]	Trial register record	Intervention
NCT01752556 2012[284]	Trial register record	Intervention
NCT02037438 2014[285]	Trial register record	Intervention
NCT02454023 2015[286]	Trial register record	Intervention
NCT02779894 2016[287]	Trial register record	Intervention
NCT03449550 2017[288]	Trial register record	Intervention
NCT04335994 2020[289]	Trial register record	Intervention
NCT05382754 2022[290]	Trial register record	Intervention
NCT05516524 2022[291]	Trial register record	Intervention
NCT05057975 2021[39]	Trial register record	Ongoing study (no results)
NCT04495062 2020[292]	Trial register record	Population
NCT01997723 2013[293]	Trial register record	Intervention (WatchPAT 200)
NCT02760680 2016[294]	Trial register record	Intervention (WatchPAT 200)
NCT05918120 2023[295]	Trial register record	Intervention
NHS Foundation Trust Cambridge unknown year[41]	Abstract	Ongoing study
O'Brien 2012[296]	Journal article	Intervention (WatchPAT 200)
Oliveira 2014[297]	Journal article	Duplicate
Oliveira 2015[298]	Journal article	Intervention
Onder 2012[299]	Journal article	Intervention (WatchPAT 200)
O'Reilly 2022[300]	Journal article	Intervention (WatchPAT 200)
Ounhasuttiyanon 2021[301]	Journal article	Intervention
Pallin 2014[302]	Journal article	Intervention
Parra 1997[303]	Journal article	Intervention
Pei 2023[304]	Journal article	Intervention
Penzel 2004[305]	Journal article	Population
Penzel 2004[306]	Conference abstract	Population
Pepin 2022[74]	Journal article	Intervention
Phua 2021[49]	Journal article	Intervention (WatchPAT 200)
Pillar 2003[307]	Journal article	Intervention (WatchPAT 100
Pinto 2015[308]	Journal article	Intervention (WatchPAT 200
Pittman 2004[309]	Journal article	Intervention (WatchPAT 100
Pittman 2000[310]	Journal article	Intervention (WatchPAT 100)
Planes 2010[311]	Journal article	Intervention
Polese 2013[312]	Journal article	Intervention
Prasad 2012[313]	Conference abstract	Intervention (WatchPAT 200)
Quintana-Gallego 2004[314]	Journal article	Intervention
Quiroga 2016[315]	Conference abstract	Intervention
Redline 1991[316]	Journal article	Intervention
Revana 2022[317]	Journal article	Intervention

Rocken 2023[318]	Journal article	Population
Rodriguez-Villegas 2014[319]	Journal article	Intervention
Rosen 2012[320]	Journal article	Intervention
Roth 2022[321]	Conference abstract	Intervention (WatchPAT 200)
SBUHB Evaluation of the Sunrise Device unknown date[322]	Video	Comparator
Scalzitti 2017[323]	Journal article	Intervention
Schafer 1997[324]	Journal article	Intervention
Schindhelm 2023[325]	Journal article	Intervention (WatchPAT 200)
Serra 2017[326]	Journal article	Comparator
Service Evaluations NHS Scotland unknown date[40]	Abstract	Ongoing study
Skomro 2006[327]	Conference abstract	Intervention
Stansbury 2022[328]	Journal article	Intervention
Stern 2022[329]	Conference abstract	Comparator
Su 2015[330]	Journal article	Language
Sun 20020[331]	Journal article	Intervention
Takama 2010[332]	Journal article	Intervention
Tanaka 2021[333]	Conference abstract	Population
Tanphaichitr 2018[334]	Journal article	Intervention (WatchPAT 200)
Tedeschi 2013[335]	Journal article	Intervention
Tellez 2022[336]	Conference abstract	Intervention
Ting 2018[337]	Journal article	Intervention (WatchPAT 200)
Tondo 2023[338]	Journal article	Intervention
Tondo 2022[339]	Journal article	Intervention
Tondo 2021[340]	Journal article	Intervention (WatchPAT 200)
Tung 2022[341]	Journal article	Intervention
van Gilst 2019[342]	Journal article	Intervention
van Veldhuisen 2022[343]	Journal article	Intervention
Veitz 2020[344]	Conference abstract	Population
Walter 2023[345]	Journal article	Comparator
Ward 2015[346]	Journal article	Intervention
Watkins 2009[347]	Journal article	Intervention
Weimin 2013[348]	Journal article	Intervention (WatchPAT 200)
Westenberg 2021[349]	Journal article	Intervention
Xian 2023[350]	Journal article	Intervention
Xu 2023[351]	Journal article	Intervention
Yadollahi 2010[352]	Journal article	Intervention
Yadollahi 2009[353]	Journal article	Intervention
Yamada 2020[354]	Journal article	Intervention
Yang 2021[355]	Journal article	Intervention
Yang 2022[356]	Journal article	Intervention
Yoon 2018[357]	Journal article	Intervention
Zandieh 2023[358]	Journal article	Intervention

Zhang 2020[359]	Journal article	Intervention (WatchPAT 200)
Zou 2006[360]	Journal article	Population

Table 57 Publications where eligibility for inclusion remains unclear after full-text screening and contacting authors

Study	Publication type	Reason unclear	Notes
Arguelles 2019[361]	Conference abstract	Unclear WatchPAT version	No author response
Campbell 2023[362]	Conference abstract	Unclear WatchPAT version and comparator	No author response
Campbell 2021[363]	Conference abstract	Unclear WatchPAT version and comparator	No author response
Fildes 2019[364]	Poster	Unclear WatchPAT version	No author response
Groshaeny 2022[365]	Conference abstract	Unclear population and comparator	No author response
Jen 2017[366]	Conference abstract	Unclear WatchPAT version	No author response
Jenad 2023[367]	Conference abstract	Unclear WatchPAT version	No author response
Kelly 2022[368]	Conference abstract	Unclear WatchPAT version	No author response
Martinot 2020[369]	Conference abstract	Unclear whether device algorithm was used for analysing data from device or applied to other patient data	
Maury 2013[370]	Journal article	Unclear intervention	
Maury 2014[261]	Journal article	Unclear intervention	
Morse 2020[371]	Conference abstract	Unclear intervention	No author response
NCT03188718 2017[372]	Trial register record	Unclear WatchPAT version	No author response
NCT04249011 2020[373]	Trial register record	Unclear population and WatchPAT version	
NCT01570738 2012[374]	Trial register record	Unclear population and WatchPAT version	
Orbea 2021[375]	Conference abstract	Unclear intervention	No author response
Pang 2007[376]	Journal article	Unclear WatchPAT version	No author response
Safadi 2014[377]	Journal article	Unclear intervention	No author response
Senny 2008[378]	Journal article	Unclear intervention	
Senny 2009[379]	Journal article	Unclear intervention	
Townsend 2007[380]	Journal article	Journal article unobtainable	Article not held by British Library. No response from Journal.
Waicus 2022[381]	Conference abstract	Unclear intervention	No author response

2c Details of ongoing studies

Table 58 Details of relevant ongoing studies

Study	Population	Intervention	Comparator	Outcomes
Sunrise				
Study: SUNSAS study; NCT05057975 2021[39] **Country:** France **Estimated completion date**: March 2024	18 to 80 years old, referred for a suspicion of OSA, **Estimated recruitment:** n=848	Sunrise (home)	PSG (In lab or as outpatient)	Non-inferiority to the normal practice (i.e., the PSG) on sleepiness at 3 months post-diagnosis. Superiority to the PSG (outpatient or in-lab) in terms of time between inclusion and diagnosis appointment and treatment initiation
Study: Service Evaluations NHS Scotland unknown date[40] **Country:** Scotland **Estimated completion date:** not reported	Adults with suspected OSAHS **Estimated recruitment:** n=100	Sunrise (unclear setting)	"Detailed sleep test" (unclear setting)	Not reported
Study: NHS Foundation Trust Cambridge unknown year[41] **Country:** UK **Estimated completion date**:2023 (ongoing at time of company submission)	Children with suspected OSAHS **Estimated recruitment:** n=100	Sunrise (Home. In-lab if significant co-morbidities or aged <9 years of age)[a]	Home cardio-respiratory polygraphy with Transcutaneous Carbon Dioxide monitoring (Home. In-lab if significant co-morbidities or aged <9 years of age)[a]	Diagnostic accuracy, specificity and sensitivity
[a] Sunrise will be used simultaneously on the same night with home cardio-respiratory polygraphy with Transcutaneous Carbon Dioxide monitoring				

Appendix 3 Data extraction template used in the systematic review of clinical effectiveness

DAP 70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

EAG Clinical effectiveness data extraction form

1. Study overview

Full data extracted by Name and date completed	
Full data checked by	
Fast tracked data extracted by	
Fast tracked data checked by	
Study lead author and year For studies with linked publications this should be the primary publication of the study.	
Rayyan database number	
Linked publications Lead author, year and Rayyan number. Confirm if any relevant data from linked publications have been included in this data extraction	
Supplemental material Confirm if any relevant data from supplemental material have been included in this data extraction	
Study design As stated by the study authors. Note whether prospective/retrospective	
Country For any UK studies record details of the region/locality/city or town	
Number of study centres State number of centres and their name(s)	
Study objective/hypothesis As stated by the study authors	
Rationale for study	

Add any information on the justification for/purpose of the study, if not already encapsulated by the objective/hypothesis. E.g. "the purpose was to assess diagnostic performance in a real world setting"	
Interim / final analysis If interim, state the expected date when the final analysis will be available	
Primary outcome(s) For diagnostic performance outcomes state the criteria used e.g. AHI/ORDI	
Secondary outcome(s) For diagnostic performance outcomes state the criteria used e.g. AHI/ORDI	
Clinical criteria for OSA State which criteria were used for diagnosing OSA	
Clinical criteria for hypopnoea State which criteria were used for diagnosing hypopnoea syndrome	

2. Participant recruitment and sampling

Number of participants enrolled State the number of participants enrolled/randomised in the study.	
Recruitment and selection of participants Summarise procedures for recruiting participants (e.g. a consecutive series/random sample). Record dates given for stating and completing recruitment.	
Study inclusion criteria	
Study exclusion criteria	
Sample size calculation	
Sample attrition/dropout Record the number who completed/withdrew from the study and the reasons for any study drop-outs/withdrawals. Note any discrepancies in the number of people starting and completing/withdrawing from the study.	

3. Participant baseline characteristics

Characteristic, Record – where applicable - unit of measurement, mean/median values, variance/confidence intervals, number and percentage of patients	**Intervention**	**Comparator / reference standard**
Age, years		
Ethnicity, n (%)		

BMI / Weight, n (%)	
Sex (M/F), n (%)	
Systolic blood pressure, n (%)	
Diabetic, n (%)	
Type of diabetes	
COPD, n (%)	
Chronic kidney disease, n (%)	
Neuromuscular disorders, n (%)	
Other comorbidities, n (%)	
Currently pregnant, n (%)	
Baseline ESS score/ other inventory score, n (%)	

4. Sleep study details

NB. Use the first table below to record details of the **intervention** and the second table for details of the **comparator / reference standard** tests.

Characteristics	**Intervention**
Part A – Device details	
Novel device name State version number, manufacturer, sleep study classification type (I-IV) if stated. Also, state any rationale given for the use of the device in this study.	
Setting of sleep study e.g. home, hospital/clinic	
Software used State the name of any software/App used to process/transmit data during the sleep study	
Signal channels used List all the channels used if applicable. If a multichannel sensor is used record the parameters included.	
Consumables used State any consumables used during the sleep study.	
Criteria for successful/failed recording State any criteria for a sleep study to be viable, such as the minimum recording time	
In attendance	

State any persons, other than the patient and their family/carer, who attended the sleep study. State if attendance was in person or virtual, and what role they had in the sleep study.	
Patient training Record details of any instruction patients were given on how to use the device	
Professional training Record details of any staff instruction on how to use the device/show others how to use the device	
Sequence of testing State the order in which tests were done	
Device transportation Record how the device and other equipment were transported to and from the patients' home	
Care pathway Record details of any care pathway for patients diagnosed with OSA used in the study	
Part B – Signal scoring details	
Signal scoring method State whether automated/manual/both.	
How long did scoring take? (e.g. mean/median/range),	
Describe any blinding procedures during scoring	
Was scoring done in duplicate? (i.e. each sleep study independently checked by a second person?)	
If automation, record any details of the algorithm used	
Who received the sleep study report?	

Characteristics	**Comparator / reference standard**
Part A – Device details	
Device name State version number, manufacturer, sleep study classification type (I-IV) if stated. Also, state any rationale given for the use of the device in this study	
Setting of sleep study e.g. home, hospital/clinic	
Software used	

State the name of any software/App used to process/transmit data during the sleep study	
Signal channels used List all the channels used if applicable. If a multichannel sensor is used record the parameters included.	
Consumables used State any consumables used during the sleep study.	
Criteria for successful/failed recording State any criteria for a sleep study to be viable, such as the minimum recording time	
In attendance State any persons, other than the patient and their family/carer, who attended the sleep study. State if attendance was in person or virtual, and what role they had in the sleep study.	
Patient training Record details of any instruction patients were given on how to use the device	
Professional training Record details of any staff instruction on how to use the device/show others how to use the device	
Sequence of testing State the order in which tests were done	
Device transportation Record how the device and other equipment were transported to and from the patients' home	
Care pathway Record details of any care pathway for patients diagnosed with OSA used in the study	
Part B – Signal scoring details	
Signal scoring method State whether automated/manual/both.	
How long did scoring take? (e.g. mean/median/range),	
Describe any blinding procedures during scoring	
Was scoring done in duplicate? (i.e. each sleep study independently checked by a second person?)	
If automation, record any details of the algorithm used	
Who received the sleep study report?	

5. Results: intermediate outcomes - measures of performance to detect OSAHS and assess severity

Use the following table to record the proportion of patients by severity category and mean AHI/RDI scores (as appropriate) by severity category (where reported). Adapt the table as appropriate to the study being data extracted.

OSA Severity	AHI criteria	Patients N (%)	Reference std Mean (SD) AHI score	Index test Mean (SD) AHI score	
No OSA	<5				
Mild	5 to 14.9				
Moderate to severe	15 to 29.9				
Severe	≥30				

Repeat the following table for each diagnostic threshold/cut off and for each NICE scope sub-group where available

	Reference standard positive	Reference standard negative	Total
Index test positive	a	b	a+b
Index test negative	c	d	c+d
Total	a+c	b+d	a+b+c+d
Accuracy			
Calculate clinical sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) if possible and note whether these agree with any values that may be reported in the paper. Use https://www.medcalc.org/calc/diagnostic_test.php to assist with calculations			
Diagnosis	**Value**		**95% CI**
Clinical sensitivity a / (a + c)			
Clinical specificity d / (b + d)			
PPV a / (a + b)			
NPV d / (c + d)			
Positive likelihood ratio [sensitivity/(1-specificity)]			
Negative likelihood ratio [(1-sensitivity)/specificity]			
Disease prevalence			
Other *(add as necessary)*			
Comments: *e.g. Calculations agree with values reported in paper. Note if any cases where 0.5 added to values to avoid division by zero when calculating diagnostic odds ratio. Add an asterisk to denote where values have been calculated by the reviewer.*			

6. Results: other intermediate outcomes

Outcome in NICE scope	Intervention	Comparator / reference standard	*P-value / CI / Other relevant statistic (e.g. ORs)*

Agreement / concordance Measures of concordance or agreement between diagnostic tests			
Impact on clinical decision-making			
Time to interpret device outputs			
Time to diagnosis or starting treatment			
Test failure rate E.g. where recordings are uninterpretable; insufficient sleep time			
Number of repeat sleep studies done Note if any other action was taken in in the event of test failure (e.g. hospital PSG).			
Use of healthcare resources E.g. number and length of hospital admissions, use of pharmacological and nonpharmacological interventions for management of OSAHS			

7. Results: patient outcomes

Outcome in NICE scope	Intervention	Comparator / reference standard
Health-related quality of life		
Ease of use and acceptability for patients and carers		
Patient and carer experience		
Add any other relevant outcomes		

8. Results: costs

Outcome in NICE scope	Intervention	Comparator / reference standard
Costs of devices (including any additional software or hardware) including any additional software or hardware		
Costs related to using the interventions (including any time analysing and storing data, communicating results, and		

arranging for use of the technology) including any time analysing and storing data, communicating results, and arranging for use of the technology		
Cost of maintenance of testing equipment		
Cost of sending testing equipment to people's homes		

9. Adverse effects

Outcome	Intervention	Comparator / reference standard
Adverse effects Report any adverse effects arising from the use of the test and any arising from subsequent treatment and care		

10. General reviewer comments (e.g. importance, methodological issues)

Comments

Appendix 4 Further information on studies included in the systematic review of clinical effectiveness

Table 59 References for included studies

Primary study reference	No. of secondary references	Secondary references
Studies in Adults		
Alsaif et al., 2023[17]	2	Alsaif et al., 2022[18], NCT05204004 2022[382]
Devani et al., 2021[19]	1	NCT03544086 2018[383]
Kelly et al., 2022[27]	1	NCT04262557 2020[384]
Lyne et al., 2023[385]	-	-
Martinot et al., 2017[21]	-	-
Massie et al., 2018[22]	-	-
Massie et al., 2022[23]	-	-
Mueller et al., 2022[29]	-	-
Pepin et al., 2020[26]	1	Martinot et al., 2022[386]
Sanchez Gomez (2024) [20]	1	Phase 1 Virgen Macarena trial 2023[15] (company submission AIC report)
Pillar et al., 2020[31]	1	NCT02369705 2015[387] (share this NCT record with Tauman)
Storey et al., 2022[28]	-	-
Tauman et al.,2020[30]	Shares same NCT record as Pillar	NCT02369705 2015[387]
Van Pee et al., 2022[24]	1	NCT04191668 2019[388]
NCT04764734 2021[16] NightOwl study. Meets inclusion criteria, is completed but does not report results. Principal Investigator has been contacted but no response.		
Studies in Children		
Martinot et al., 2015[42]	-	-
Martinot et al.,2022[35]	-	-
NCT04031950 2019[32]	3 (company submission AIC reports)	[redacted]37 [redacted]71 [redacted]72
The '-' symbol means that there were no secondary references		

Table 60 Characteristics of study participants in the systematic review of clinical effectiveness (people over 16 years)

Study. Sample size	Age, years, mean ± SD	Sex, n (%)	Ethnicity & race, n (%)	BMI kg/m^2 mean ± SD	Co-morbidities, n (%)	ESS score, mean ± SD (0-24)
Novel devices compared to RP						
Devani et al (2021)[19] AcuPebble Sample n=150	44 ± 11. Range 21 - 65 yrs	M= 10 (71.3) F= 43 (28.7)	White British, 47 (31) White other, 19 (12.67) Asian or Asian British (excluding the ones below), 31 (20.67) Black or Black British (excluding the ones below), 3 (2), Indian, 2 (1.33) Pakistani, 2 (1.33) White or Black African, 2 (1.33) Chinese, 1 (0.67) White or Black Caribbean, 5 (3.33) Other 38 (25.34)	31.2 ± 7.6 Range 17.6 - 56.6[a] Weight (kg) 95.3 ± 25.7 Range (45.7-190)[b]	High blood pressure= 38 (25.3) Diabetic= 17 (11.3)	Not reported
Alsaif et al (2023)[17] Sunrise [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mueller et al (2022)[29] WatchPAT 300 Sample n=56	44 ± 12. Range 19 - 76	M= 39 (70)[d] F= 17 (30)	Not reported	28.1 kg/m^2. Range 19.2 - 41.4	Not reported	Not reported
Storey et al (2022)[28] WatchPAT ONE Sample n=600	49 ± 13.6 (novel device) 50 ± 13 (RP)	Not reported	Not reported	Not reported	Not reported	Not reported
Novel devices compared to PSG						
Sanchez Gomez et al (2024) [20] AcuPebble Sample N=63[c]	50 ± 12 Range 25 - 78	M= 40 (63) F= 23 (37)	Not reported	50±7.0 (16.7 to 54.0)	Hypertension = 17 (26.98); Cardiac diseases =7 (11); Diabetic = 6 (9.52)	10.4 ± 6.8

Study. Sample size	Age, years, mean ± SD	Sex, n (%)	Ethnicity & race, n (%)	BMI kg/m^2 mean ± SD	Co-morbidities, n (%)	ESS score, mean ± SD (0-24)
Martinot et al (2017)[21] Brizzy, Sample n=79[d]	48.8 ± 14.6. Range 18 - 80	"Predominantly Males" (p.569)	Not reported	31.8 ± 7.7 Weight (kg), 95.1 ± 25.8	Not reported	10.0 ± 5.9
Massie et al (2018)[22]	53 ± 13	M= 56% F= 44%	Not reported	28.2 ± 4.9	Not reported	Not reported
Massie et al (2022)[23] , NightOwl, Sample n=261	54 ± 14 Range 21 - 84	M= 60% F= 40%	*"Persons of diverse racial and ethnic backgrounds were included",* no further information given.	30.0 ± 5.9. Range 18.2 - 53.8	Not reported	Not reported
Van Pee et al (2022)[24], NightOwl, Sample n=167	56 ± 15 Range 21-84	M= 63% F= 37%	Black, 43 (26) White, 124 (74) Hispanic, Latino, or Spanish, 92 (55)[e]	30.7 ± 6.3. Range 18.2 - 53.8	Not reported	Not reported
Lyne et al (2023)[25] , NightOwl, Sample n=100	48 ± 11.8	M= 57 (57) F= 43 (43)	Not reported	33.6 ± 9.1	Not reported	7.7 ± 4.7
Pepin et al (2020)[26], Sunrise, Sample n=376	49.7 ± 13.2	M= 207 (55.1) F= 169 (44.9)	Not reported	31.0 ± 7.1	Not reported	Not reported
Kelly et al (2022)[27], Sunrise, Sample n=31	48 ± 15	M= 58% F= 42%	Not reported	30.4 ± 7.6	Not reported	Not reported
Supporting evidence [f]						
Pillar et al (2020)[31], WatchPAT 200U, N=84	57 ± 16 Range 22 - 93	M= 54 (64) F= 30 (36)	Not reported	29.8 ± 5.7; Range 17–45	Diabetic = 19 (23) Heart failure = 41 (48)	Not reported
Tauman et al (2020)[30], WatchPAT 200U, N=101	68 ± 12 Range 17–48	M= 70 (69.3), F= 31 (30.69)	Not reported	31 ± 5.2 Range 22–87	Hypertension = 80 (79) Type 2 Diabetes = 42 (42) Atrial fibrillation (permanent) = 12 (11.8)	Not reported

ESS Epworth Sleepiness Scale
[a] Self reported. Data available from 128 patients (84.2%)
[b] Self reported. Data available from 129 patients (84.9%)
[c] Baseliine characteristics are for 63 patients evaluated in the study
[d] Baseline characteristics are for the 79 patients evaluated in the study
[e] Figures only available for the US study centres. Also, the numbers of participants across the, presumably mutually exclusive ethnicity categories, (n=259) far exceeds the total number of participants in the study sample (n=167)
[f] see section 4.1 for an explanation of supporting evidence.

Table 61 Characteristics of study participants in the systematic review of clinical effectiveness (children and young people aged 2 to 16 years)

Study Sample size	Age, years, median	Sex, n (%)	Ethnicity and race	BMI kg/m^2 median	Co-morbidities, n (%)	ESS score, mean ± SD (0-24)
NCT04031950 (2019)[32] AcuPebble Sample size N= 199	Not reported[a]	Not reported[a]	[redacted]	Not reported[a]	[redacted]	Not reported[a]
Martinot et al (2015)[33] Brizzy Sample size N=33	5.0 Range 2 -15 yrs	Sex ratio (M/F) = 2.3 (23/10).	Not reported	15.8 Range 12.4-38.0	Not reported	Not reported
Martinot et al (2022)[35] Sunrise Sample size N=140	6.90 (IQR 5.6)	M= 63 (45%) F= 77 (55%)	Not reported	Median (IQR) 16.46 (3.57)	Not reported	15.5 ± 3.8

[a] Data were provided for both trial sites combined, but not separately for the site for which preliminary results have been reported.
[b]Data were provided for both trial sites combined. In addition, individual patient data ,with site identifiers was provided. This allowed the EAG to calculate the number of patients with specific co-morbidities in the site for which preliminary results were reported.
IQR Inter-quartile range

Table 62 Device failure rates reported in studies included in the systematic review of clinical effectiveness (people over 16 years)

Study ID	Sleep study failures by novel device and comparator
Novel devices compared to RP	
Devani (2021)[19] AcuPebble, N=182	**AcuPebble SA100** Total exclusions n=16/182 (11.25%) • Forgot to do test, n=7 • Played with the App before going to sleep, logging themselves out then unable to log back into App without password, n=4 • Started the test but forgot the phone was in another room before going to sleep, n=1 • Given the wrong kit for the study, n=4 **Home RP** Total exclusions n=16/182 (11.25%) • Suffered problems with CR-PG[a] to the point that data recorded was not valid for diagnosis, n=16 Of the remaining 150 participants included in the evaluation, all were able to use the AcuPebble SA100 and all the signals from AcuPebble SA100 were valid and analysable.
Alsaif (2023)[17b] Sunrise N=17	[redacted]
Storey (2022)[28] WatchPAT ONE N= 600	Not reported
Mueller (2022)[29], WatchPAT 300, N=61	**WatchPAT "200/300"** Needed repeat examination, n=2/61 (3%) • Operating error by the patient (forgot to switch on device), n=2 • Early termination due to pain from the finger probe in a patient with unusually large fingers, n=1

	No technical failures related to insufficient recording time or loss of sensors. **RP n=4/61** Needed repeat examination, n=8/61 (13%) • Complete testing failure due to miscommunication with the technical staff, n=1 • Recording time under 6h due to unknown technical failures, n=2 • Loss of the nasal pressure sensor or inadequate examination time (< 4 h) achieved, n=3 • device did not start recording due to an operating error by the technical assistant, n=2 Of 61 enrolled participants, 5 were excluded from the analysis because they were missing data for 1 of the 2 study tests due to logistics and loss to follow-up). Technically adequate RP testing defined as all channels providing non-interrupted information for at least 6 h (2012 update of 2007 AASM criteria). Authors report no statistical differences in the need to repeat the tests (p =0.22) but they report "a tendency" for "less failures" with WatchPAT testing.
Novel devices compared to PSG	
Sanchez Gomez et al (2024) [20] N=80	**AcuPebble SA100** Total exclusions n=9/80 (11.25%) • n=1 No AcuPebble SA100 diagnostic output (defective mobile phone) • n=3 no AcuPebble SA100 recording (i.e. test was not started). • n=5 output of the AcuPebble SA100 "Inconclusive" **In-hospital PSG** Total exclusions n=8/80 (10%) • n=2 did not have the gold standard reference output. • n=6 short reference recording length 17/80 (21%) sleep studies excluded due to one or both devices not producing a valid diagnostic output.
Martinot (2017)[21], Brizzy, N=100	**Brizzy** Technically unacceptable tests n=4/100 (4%) • n=4 (4%) failure to capture the Brizzy MM signal into the PSG **PSG** Technically unacceptable tests n=4/100 (4%) • n=3 (3%) poor oximetry recording • n=1 (1%) loss of belts signal. Authors comment Brizzy has an "acceptable failure rate" (<10%) in keeping with published recommendations for unattended PSG.

Massie et al (2018)[22] N=101	Not reported	
Massie (2022)[23], NightOwl, N=261	Study inclusion criteria was restricted to participants with a valid sleep study (n=261). The study, therefore, does not report failure rates. Criteria for adequate test: • All PSG channels interpretable by technicians (e.g. no detachments of the nasal cannula or pulse oximeter), • At least 4hr of analysable signal obtained for the PAT (Sunrise) test	
Van Pee et al (2022)[24] N=228	**NightOwl** Data acquisition failure or technically inadequate recording, n=41/228 (18%) • Defect in PAT sensor causing it not to acquire data: n=3. • Patient detached PAT sensor: n=2 • PAT data acquisition not started: n=7 • Data acquisition App loss of data: n=16 • PAT recordings technically inadequate: n=13 **PSG** Data acquisition failure or technically inadequate recording, n=20/228 (8.7%) • PSG flow issue (scoring not possible): n=4 • PSG Sp02 issue (scoring not possible): n=7 • PSG admin error (PSG was only single-scored or the link between the PAT and PSG could no longer be retrieved): n=9	*Criteria for valid test* NightOwl technically adequate if at least 4 h of analysable signal obtained. PSG technically inadequate if: • one of the channels could not be interpreted by the technicians; • PSG data or any annotations of the two scoring centers were missing due to administrative errors Technically inadequate recordings were excluded from further analysis
Lyne et al (2023)[25] NB. 115 enrolled	**NightOwl Mini (disposable)** signal acquisition failures, n=6/107 **NightOwl Reusable** signal acquisition failures, n=4/107 **NightOwl and PSG** signal acquisition failures, n=7/107 **NightOwl Mini (disposable)** failure rate = 12.1% (13/107 applications) **NightOwl Reusable failure rate** = 10.3% (11/107 applications).	*Criteria for valid test* Participants who failed to attain a total sleep time of less than four hours as recorded by the in-laboratory PSG on the study night were not included in the analysis. In contrast, NOM and NOR results were included if three hours of data were acquired, in accordance with the proprietor's determined threshold for an interpretable study
Pepin et al (2020) [26] N=376	**Sunrise** - No technical failures occurred with use of Sunrise system. **PSG** not reported (no further information provided)	
Kelly et al (2022)[27] N=31	**Sunrise** Technical failures, n=4/38 (10.5%) • Bluetooth connection lost, n=3 • Sunrise sensor disconnected, n=1. **Home PSG**	

	Technical failures (all poor quality signals), n=3/38 (8%)
Supporting evidence	
Pillar et al (2020)[31]	Not reported
Tauman et al (2020)[30]	The presence of AF did not cause significant non-valid PAT signal. The mean excluded sleep time was 4.4% in the 46 patients with active AF compared with 6.4% in patients without active AF.
[a] CR-PG cardiorespiratory polygraphy	

Table 63 Assessment of health care resources and costs associated with AcuPebble of (Devani et al 2021)

Use of healthcare resources	Intervention	Comparator
Cleaning	0.5 min	2 min
Device preparation	0.5 min	10 min
Time of healthcare professional training patient on using the device	0	30 min
Analysis of signals by experts to issue a diagnosis	0	60–120 m
Cost	~£1	£250–£500[a]
[A] This range has been calculated taking into account the variation in the time spent by different healthcare professionals analysing signals (60–120 min), as well as their cost in the UK NHS. The numbers have been obtained using the tool provided by the UK National Institute for Health Research, version 2019. It has been assumed that the training of the patient is done by a nurse or allied health professional and the analysis by a clinician.		

Table 64 Assessment of health care resources and costs associated with WatchPAT ONE versus NOX T3 (home RP) (Storey et al 2022)

Use of healthcare resources	Intervention[a]	Comparator[a]
Number of appointments not attended by patients	13	42
Cost per appointment (including equipment, room, staff and postage)	£73.16	£39.91
Mean staff time take per appointment (from check in to check out but excluding analysis) in minutes	12	21
[a]300 patients were enrolled in each study arm however, it was not reported whether data for all 600 patients were included in the analyses		

Table 65 Patient voluntary App usability questionnaire results of AcuPebble SA100 versus RP (Devani et al., 2021)

Question	Response (n =123/150 respondents)
Q1. I managed to follow all the steps on the mobile app without assistance.	Yes = 97.6% No = 2.4%
Q2. I understood all instructions in the phone/tablet.	Strongly agree/Agree = 99.2% Disagree = 0.8%
Q3. I felt confident using the app on the phone/tablet.	Strongly agree/Agree = 99.2% Disagree = 0.8%
Q4. It was easy to attach the sensor to my neck.	Strongly agree/agree = 98.4% Neither Agree nor Disagree = 1.6%
Q5. I had no problem replacing the adhesive (sticky paper) on the sensor.	Strongly agree/Agree = 77.7% Neither Agree nor Disagree = 22.3%
Q6. The sensor on the neck was more comfortable than the other sensors on my body	Strongly agree/Agree = 90.2% Neither Agree nor Disagree = 8% Disagree = 1.6%
Q7. The sensor on the neck was easier to attach than the combination of all the other sensors on my body.	Strongly agree/Agree = 96.7% Neither Agree nor Disagree = 3.3%

Table 66 Questionnaire responses evaluating the quality of sleep and discomfort levels (Mueller et al., 2022)

	RP			WatchPAT	
	Yes	**No**		**Yes**	**No**
Questions asked after a night of testing with RP (n=55 patients) and WatchPAT (n=54 patients), respectively					
Slept well during the night	20 (39%)	32 (61%)		39 (74%)	15 (26%)
Falling asleep disturbed by the test	33 (60%)	22 (40%)		6 (12%)	46 (78%)
Lost sensors during the night (e.g. nasal canula, finger sensor)	11 (20%)	44 (80%)		3 (6%)	51 (94%)
Experienced pain during the test	3 (5%)*	52 (95%)		7 (13%)**	47 (87%)
Woken up by the test***	28(50%)	27 (50%)		16 (30%)	38 (70%)
Number of awakenings	Mean 1.8, median 1 Range 0, 10			Mean 0.62, median 0 Range 0, 6****	
Questions asked after both nights testing					
Slept better (RP vs WatchPAT)	11 (20%)			45 (80%)*****	
Preference for future testing? (RP vs WatchPAT)	6 (12%)			49 (88%)*****	

NB. Some patients abstained from answering some of the questions
* pain due to the nasal cannula dynamic pressure measurement and due to the device itself.
** described as "pain in the exposed finger during PAT" "at the side of the finger probe"
*** described as "awakenings subjectively related to the testing device"
**** The subjective perceived number of awaking was significantly lower during PAT compared to RP (p=0.004).
***** Described by authors as "Significantly more patients" (RP vs WatchPAT). It is not stated if they mean statistically significant.

Table 67 Patient self-reported overall sleeping comfort for RP (n=56) and WatchPAT (n=56) (Mueller et al., 2022)

Device	**Sleep grade**					
	1 (best)	**2**	**3**	**4**	**5 (worst)**	**6 (no rating reported)**
RP, n	3	11	19	12	8	3
WatchPAT, n	24	24	5	2	0	1

NB. These data are presented as barcharts in the study publication (Figure 2); for ease of interpretation they have been transposed here by the reviewer into table format

Table 68 Responses to voluntary usability questionnaire (NCT04031950, 2019)

Ease of use and acceptability for patients and carers
[redacted]

Appendix 5 Critical appraisal of studies included in the systematic review of clinical effectiveness

Individual Risk of Bias and Applicability assessments for included studies using the QUADAS-2 instrument[13]

Alsaif et al., 2023[17]

Device: Sunrise **Secondary papers:** Alsaif et al., 2020[18], NCT05204004 2022[382], Sunrise stakeholder comments		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Unclear	None of the information for this study states whether the sample enrolled is consecutive or random.
Signalling question 2: Was a case-control design avoided?	Yes	[redacted]
Signalling question 3: Did the study avoid inappropriate exclusions? *(Note: Remember that the device may be contraindicated in certain patient populations)*	Unclear	
Judgment: Could the selection of patients have introduced bias?	**RISK:** UNCLEAR	
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** UNCLEAR	[redacted]

		[redacted]
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard? *(Note: Consider whether the index test was automatically scored by the software only, and could therefore be considered independent of the results of the reference standard)*	Yes	States "*MM monitor sleep reports were automatically generated"*
Signalling question 2: If a threshold was used, was it pre-specified? *(Note: for AHI and ODI, the following thresholds are standard (NICE scope, EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per. If these specific thresholds are used but NOT prespecified we will not consider this an increase risk of bias)*	Unclear	Response 18 from Sunrise stakeholder comments states *"For scoring OSA severity, the Sunrise device utilises pre-specified thresholds established in Pepin et al., 2020. These thresholds are clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included"* We acknowledge that the Sunrise device utilises pre-specified thresholds established in Pepin et al., 2020. However, it cannot necessarily be assumed that these thresholds were used in the Alsaif study. We note that Pepin et al., 2020 was not cited in the draft manuscript.
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** UNCLEAR	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Unclear what thresholds were used. Index was conducted in the home
DOMAIN 3: REFERENCE STANDARD		
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"Home respiratory polygraphy has been validated in the screening and diagnosis of sleep apnoea (18).*

		ApnoeaLink Air (ResMed Ltd, Australia) was used in this study."
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	[redacted]
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING		
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Simultaneous recording.
Signalling question 2: Did all patients receive a reference standard?	Yes	No comment.
Signalling question 3: Did patients receive the same reference standard?	Yes	No comment
Signalling question 4: Were all patients included in the analysis?	No	[redacted]
Judgment: Could the patient flow have introduced bias?	**RISK:** UNCLEAR	[redacted]

		[redacted]

[a] with caveat that the device may be contraindicated in certain patient populations
[b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard
[c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.

Devani et al., 2021[19]

Device: AcuPebble SA100 **Secondary papers:** NCT03544086 2018[383]		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"Consecutive patients aged between 18 and 70 who were referred for evaluation of possible OSA to the Sleep and Ventilation clinic at the Royal Free London Hospital NHS Foundation Trust were recruited for the study."* (p2)
Signalling question 2: Was a case-control design avoided?	Yes	*"Patients who consented to participate in the study were issued with both a CR-PG device as per usual clinical care and an AcuPebble SA100 device with a smart phone. Both the CR-PG and AcuPebble SA100 were to be used simultaneously."* (p2)
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	*"All adult patients were included except for those above the age of 70; subjects who were not fluent in English or had special communication needs; those with a known allergy to adhesive dressings; subjects with physical or mental impairments, which would make them unable to use the new technology on their own; subjects with electronic body implants and subjects with extremely lose skin in the neck area, which would make the device swing if the neck moved."*(p2)
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment:	**CONCERN:** LOW	Note that patients above 70 were excluded.

Is there concern that the included patients do not match the review question?		
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	*"In order to test the efficacy of AcuPebble SA100 for the four different diagnostic outputs, the CR-PG signals were manually and independently scored by two experienced clinician scorers, resulting in these four diagnostic indices and corresponding diagnosis. The clinicians were blinded to the outputs from the AcuPebble SA100. At the end of the study, the diagnostic results derived from the experts manual scoring were compared with AcuPebble SA100 automated diagnosis."* (p4)
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	"using the current AASM diagnostic criteria."(p4)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** LOW	Index test was conducted at home
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"The domiciliary CR-PG system used in this study was the type III monitor; Embletta MPR Sleep System (Natus Medical, California) and accompanying Embla Remlogic software (Natus Medical, California)…. This device was chosen for two reasons: the system is routinely used in the Sleep and Ventilation clinic at Royal Free London NHS Foundation Trust for diagnosis of sleep disordered breathing and meets the American Academy of Sleep Medicine's (AASM) technical adequacy requirements to be considered a gold standard for ambulatory diagnosis of the disease"* (p3)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	*"In order to test the efficacy of AcuPebble SA100 for the four different diagnostic outputs, the CR-PG signals were manually and independently scored by two experienced clinician scorers, resulting in these four diagnostic indices and corresponding diagnosis. The clinicians were blinded to the outputs from the AcuPebble SA100. At*

		the end of the study, the diagnostic results derived from the experts manual scoring were compared with AcuPebble SA100 automated diagnosis." (p4)
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	*"Both the CR-PG and AcuPebble SA100 were to be used simultaneously."* (p2)
Signalling question 2: Did all patients receive a reference standard?	Yes	*"Patients who consented to participate in the study were issued with both a CR-PG device as per usual clinical care and an AcuPebble SA100 device with a smart phone. Both the CR-PG and AcuPebble SA100 were to be used simultaneously."* (p2)
Signalling question 3: Did patients receive the same reference standard?	Unclear	*"The domiciliary CR-PG system used in this study was the type III monitor; Embletta MPR Sleep System (Natus Medical, California) and accompanying Embla Remlogic software (Natus Medical, California). The following signal channels from the CR-PG device were used for analysis: thoracic and abdominal piezoelectric respiratory movement sensors peripheral pulse oximetry, nasal thermistor air flow sensor, snore and body position."* (p3) Figure 2 *"4 patients were given the wrong kit for the study"* (these were excluded from the analysis)(p3)
Signalling question 4: Were all patients included in the analysis?	No	*"One hundred and eighty-two consecutive patients were recruited for evaluation over an 8-month period between November 2018 and July 2019. One hundred and twenty-nine (71%) were men. Of these studies, 150 could be used for evaluation purposes*" (p4) Flow diagram on next page (Figure 2 from paper) shows the reasons for exclusions. None appear unreasonable.
Judgment:	**RISK:** LOW	No comment

Could the patient flow have introduced bias?		
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Kelly et al., 2022[27]

Device: Sunrise **Secondary papers:** NCT042662557 2020[384]		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"Forty consecutive adult patients."* (p.2)
Signalling question 2: Was a case-control design avoided?	Yes	*"All 40 participants underwent an overnight PSG (the reference method) with simultaneous MM recordings using the Sunrise system (Sunrise SA, Belgium)."* (p.2)
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	*"Forty consecutive adult patients undertaking a diagnostic home sleep study for suspicion of OSA were invited to participate. Participants had to be able to use portable devices and smartphones. All 40 participants underwent an overnight PSG (the reference method) with simultaneous MM recordings using the Sunrise system (Sunrise SA, Belgium)."* (p.2)
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	*"The recorded MM data were automatically transferred from the smartphone to a cloud-based infrastructure at the end of the*

		night. These data were then analysed using a dedicated machine learning algorithm." (p.3)
Signalling question 2: If a threshold was used, was it pre-specified?[c]	No	*"We performed an area under the curve (AUC), and a post hoc analysis to optimise the cut-off points of MM-ORDI for diagnostic decisions, compared with the criterion-standard cut-off values of obstructive PSG-ORDI recommended in ICSD-3 (5 events/hour and 15 events/hour). The optimal MM cutoffs were assessed at the highest value of the Youden index (sensitivity plus specificity minus 1). Finally, we calculated the metrics of clinical utility and accuracy for the optimal detection thresholds and the post-test probability for each cut-off point recommended by the Portable Monitoring Task Force of the AASM (Collop et al., 2007)."* (p.3)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** HIGH	A post-hoc analysis optimised the cut-off points of MM-ORDI for diagnostic decisions
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** HIGH	A post-hoc analysis optimised the cut-off points of MM-ORDI for diagnostic decisions
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"In-home PSG was recorded with a portable acquisition system (Nox A1, ResMed, Saint-Priest Cedex, France)." (p.2)* *"Polysomnography recordings were initially scored by experts from the recruiting centre (Grenoble Alpes University Hospital, France). PSG were anonymized, converted in European data format (EDF) and sent via a secured platform for blinded scoring to the second reference centre (Imperial College London, United Kingdom). Scoring was performed according to the recommended criteria established by the American Academy of Sleep Medicine (AASM) Manual for the Scoring of Sleep and Associated (Berry et al., 2012)."* (p.3)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	*"PSG recordings were analysed blinded to the MM data*" (p.3)
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment

B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	*"an overnight PSG (the reference method) with simultaneous MM recordings using the Sunrise system (Sunrise SA, Belgium)."* (p2)
Signalling question 2: Did all patients receive a reference standard?	Yes	*"All 40 participants underwent an overnight PSG (the reference method) with simultaneous MM recordings using the Sunrise system (Sunrise SA, Belgium)."* (p.2)
Signalling question 3: Did patients receive the same reference standard?	Yes	*"All 40 participants underwent an overnight PSG (the reference method) with simultaneous MM recordings using the Sunrise system (Sunrise SA, Belgium)."* (p.2)
Signalling question 4: Were all patients included in the analysis?	No	*"data from 31 participants were included in the analysis. Two participants withdrew, and there were three technical failures of PSG (poor quality signals) and four technical failures of the Sunrise device (Bluetooth connection was lost for three patients and for one patient the Sunrise sensor became disconnected)."* (p.3)
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	Exclusions seem reasonable with the majority due to technical failures of a device.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Lyne et al., 2023[385]

Device: NightOwl **Secondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		

Signalling question 1: Was a consecutive or random sample of patients enrolled?	Unclear	This is a prospective cohort study. The methods state *"A maximum of two eligible participants per night"* (p.1430) but it is unclear if these participants were consecutive or not.
Signalling question 2: Was a case-control design avoided?	Yes	Although not explicitly stated text indicates PSG, NOM and NOR done simultaneously "The trial involved single-night in-laboratory testing" (p.1434)
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	*"Exclusion criteria included patients with a known diagnosis of OSA undergoing a treatment PSG or patients unable to provide informed consent."* (p.1430)
Judgment: Could the selection of patients have introduced bias?	**RISK:** UNCLEAR	The reasons for exclusion from the study seem appropriate but it is unclear how many patients may have been excluded because of the limit of two eligible participants per night and how the two eligible patients were chosen if there were more than two available for any one night.
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	The NightOwl devices (mini and reusable) acquire data that is interpreted by a software algorithm so that AHI can be calculated. Results are retrieved from cloud-based storage.
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	Used standard thresholds for OSA severities (no reference cited for these) "Secondary outcomes included the agreement between NOM and NOR and PSG with respect to diagnostic classification of OSA across four categories: no-OSA group (AHI < 5 events per hour), mild OSA (AHI 5–14 events per hour), moderate OSA (AHI 15–29 events per hour), and severe OSA ≥ 30 events per hour." (p.1430)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index test conducted in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**

A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	"PSG data was collected using Compumedics Grael Profusion PSG system (version 4.5, build 531, Compumedics Limited). PSG studies were scored by an experienced sleep scientist according to the latest American Academy of Sleep Medicine criteria." (p.1430)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	Methods do not explicitly state whether the sleep scientists scoring the PSG studies had knowledge of the results of the index tests or not.
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	Sleep scientists were experienced and used American Academy of Sleep Medicine criteria and "participate in the American Academy of Sleep Medicine concordance program" (p.1434) so this should have limited any bias if the sleep scientists had knowledge of the index tests.
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment.
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Although not explicitly stated text indicates PSG, NOM and NOR done simultaneously "The trial involved single-night in-laboratory testing" (p.1434)
Signalling question 2: Did all patients receive a reference standard?	Yes	No comment.
Signalling question 3: Did patients receive the same reference standard?	Yes	"PSG data was collected using Compumedics Grael Profusion PSG system (version 4.5, build 531, Compumedics Limited)" (p.1430)
Signalling question 4: Were all patients included in the analysis?	No	Of 115 participants, 100 were included in the final analysis with fewer analysable from NOM (n=94) and NOR (n=96) devices. The 15 exclusions were for withdrawn consent (n=1), administrative error (n=1), PSG less than 4 hours of recording (n=6), both study devices had signal acquisition failure (n=7).
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	Patient flowchart provided (Fig 1 reproduced below). Aside from one exclusion for administrative reasons which is not further explained, the reasons for exclusion are appropriate.
[a] with caveat that the device may be contraindicated in certain patient populations		

[b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard
[c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.

Martinot et al., 2015[42] (child)

Device: Brizzy **Secondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	Abstract states consecutive children (also stated on p. 570).
Signalling question 2: Was a case-control design avoided?	Yes	PSG and MM were concurrent.
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	Consecutive children and no exclusion criteria given so presume there were no exclusions.
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment.
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Unclear	No mention of separate software for the Brizzy device which appears to have been connected directly to the PSG equipment *"The probes were connected to an electronic module and a measure of the distance was computed before transmission to the PSG*." (p.568) Figure 3A shows polysomnographic tracing which includes the MM signal suggesting a reader would see all the information at the same time but this is not explicitly stated.

Signalling question 2: If a threshold was used, was it pre-specified?[c]	Unclear	Table 1 shows four ways to analyse MM but it is not clear how these related to OSA diagnosis or whether the MM variables in Table 1 were pre-specified.
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** UNCLEAR	Focus of the paper is analysis of patterns of mandibular movement to detect respiratory effort associated with obstructive sleep apnoea-hypopnoea. Study doesn't report diagnostic accuracy results for Brizzy device against PSG or any other reference standard.
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** HIGH	Study doesn't report diagnostic accuracy results for Brizzy device against PSG or any other reference standard. Note index test was conducted in a sleep laboratory not home setting.
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"Routine laboratory-based PSG was recorded with a Dream Medatec device, Brussels, Belgium"* (p.568)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	page 570 of the journal article states *"All respiratory events were scored separately by 2 experienced observers who were blinded to the mandibular and the PTT records according to the American Academy of Sleep Medicine rules for the scoring of sleep and associated events reported in 2012"*
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	page 570 of the journal article states *"All respiratory events were scored separately by 2 experienced observers who were blinded to the mandibular and the PTT records according to the American Academy of Sleep Medicine rules for the scoring of sleep and associated events reported in 2012"*
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** UNCLEAR	No comment.
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Concomitant testing.
Signalling question 2: Did all patients receive a reference standard?	Yes	No comment.

Signalling question 3: Did patients receive the same reference standard?	Yes	*"Routine laboratory-based PSG was recorded with a Dream Medatec device, Brussels, Belgium*" (p.568)
Signalling question 4: Were all patients included in the analysis?	Yes	Paper does not suggest that any data were missing for any reason.
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	No comment.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Martinot et al., 2017[21]

Device: Brizzy **Secondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"The patients eligible to participate were consecutive subjects"* (p.568). Thirteen participants with "no specific sleep complaints were recruited by word of mouth" (p. 568), but the reviewers do not expect that this would have introduced bias.
Signalling question 2: Was a case-control design avoided?	Yes	The computed distance from the MM probes was transmitted to the PSG when PSG was conducted so both measures done on the same patient.
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	Consecutive patients were consenting adults "18 years and older with symptoms suggestive of sleep-disordered breathing (SDB) undergoing a single PSG."
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment

DOMAIN 2: INDEX TEST(S)	Assessment	Comments
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	*"Scoring for MM was performed by two blinded independent readers who had been trained to read MM tracings, while a different experienced reader analysed the standard PSG"* (p.568).
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	*"The outcome variable related to the diagnostic of the disease was based on a sensitivity/specificity analysis of MM device with the two different polysomnographic pre-specified cut-off values of RDI recommended in ICSD-3 (PSG-RDI ≥ 5 and ≥15/h TST). OSAS severity was evaluated from AHI, with <5, 5–15, 15–30 and >30/h TST representing the four severity categories*." (p.569)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index test carried out in a lab not a home setting
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	Study used "standardized in-laboratory PSG for the diagnosis of OSAS (ICSD-3, International Classification of Sleep Disorders, Third Edition)" and "a different experienced reader analysed the standard PSG, after de-identification of records." (p.568)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	Does not explicitly state that the PSG reader interpreted the PSG without knowledge of the index test results, but the PSG was analysed by a different reader. States PSG analyses were conducted after 'de-identification of records' but unclear what aspect of identification was removed (i.e. MM results or patient name/hospital number).
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	Although not explicit that the PSG reader didn't have knowledge of the index test results this seems likely to be the case.
B. Concerns regarding applicability		
Judgment:	**CONCERN:** LOW	No comment.

Is there concern that the target condition as defined by the reference standard does not match the review question?		
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Index and reference standard tests occurred at the same time during a single sleep test.
Signalling question 2: Did all patients receive a reference standard?	Yes	No comment.
Signalling question 3: Did patients receive the same reference standard?	Yes	"PSG was recorded with a Somnoscreen Plus®, Somnomedics, Randersacken, Germany." (Supplement)
Signalling question 4: Were all patients included in the analysis?	No	No patient flow diagram. 13 health subjects included to be true negatives & data presented for all 13. Data presented for 79 patients with suspected OSAS but states 8 recordings were "technically unacceptable" due to failure to capture MM signal into the PSG (4 participants), poor oximetry recording (3 participants) and loss of belts signal (1 participant). Therefore presume there were 87 patients with suspected OSAS at the start.
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	The reasons for excluding patients from the analysis all due to missing or poor data.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Martinot et al., 2022[35] (child)

Device: Sunrise **Secondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"consecutive pediatric patients clinically referred for a suspicion of OSA"* (p.1905)
Signalling question 2: Was a case-control design avoided?	Yes	Figure 1 shows concomitant recording by PSG and Sunrise.
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Unclear	*"Children with significant, chronic medical conditions, such as genetic syndromes, diabetes mellitus, craniofacial anomalies, or neurologic disease were excluded. Children receiving medications that could affect sleep (sedatives, or systemic corticosteroids) were also excluded."* (p.1905) Whilst many of the exclusions described seem appropriate this will limit the generalisability of the study.
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	Patient selection unlikely to have introduced bias but may have limited the generalisability of the study.
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment.
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	Sunrise device MM data automatically transferred to cloud-based infrastructure and a machine learning algorithm conducted the data analysis. (p.1906)
Signalling question 2: If a threshold was used, was it pre-specified?[c]	No	"*a post hoc analysis to optimize the cutoff points of Sr_RDI for diagnostic decisions compared to the goldstandard cutoff values of obstructive PSG_OAHI recommended in ICSD-3 (1 events/h, 5 events/h, and 10 events/h) was performed. The optimal MM cutoffs were determined at the highest value of the Youden index."*

		"After cutoff point optimization, we found that at the best thresholds of 5.75 events/h, 9.60events/h, and 13.07 events/h, Sr_RDI allowed for detection of patients with PSG_OAHI ≥ 1, PSG_ OAHI ≥ 5, or PSG_OAHI ≥ 10 with accuracy levels of 66%, 85%, and 94%, respectively." (p.1908)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** HIGH	A post-hoc analysis used to optimise cut off points.
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** HIGH	A post-hoc analysis used to optimise cut off points and the Index test was conducted in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"PSG scoring was performed by experienced technician who was blind to the study hypothesis and aims. The PSG data were manually scored in accordance with the recommended criteria in the scoring manual published by the AASM Manual for the Scoring of Sleep and Associated Events and using Domino version 3.0.0.1 software"* (p.1905)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	*"technician who was blind to the study hypothesis and aims"* (p.1905)
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Tests conducted concurrently.
Signalling question 2: Did all patients receive a reference standard?	Yes	No comment.

Signalling question 3: Did patients receive the same reference standard?	Yes	*"Routine laboratory based PSGs were recorded with XDream Medatec device (Medatec, Belgium)."* (p. 1905)
Signalling question 4: Were all patients included in the analysis?	No	*"Fifteen patients were excluded for the following reasons: incomplete data (7 patients), total sleep time inferior to 4 h (3 patients), and technical failures (5 patients)."*
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	Reasons for exclusion of patients were due to missing or incomplete data.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Massie 2018[22]

Device: NightOwl **Seondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Unclear	*"Data of 101 patients who underwent a diagnostic in-hospital PSG in the sleep laboratory of ZOL"* (p1971) *"A cohort of 101 participants were included"* (p1792). Unclear if this is a consecutive sample of patients
Signalling question 2: Was a case-control design avoided?	Yes	No comment
Signalling question 3: Did the study avoid inappropriate exclusions? *(Note: Remember that the device may be contraindicated in certain patient populations)*	Unclear	Exclusion criteria were not stated
Judgment: Could the selection of patients have introduced bias?	**RISK:** UNCLEAR	
B. Concerns regarding applicability		
Judgment:	**CONCERN:** UNCLEAR	Poor reporting of study. The authors have performed similar studies of NightOwl which were better reported (Massie 2022 and Van Pee

Is there concern that the included patients do not match the review question?		2022) and for which we judged low risk of bias and low concerns over applicability.
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard? *(Note: Consider whether the index test was automatically scored by the software only, and could therefore be considered independent of the results of the reference standard)*	Yes	NightOwl uses automated analysis
Signalling question 2: If a threshold was used, was it pre-specified? *(Note: for AHI and ODI, the following thresholds are standard (NICE scope, EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per. If these specific thresholds are used but NOT prespecified we will not consider this an increase risk of bias)*	Yes	Night owl REI diagnostic performance was referenced against PSG AHI severity cut-offs using AASM criteria which are standard. It is not stated what the REI cut offs were used (e.g. whether these were optimised)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index test conducted in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD		
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"diagnostic in-hospital PSG in the sleep laboratory"* (device name not stated)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	Not reported whether the PSG scorers were blind to the results of the index test. Note there was double scoring of PSG by local (manual) and external (automated with manual editing) somnologists
Judgment:	**RISK:** LOW	Although it is not stated that the PSG scorers were blind to the results of the index test, there was double scoring of PSG by local

Could the reference standard, its conduct, or its interpretation have introduced bias?		and external somnologists. Furthermore, the external scoring was automated with manual editing
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING		
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	PSG and NightOwl were used simultaneously
Signalling question 2: Did all patients receive a reference standard?	Yes	*"101 participants who underwent an in-laboratory polysomnography (PSG)"* (abstract)
Signalling question 3: Did patients receive the same reference standard?	Yes	Not reported, but all PSGs were carried out in the same centre in Belgium.
Signalling question 4: Were all patients included in the analysis?	Yes	From results can tell that sample size is 101, which equates to the number of participants in the study.
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Massie et al., 2022[23]

Device: NightOwl **Secondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"prospectively recruited in a consecutive cohort across four different centres"* (p3)
Signalling question 2: Was a case-control design avoided?	Yes	See response to signalling question 1
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	*"Underaged or mentally impaired participants were excluded from participating in the study"* (p3)
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	NightOwl uses automated analysis
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	Standard threshold was used (*"Sleep apnea severity was defined as follows: normal (AHI<5),mild (AHI 5–14.9),moderate(AHI 15–29.9)and severe(AHI≥30)"* p4)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment:	**CONCERN:** UNCLEAR	Index test was performed in a sleep laboratory not in a home setting.

Is there concern that the index test, its conduct, or interpretation differ from the review question?		
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"Qualified lab technicians at each participating study centre were responsible for setting up the equipment and capturing PSG data"* (p3) *"For the European and USA centres, respectively, the Alice 6 PSG (Philips Respironics, USA) and Cadwell Easy PSG (Cadwell, USA) were used"* (p3)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	*"All PSG data were double-scored by two independent centres that were blinded from one-another's analysis"* (p3) Not reported if the scorers were blind to the NightOwl results when scoring the corresponding PSG data
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	Although it is not stated that the PSG scorers were blind to the results of the index test, the double scoring of PSG by the independent centers blind to one-another's analysis should mean the risk of bias is low.
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	*"The strength of this study lies in its analysis of a large and independently double-scored dataset, which includes both PSG data and synchronously acquired PAT HSAT data."* (p8)
Signalling question 2: Did all patients receive a reference standard?	Unclear	"*Routine PSG was performed for all study participants"* (p3)
Signalling question 3: Did patients receive the same reference standard?	Unclear	Although all patients received PSG, two different devices were used "*For the European and USA centres, respectively, the Alice 6PSG (Philips Respironics, USA) and Cadwell Easy PSG (Cadwell, USA) were used*."(p3) The paper does not state whether these different PSG devices would give identical results.

Signalling question 4: Were all patients included in the analysis?	Yes	*"All 261 participants included in the dataset met the following data adequacy criteria: all PSG channels could be interpreted by the technicians (e.g. there were no detachments of the nasal cannula or pulse oximeter), and at least 4hr of analysable signal could be obtained for the PAT HSAT, as recommended by the AASM (Kapur et al., 2017).Participants with missing patient characteristics, such as age and gender data, were omitted from the analysis of population demo-graphic statistics. These participants were still included for the calculation of endpoint parameters"* (p4). All participants recruited were included in the analysis.
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	Seems likely that the different PSG devices would give the same results.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Mueller et al., 2022[29]

Device: WatchPAT 300 **Secondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	"participation was offered to all adult patients needing HST for suspected sleep-disordered breathing, in terms of loud irregular snoring, witnessed apneas, and daytime sleepiness." (p. 1674) "RP and PAT were compared in 56 patients receiving two nights of HST with either RP or PAT in a randomized fashion." (p. 1673) "we performed a randomized controlled study" (p. 1674)
Signalling question 2: Was a case-control design avoided?	Yes	Prospective, randomised controlled study
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	"Exclusion criteria included physical or mental restrictions interfering with independently installing the devices. In addition, patients with musculoskeletal diseases were not included."
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	Note that patients with physical or mental restrictions interfering with independently installing the devices and patients with musculoskeletal diseases were not included
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Unclear	Not reported
Signalling question 2: If a threshold was used, was it pre-specified?[c])	Yes	Standard thresholds were used "Mild OSA was defined as an AHI of 5 to <15, moderate OSA was defined as an AHI of 15 to <30, and severe OSA was defined as an AHI ≥ 30."
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** UNCLEAR	No comment.

B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** LOW	The study does not present results for the diagnostic performance of WatchPAT compared to respiratory polygraphy (e.g. accuracy). It does not describe respiratory polygraphy as a reference standard (the aim was to observe whether WatchPAT is superior to respiratory polygraphy)
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	If we are referring to the comparator (RP) rather than a reference standard
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	Not reported
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** UNCLEAR	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	If we are referring to the comparator (RP) rather than a reference standard
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	If we are referring to the index test in relation to comparator rather than a reference standard. “All participants received a RP- and a PAT-based HST device in a randomized order in two consecutive nights”. (p. 1674)
Signalling question 2: Did all patients receive a reference standard?	Yes	Five of the 61 enrolled participants were excluded from the analysis, as they only received one test. It therefore appears that the remaining 56 participants received the ‘reference standard’, RP.
Signalling question 3: Did patients receive the same reference standard?	Yes	If we are referring to the comparator rather than a reference standard. All participants appeared to receive RP (the ‘reference standard’).
Signalling question 4: Were all patients included in the analysis?	No	“Sixty-one patients were included in the study. Five patients were excluded from the analysis, as only one test was performed due to logistics and loss to follow-up.” (p. 1675)

		4 participants had missing RP data and 1 participant had missing WatchPAT data
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	No comment.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

NCT04031950 2019[32] (child)

Device: AcuPebble SA100 **Secondary Papers:** Three academic in confidence company reports[37 71 72] obtained as part of the company submission		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	[redacted]
Signalling question 2: Was a case-control design avoided?	Yes	No comment
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	Exclusion criteria are: *"• Patients whose parents/legal guardians/carers are not fluent in English, or who have special communication needs.* *• Known allergy to the adhesive dressing.* *• Patients with physical or mental impairments who would be too distressed with additional sensors on themselves* *• Patients with not enough space on the neck area to fit the sensor.* *• Clinical problem in the area in which the device will be attached, eg skin condition"*
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		

Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	Inclusion criteria are: *"• Children older than one year old (1 year to 18 years)* *• Children who have been referred to a sleep clinic due to suspicion of sleep apnoea"*
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	Assume automated scoring of index test (terms used are [redacted] and therefore results would be interpreted without knowledge of the results of the reference standard.
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	AASM 2020 paediatric rules
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index test conducted in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	[redacted]
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	[redacted]
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment

DOMAIN 4: FLOW AND TIMING	Assessment	Comments
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	[redacted]
Signalling question 2: Did all patients receive a reference standard?	Yes	[redacted]
Signalling question 3: Did patients receive the same reference standard?	Yes	No comment
Signalling question 4: Were all patients included in the analysis?	No	[redacted]
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	Reasons for being omitted from analysis were appropriate
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Pepin et al., 2020[26]

Device: Sunrise **Secondary papers:** Martinot et al., 2022[386]		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1:	Yes	"*376 consecutive adults with suspected OSA were enrolled"* (p4)

Was a consecutive or random sample of patients enrolled?		
Signalling question 2: Was a case-control design avoided?	Yes	*"376 consecutive adults with suspected OSA were enrolled*" (p4)
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	Consecutive adults with suspected OSA and no exclusion criteria given so presume there were no exclusions .Furthermore, "The final data set included all 376 patients recruited" (p4)
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	Automated scoring
Signalling question 2: If a threshold was used, was it pre-specified?[c]	No	"*Receiver operating characteristic (ROC) curve analysis was used to evaluate the overall clinical effectiveness of the new diagnostic tool via area under the curve (AUC), and a post hoc analysis was performed to optimize the cutoff points of Sr-RDI for diagnostic decisions compared with the criterion-standard cutoff values of obstructive PSG-RDI recommended in ICSD-3(5 events/h and 15 events/h*)"(.p4)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** HIGH	Post-hoc analysis was performed to optimize the cut-offs of Sr-RDI
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** HIGH	Post-hoc analysis was performed to optimize the cut-offs of Sr-RDI and index test conducted in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**

A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	"In-laboratory PSG was recorded with a digital acquisition system (Somnoscreen Plus; Somnomedics)" (p3)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	"In this prospective, diagnostic study of adult patients who were referred for a single overnight in-laboratory PSG, the PSG was used as the reference method and, with blinding, was compared with simultaneous MM recordings using the Sunrise system" (p2) "In a large, prospective cohort of patients with and without OSA, we evaluated the agreement between MM-derived Sr-RDI and blindly scored PSG-RDI" (p7)
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Simultaneous PSG and Sunrise use
Signalling question 2: Did all patients receive a reference standard?	Yes	"*The final data set included all 376 patients recruited"* (p4) so assume all received a reference standard
Signalling question 3: Did patients receive the same reference standard?	Yes	*"The final data set included all 376 patients recruited" (p4) and "In-laboratory PSG was recorded with a digital acquisition system (Somnoscreen Plus; Somnomedics)"* (p3) so assume so.
Signalling question 4: Were all patients included in the analysis?	Yes	*"The final data set included all 376 patients recruited*" (p4)
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	No comment
[a] with caveat that the device may be contraindicated in certain patient populations		

[b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard
[c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.

Pillar et al., 2020[31]

Device: WatchPAT 200 Unified **Secondary papers:** NCT02369705 2015[387]		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	No	*"Selectively recruited heart-failure patients in this study" (p. 389) as well as those without cardiac disorders.* *"The local staff at each participating center attempted recruiting participants in whom they estimated that there was a relatively high risk of having central apneas."* (p. 389) "Eighty-four (84) patients suspected of having SDB with selective bias toward recruiting patients with congestive heart failure (CHF) were recruited" (p. 389).
Signalling question 2: Was a case-control design avoided?	Yes	*"All participants underwent a full in-lab sleep study with simultaneous recording of PSG (which is in use in each center) and WP200U (Itamar-Medical, Caesarea, Israel)."* (p.389)
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Unclear	The study aimed to have a selection bias towards including heart-failure patients in their study but it is not clear, if in doing this, inappropriate exclusions were avoided. Listed criteria for exclusion (finger deformity and use of alpha blockers or short-acting nitrates less than 3 hours before the study) appear appropriate.
Judgment: Could the selection of patients have introduced bias?	**RISK:** HIGH	The intentional selection bias for patients with congestive heart failure may have introduced other unintentional biases.
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** UNCLEAR	Participants were included in the study if centre staff judged that they had "a relatively high risk of having central apneas" but participants would also have included people with suspected OSAHS.
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**

A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	*"The WP200U signals were analyzed for a WP-derived total apnea-hypopnea index (WP AHI) and central apnea-hypopnea index (AHIc) using its automatic software"* (p. 389)
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	Algorithm for the WP software had been determined by a prior study so was prespecified (p.392). The AHI and AHIc parameters from the WP200U device and the PSG were compared for thresholds of 10 and 15, for sensitivity analysis (p.392).
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index test was conducted in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"The reference used for comparison in this study was an FDA approved in-lab PSG from multiple manufacturers used in the eleven study sites."*(p.389)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	*"the PSG's manual scoring which was performed centrally by an external experienced PSG technologists blinded to the WP200U analysis."* (p.389)
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Tests conducted simultaneously

Signalling question 2: Did all patients receive a reference standard?	Yes	No comment.
Signalling question 3: Did patients receive the same reference standard?	Unclear	The study took place at 11 different centres, and PSG from 'multiple manufacturers" were used. So although all patients received PSG the paper does not state whether these different PSG devices would give identical results.
Signalling question 4: Were all patients included in the analysis?	Yes	Results for all 84 participants included.
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	No comment.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Storey et al., 2022[28]

Device: WatchPAT ONE **Secondary papers:** none		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"Patients were selected at random within the period from 01/01/2019 to 01/06/2022"*
Signalling question 2: Was a case-control design avoided?	Yes	*"Patients were selected at random within the period from 01/01/2019 to 01/06/2022"*
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Unclear	Inclusion and exclusion criteria were not reported
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	States "*Patients were selected at random"*
B. Concerns regarding applicability		

Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** UNCLEAR	States "*Patients were selected at random within the period from 01/01/2019 to 01/06/2022. Patients were referred from different consultants specialising within 5 disciplines. These included Sleep, ENT, Insomnia, Dental and Respiratory*" therefore unclear if these patients were all suspected having OSA or other sleep disorders
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Unclear	This was an audit to compare DNA rate, cost efficiency, staff time and patient time
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Unclear	This was an audit to compare DNA rate, cost efficiency, staff time and patient time
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** UNCLEAR	Details on conduct and interpretation of index test report were not reported
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Details on conduct and interpretation of index test report were not reported
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	NOX T3 is listed in the NICE scope (section 4 Comparators page 21) as one of the home respiratory polygraph technologies currently used across the NHS
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	This was an audit to compare DNA rate, cost efficiency, staff time and patient time
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** UNCLEAR	This was an audit to compare DNA rate, cost efficiency, staff time and patient time
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** UNCLEAR	Details on interpretation of results of reference standard were not reported
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**

A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Unclear	This was an audit to compare DNA rate, cost efficiency, staff time and patient time
Signalling question 2: Did all patients receive a reference standard?	Unclear	This was an audit to compare DNA rate, cost efficiency, staff time and patient time
Signalling question 3: Did patients receive the same reference standard?	Unclear	This was an audit to compare DNA rate, cost efficiency, staff time and patient time
Signalling question 4: Were all patients included in the analysis?	Unclear	Not reported if data from all 600 patients were included in the analyses
Judgment: Could the patient flow have introduced bias?	**RISK:** UNCLEAR	Not reported if data from all 600 patients were included in the analyses
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias		

Tauman et al., 2020[30]

Device: WatchPAT 200 Unified **Secondary papers:** NCT02369705 2015[387]		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Unclear	Participants were recruited as part of a larger study and had to have atrial fibrillation. Paper does not state if consecutive patients with atrial fibrillation were enrolled.
Signalling question 2: Was a case-control design avoided?	Yes	*"Full in-lab PSG and WP200U (Itamar Medical, Caesarea, Israel) were conducted simultaneously*." (p.1116)
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	Study is focussing on patients with atrial fibrillation (a population excluded from other studies using the WP device). Exclusions listed in the paper ("finger deformity that precluded adequate sensor

		appliance, use of short acting nitrates for less than 3 hours before the study and use of alpha blockers"; p. 1116) appear appropriate.
Judgment: Could the selection of patients have introduced bias?	**RISK:** UNCLEAR	No information is provided about how the participants were selected to take part in this study from the larger study.
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	But note that study focus is patients with atrial fibrillation who would be a subgroup of the overall population for the review question.
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	*"The WP signals were analyzed using its automatic software"* (p.1116)
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	Paper states that the AHI was obtained from the WP device but does not indicate if the device has any in-built thresholds defined. For polysomnography Moderate-severe sleep apnoea was defined as AHI ≥15/hour and paper states "The AHI obtained from the WP device and from PSG were compared as descriptive statistics and compared using Pearson correlation and Bland-Altman. Sensitivity and specificity of the WP in diagnosis moderate-severe sleep apnea were compared to the PSG, ROC and AUC were used." It therefore appears that the same threshold as for PSG was used.
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index test conducted in a sleep laboratory not home setting. Again noting that the population is limited to those with atrial fibrillation.
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	"*Moderate-severe sleep apnea was defined as AHI ≥15/hour"* (p.1116) This is a standard threshold and the paper is grouping moderate and severe together (i.e. lower boundary AHI ≥15/hour and no upper boundary given).
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	*"External experienced PSG technologist who was blinded to the automatic scoring analysis of the WP manually scored the PSG"* (p.1116)

Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment.
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	Tests conducted simultaneously.
Signalling question 2: Did all patients receive a reference standard?	Yes	No comment.
Signalling question 3: Did patients receive the same reference standard?	Unclear	Eleven centres were involved in this study and the "In-lab PSG was FDA-approved from multiple manufacturers used in the eleven study sites, all according to the site regular clinical practice and compliant with accepted standards" (p.1116) Although all patients received PSG the paper does not state whether these different PSG devices would give identical results.
Signalling question 4: Were all patients included in the analysis?	Yes	States 101 recruited and 101 were included in the analysis.
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	No comment.
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Van Pee et al., 2022[24]

Device: NightOwl **Secondary papers:** NCT04191668 2019[388]

DOMAIN 1: PATIENT SELECTION	Assessment	Comments
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"One hundred sixty-seven participants suspected of having OSA were consecutively included in a cohort"* (p.2)
Signalling question 2: Was a case-control design avoided?	Yes	*"During the setup of PSG, the PAT HSAT (NightOwl, reusable version, software version 1.202.1) was attached to the middle finger of the hand to which the pulse oximeter of PSG was applied."* (p.2)
Signalling question 3: Did the study avoid inappropriate exclusions?[a]	Yes	Consecutive sample and "*Underaged or mentally impaired participants were excluded from participation in the study*" – these seem appropriate exclusions for the purposes of the study.
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?[b]	Yes	Used the reusable NightOwl device. Although not explicitly stated in the paper the device records sleep data and provides the AHI score.
Signalling question 2: If a threshold was used, was it pre-specified?[c]	Yes	Although not explicitly stated the NightOwl device will have a pre-specified threshold that the algorithm uses.
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment.
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index text performed in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	*"Qualified lab technicians at each participating study center were responsible for setting up the equipment and capturing PSG data*." (p. 2) "For the European center, the Alice 6 PSG (Philips

		Respironics, USA) was used, whereas a Cadwell Easy PSG (Cadwell, USA) was applied in the US centers." (p. 2)
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	*"All PSG data were double-scored by two independent centers which were blinded from one-another's analysis."* (p.2) but it is not stated whether they were also blind to the results of the index test.
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	Although it is not stated that the PSG scorers were blind to the results of the index test, the double scoring of PSG by the independent centers blind to one-another's analysis should mean the risk of bias is low.
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 4: FLOW AND TIMING	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	The two tests were carried out concurrently.
Signalling question 2: Did all patients receive a reference standard?	Yes	*"Routine PSG was performed in all study participants"* (p. 2).
Signalling question 3: Did patients receive the same reference standard?	Unclear	Although all patients received PSG two different devices were used "For the European center, the Alice 6 PSG (Philips Respironics, USA) was used, whereas a Cadwell Easy PSG (Cadwell, USA) was applied in the US centers." The paper does not state whether these different PSG devices would give identical results.
Signalling question 4: Were all patients included in the analysis?	No	Patient flow provided in Fig 2 of journal article. Of 228 recruited, acquisition failed for 48 (reasons given: PSG flow failed: 4; PSG SpO2 issue: 7; PSG admin error:9; Defect PAT HSAT:3; Detached PAT HSAT: 2; PAT HSAT :2; PAT HSAT not included:7; PAT HSAT app issue:16)) and 13 of 180 successful acquisitions were deemed to be "technically inadequate" (less than 4 hours of interpretable data).
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	Reasons for exclusion of patients are due to absent or inadequate data. Seems likely that the different PSG devices would give the same results.
[a] with caveat that the device may be contraindicated in certain patient populations		

[b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard
[c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.

Sanchez Gomez et al., 2024

Device: AcuPebble SA100 **Secondary papers:** Phase 1 Virgen Macarena trial 2023[15]		
DOMAIN 1: PATIENT SELECTION	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	*"A total of 80 consecutive participants meeting the eligibility criteria were consecutively enrolled in this prospective study"*
Signalling question 2: Was a case-control design avoided?	Yes	No comment
Signalling question 3: Did the study avoid inappropriate exclusions? *(Note: Remember that the device may be contraindicated in certain patient populations)*	Yes	Inclusion and exclusion criteria appear appropriate
Judgment: Could the selection of patients have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** LOW	No comment
DOMAIN 2: INDEX TEST(S)	**Assessment**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard? *(Note: Consider whether the index test was automatically scored by the software only, and could*	Yes	Scoring was automated

therefore be considered independent of the results of the reference standard)		
Signalling question 2: If a threshold was used, was it pre-specified? *(Note: for AHI and ODI, the following thresholds are standard (NICE scope, EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per. If these specific thresholds are used but NOT prespecified we will not consider this an increase risk of bias)*	Yes	Standard thresholds used with no post-hoc adjustment (Source: DAP70 EAG request to AcuPebble November 2023_responses [AIC].docx)
Judgment: Could the conduct or interpretation of the index test have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN:** UNCLEAR	Index test conducted in a sleep laboratory not home setting
DOMAIN 3: REFERENCE STANDARD		
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	In-hospital PSG
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	*"Blind manual scoring was carried out by a trained sleep physician following the 2017 American Academy of Sleep Medicine (AASM) guidelines"* *"Clinicians did this blindly and at no point had any information about the automatic diagnosis produced by AcuPebble SA100."*
Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** LOW	No comment
B. Concerns regarding applicability		
Judgment: Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN:** LOW	No comment

DOMAIN 4: FLOW AND TIMING		
A. Risk of Bias		
Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?	Yes	AcuPebble and PSG were used simultaneously
Signalling question 2: Did all patients receive a reference standard?	Yes	PSG using Philips Sleepware G3 version 2.8.78
Signalling question 3: Did patients receive the same reference standard?	Yes	No comment
Signalling question 4: Were all patients included in the analysis?	No	80 participants recruited, but 63 included in analysis. *"Out of the total, 63 studies could be used for comparison. The 17 remaining patients could not be included due to 1 third-party equipment malfunction only observed post-recruitment, 6 short duration of reference PSG recording, 2 due to lack of output from the reference recording due to software-related failure in the computer running the reference standard system (of these one of the patients had also been annotated in the CRF not to have slept at all), 3 due to human error in the form of forgetting to start the AcuPebble SA100 test, and 5 due to the algorithms of AcuPebble SA100 not producing any diagnosis, but rather an output determining there was lack of confidence from the signals to produce this. (This is a safety-by-design feature of the system to minimize the risk of a wrong diagnosis, which can trigger for a variety of reasons, such as signals not representative of intended use as per the user manual, signals indicative of instructions of use not followed, etc.)."*
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	No comment
[a] with caveat that the device may be contraindicated in certain patient populations [b] with caveat that if the index test was automatically scored by the software only, it could be considered independent of the results of the reference standard [c] with caveat that for AHI and ODI, the following thresholds are standard (as per NICE scope, and EAG protocol): Mild OSAHS: 5 or more to less than 15 events per hour; Moderate OSAHS: 15 or more to less than 30 events per hour; Severe OSAHS: 30 or more events per hour. If these specific thresholds are used but NOT prespecified this is not considered an increased risk of bias.		

Appendix 6 Systematic review of cost-effectiveness studies

Table 69 Inclusion and exclusion criteria for the CEA systematic review

	Included	Excluded
Population	People suspected of OSAHS suitable for home testing (adults and children)	People already diagnosed with OSAHS
Intervention	Acupebble SA100 (Acurable) NightOwl (ResMed) Sunrise (Sunrise) WatchPAT 300 WatchPAT ONE Brizzy	Any intervention not listed
Comparison	Home respiratory polygraphy Home oximetry	Hospital respiratory polygraphy or PSG
Outcome	Cases detected Life-years QALYs Time to diagnosis/treatment	Only resource use or costs
Study design	CEA CUA CBA	Cost studies Non-comparative BIA CMA
Publication type	Full-text Abstracts (prioritise those within 3 years) English-language only	

Abbreviations: BIA budget impact assessment; CBA cost-benefit analysis; CEA cost-effectiveness analysis; CMA cost-minimisation analysis; CUA cost-utility analysis; PSG polysomnography; QALY quality-adjusted life-years

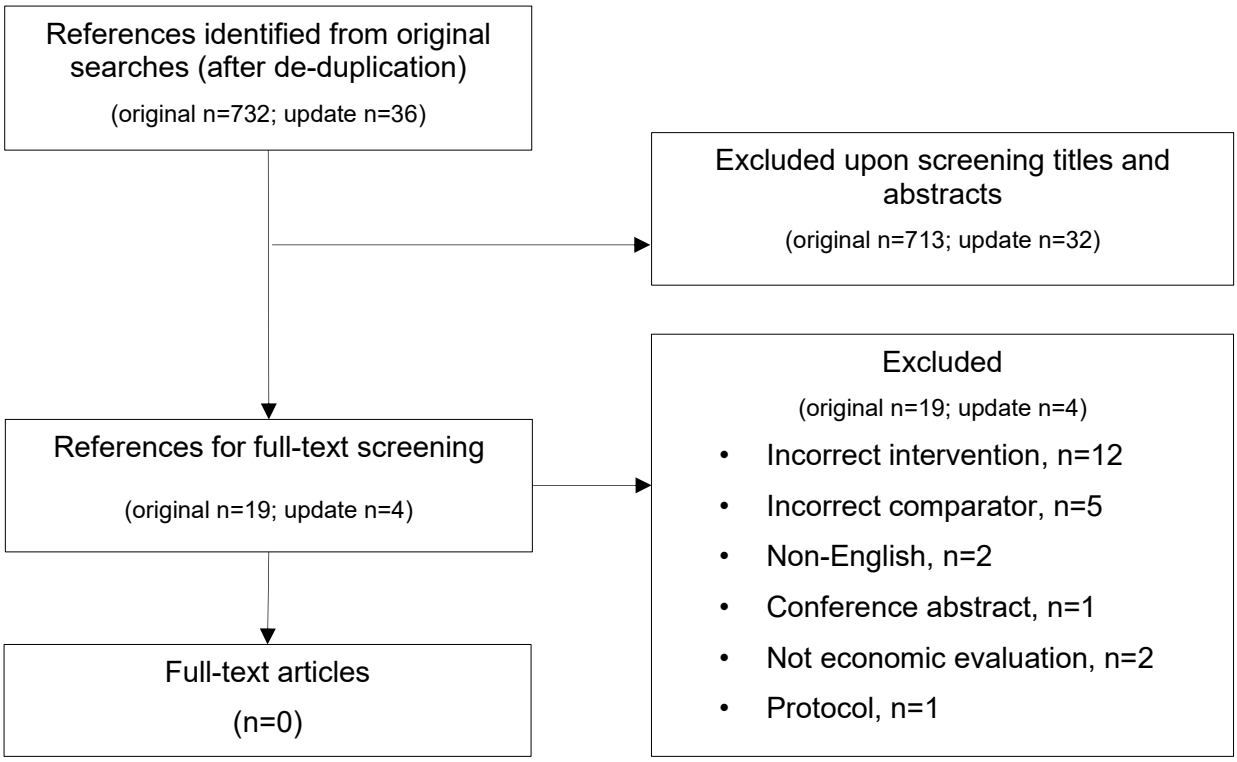


Figure 10 Flow chart for the identification of economic studies

Table 70 Full text publications excluded from systematic review of cost effectiveness

Publication	Reason for exclusion
Mills et al 2023 [389]	Not an economic evaluation
NCT 2023 [295]	Protocol
Pei 2023 [304]	Not a named intervention
Ssegonja 2023 [390]	Not a named intervention
Jurado Gamez 2007 [217]	Not in English
Perleth 2003 [391]	Not in English
Ayas 2010 [392]	Not a named intervention
Ayas 2021 [393]	Not a named intervention
Barriuso 2020 [394]	Not a named intervention
Bravata 2018 [395]	Not a named intervention
Deutsch 2006 [396]	Not a named intervention
Fletcher 2000 [397]	Not a named intervention
Pietzsch 2011 [398]	Not a named intervention
Veloso 2021 [399]	Not a named intervention
Natsky 2021 [400]	Not a named intervention
Natsky 2022 [401]	Not a named intervention
Duran-Cantolla 2017	Incorrect comparator
Duran-Cantolla 2017	Duplicate

Publication	Reason for exclusion
Masa 2011 [402]	Incorrect comparator
NCT 2012 [284]	Incorrect comparator
NCT 2016 [403]	Incorrect comparator
Atwood 2014 [404]	Conference abstract
Tondo 2021 [340]	Not an economic evaluation

Table 71 Description of economic studies of interest not meeting inclusion criteria

Author, year	Study summary
Phua 2021 et al [49]	An evaluation of an earlier version of WatchPAT. Although the version of WatchPAT evaluated is not stated explicitly in the article, the sleep studies were conducted between 2014 and 2017, and the authors refer to accuracy data from WatchPAT 200. We therefore assumed that this is the version of WatchPAT evaluated in this study. Phua and colleagues compared the costs associated with the use of WatchPAT and PSG for diagnostic sleep studies in Singapore. Estimates of the test costs, the number of repeat tests, the prevalence and severity of OSA, and the time to treatment were obtained from a retrospective review of patients from a single clinic, over a period of 3.5 years. Individuals prescribed WatchPAT had to have a high probability of moderate to severe OSA, and no contraindications. Patients with contraindications for WatchPAT were offered in-laboratory PSG. PSG was assumed to have perfect sensitivity and specificity, with accuracy for WatchPAT taken from an earlier study for WatchPAT 200. We did not use data from this study to inform the EAG model due to the setting in Singapore, the number of assumptions made, the lack of detailed data, and the fact the analysis corresponded to an earlier version of WatchPAT.
Di Pumpo et al 2022 [47]	Report a cost-minimisation analysis for use of WatchPAT 200 at home for diagnosis of OSAHS, comparing two scenarios: 1) assuming that patients would need to attend hospital appointments and collect and return equipment; and 2) assuming online/telephone appointments and delivery and return of equipment by post (telemedicine). The analysis is based in Italy, and only estimates costs associated with OSA diagnosis, including the costs of “negative studies” which were assumed to incur further hospital diagnostics. Three perspectives were assumed: patient, hospital and society. From the patient and the societal perspective, the telemedicine strategy is associated with the lowest cost: €289 compared

	to €456 (patient) and €517 compared to €636 (societal perspective). However, from the hospital perspective, the telemedicine strategy is more costly strategy €228 compared to €180. This is due to greater equipment costs for the telemedicine approach, but it is unclear from the article what exactly is included in these costs. Although this study concerns a named device (albeit an earlier version), due to the lack of detail on the estimates of resource use used to obtained relevant costs, and its setting in Italy, we did not use any data from this study to inform our model.
Geessinck et al 2018 [48]	A cost-utility analysis of a screening tool for OSAHS (DiagnOSAS) in men aged 50 years based in the Netherlands. It uses a Markov model with 5 and 10 year time horizons, capturing cardiovascular disease (CVD) and road traffic accident (RTA) risks associated with un- or under-treated OSAHS. There is no mention of failure rates for the diagnostic tests, and although the model diagram indicates that false negatives return to the diagnostic state, no details are provided on this. Many of the parameter sources for this model were noted in previous evaluations, including the recent NICE guidelines (NG202).[2]
McMillan et al 2015 [51]	A UK-based RCT that evaluated the impact of CPAP and best supportive care compared to best supportive care in 278 individuals newly diagnosed with OSAHS age ≥ 65 years. A CUA was conducted based on the RCT, with an NHS perspective and a one year time horizon. The authors report that both EQ-5D and SF-6D utilities obtained from participants were used to inform the CUA. Resource use costs were informed by patient diaries and included appointments with a GP or nurse, calls to NHS Direct, costs associated with ambulance services, attendance at Accident and Emergency (including admissions), as well as outpatient clinics. The cost of one year of CPAP usage was also included. McMillan and colleagues also used a Markov model to evaluate the longer-term impacts of CPAP treatment. The base case analysis only included two health states: treated OSAHS and death. Scenario analyses also included states reflecting (acute- and post-) coronary heart disease and stroke events, with the assumption that CPAP treatment led to reductions in the risk of cardiovascular events. The Framingham risk prediction model was used to estimate transition probabilities for cardiovascular events. Information on model structure

	and parameter values from McMillan et al 2015 are used in a number of scenario analyses for the EAG economic model for adults, including the assumption that treatment has no impact on cardiovascular or RTA risk, and evidence on adherence to CPAP.
Guest et al 2008 [50]	Report a Markov model to evaluate the cost-utility of CPAP compared to no treatment in the UK. As well as assuming impacts on cardiovascular risk, the Guest model also assumes treatment impacts on RTAs. Unlike the NG202 Markov model, the Guest model allows for the probability that an individual can experience more than one cardiovascular event. Although this study is relevant in terms of being based in the UK and evaluating a population with OSAHS, due to its age, it did not contribute any evidence to the EAG model.

Appendix 7 Systematic review of health-related quality of life studies

Table 72 Inclusion criteria for the HQRoL systematic review

	Included
Population	Adults and children - People suspected of OSAHS - People diagnosed with OSAHS (Adults who are untreated, treated with CPAP, MAD, conservative management; Children who are untreated or receiving any treatment)
Outcome	Prioritise EQ-5D - Adults: EQ-5D (-3L and -5L), SF-36, SF-12, SF-6D, HUI-1, HUI-2, HUI-3, 15D, QWB (all can be mapped to EQ-5D) - Children: generic measures to value utility in children, including HUI-2 specific to children, EQ-5D-Y, CHU-9D
Value-set	Any country Prioritise those from UK
Study design	Primary studies Systematic review
Publication type	Full-text research studies only English-language only
Abbreviations: CPAP continuous positive airway pressure; MAD mandibular advancement device; EQ-5D Euro-Qol 5 dimensions; EQ-5D-Y Euro-Qol 5 dimensions-youth; SF Short-Form; HUI Health Utilities Index; QWB Quality of Wellbeing; CHU-9D Child Health Utility Index – 9 Dimensions	

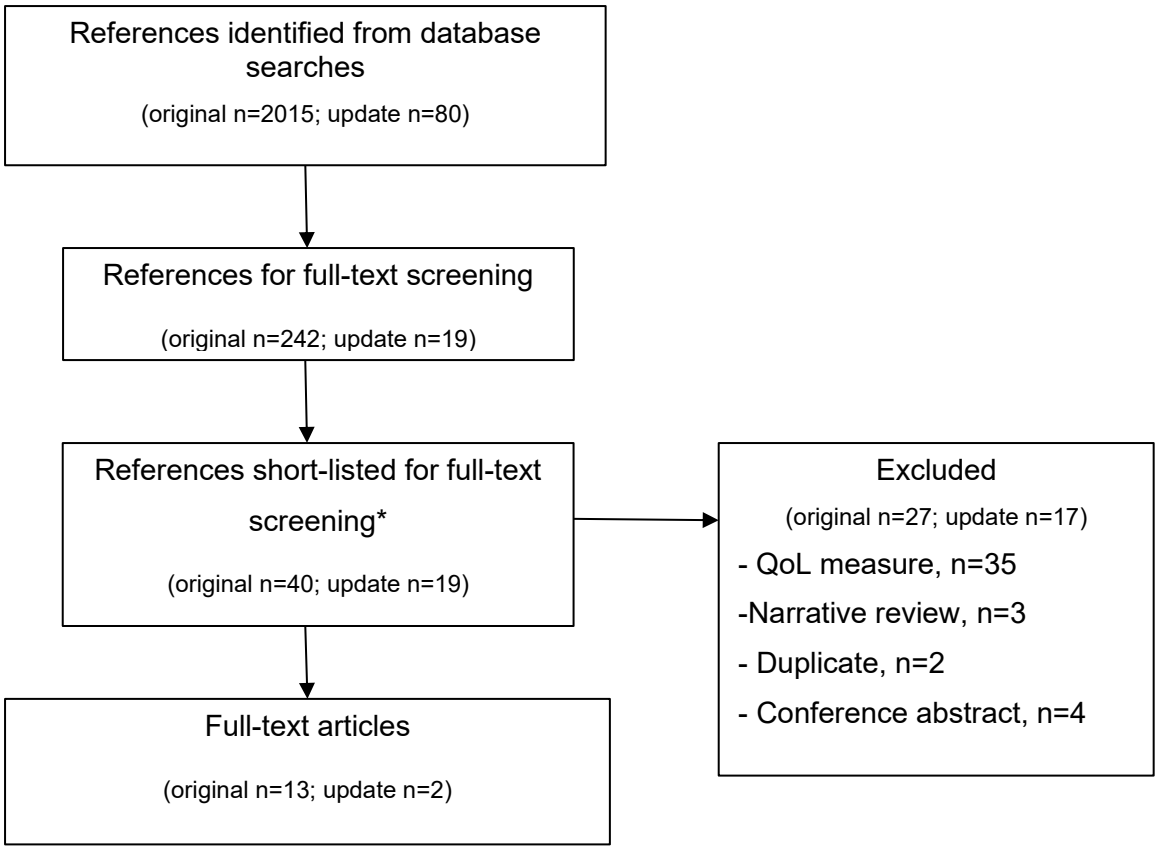


Figure 11 Flow chart for the identification of HRQoL studies

* All studies in children were included in full-text screening. For adult studies, only those reporting EQ-5D, SF-6D, HUI, QWB or 15D and/or set in the UK were included in full-text screening.

Table 73 Full text publications excluded from systematic review of HRQoL

Study	Reasons for exclusion
Krishnan 2022 [405]	No utility outcomes
Wireklint 2012 [406]	No utility outcomes
Jennum 2021 [407]	No utility outcomes
McMillan 2014 [408]	Duplicate
De Vries 2019 [409]	No utility outcomes
Tarraubella 2018 [410]	Duplicate
Batool-Anwar 2018 [411]	Conference abstract
Amin 2021 [412]	No utility outcomes
Betavani 2022 [413]	No utility outcomes
Carvalho 2016 [414]	No utility outcomes
Clements 2021 [415]	No utility outcomes

Study	Reasons for exclusion
Decuzzi 2022 [416]	No utility outcomes
End 2022 [417]	No utility outcomes
Hamid 2020 [418]	No utility outcomes
Idris 2018 [419]	No utility outcomes
Jackman 2013 [420]	No utility outcomes
Kaditis 2022 [421]	Narrative review
Kanona 2015 [422]	No utility outcomes
Kao 2017 [423]	Not a systematic review
Marcus 2013 [113]	No utility outcomes
Nazir Othman 2016 [424]	No utility outcomes
Roy 2023 [425]	No utility outcomes
Saju 2022 [426]	No utility outcomes
Todd 2017 [427]	No utility outcomes
Torretta 2017 [428]	Not a systematic review
Venekamp 2015 [123]	No utility outcomes
Zhang 2017 [429]	No utility outcomes
Bhushan et al 2023 [430]	No utility outcomes
Bironneau et al 2023 [431]	No utility outcomes
Chen et al 2023 [432]	No utility outcomes
De Weerdt 2023 [433]	No utility outcomes
Dundervill 2023 [434]	No utility outcomes
Horne et al 2023 [435]	No utility outcomes
Jensen 2023 [436]	Conference abstract
Jensen 2023 [437]	No utility outcomes
Lee et al 2023 [438]	No utility outcomes
Luz 2023 [439]	No utility outcomes
Samota 2023 [440]	No utility outcomes
Schwartz 2023 [441]	Conference abstract
Shen 2023 [442]	No utility outcomes
Sudarsan 2023 [443]	No utility outcomes

Study	Reasons for exclusion
Vilarrasa 2023 [444]	Conference abstract
Vilma Fidelis 2023 [445]	No utility outcomes
Xiao 2023 [446]	No utility outcomes

Appendix 8 Resource use and cost estimates

Cost of devices

Table 74 Cost components for a failed sleep study

	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Sleep study	£1	£0	£0	£62	£0	£0
Sleep physiologist data review time	£0[a] £4.42[b]	£4.42	£4.42	£4.42	£4.42	£4.42
Sleep physiologist 10 minute call	£0 [a] £8.83 [b]	£8.83	£8.83	£8.83	£8.83	£8.83
Preparation of device	NA[c]	£2.83	NA[c]	£0	£5.67	£2.83
Organisation of device for repeat	NA[c]	£5.67	NA[c]	£5.67	£5.67	£5.67
Postage	NA[c]	£5.98	NA[c]	£2.99	£5.98	£2.99
Total cost[d]						
Device still with patient	£1 [a] £14.25 [b]	NA	£13.25	NA	NA	NA
Total cost[d] – second device required						
Posted	NA	£27.73	NA	£83.91	£30.56	£24.74
Collected (and returned) in person	NA	£21.75	NA	£80.92	£24.58	£21.75

[a]if function automated; [b]if function not automated; [c] Device still with patient; [d] **does not include the costs of 20 minutes to review data and prepare report (as this is accounted for in the initial sleep study)**

Table 75 Cost components for a second sleep study due to misdiagnosis

	AcuPebble SA100	Brizzy	NightOwl	Sunrise	WatchPAT 300	WatchPAT ONE
Sleep study	£44	£0	£0	£62	£50	£80
Consumables	£1	£0	£0	£0	£0	£0
Preparation of device	£5.67	£2.83	£0	£0	£5.67	£2.83
Organisation of device for repeat	£5.67	£5.67	£5.67	£5.67	£5.67	£5.67
Postage	£5.98	£5.98	£0	£2.99	£5.98	£2.99
Data and report review; 20 minutes	£17.67	£17.67	£17.67	£17.67	£17.67	£17.67
Total cost						
Posted; 20 minutes to review	£79.98	£32.15	£23.33	£88.32	£84.98	£109.16
In person; 20 minutes to review	£74.00	£26.17	NA	£85.33	£79.00	£106.17

Table 76 First year CPAP costing assumptions

Type of CPAP	Cost components			
	Annuitised device cost	**Retitration**	**Telemonitoring**	**Annuitised consumable costs**
CPAP with autotitration	Average cost (excluding VAT) for: - ResMed S9 Escape - ResMed AirSense10 Elite - ResMed AirSense Elite - DeVilbiss Healthcare Standard sleepcube - Philipps Respironics SystemOne Pro - Philipps Respironics Dreamstation Pro 7 year device lifetime 3.5% discount rate	Average cost (excluding VAT) for: - ResMed AirSense AutoSet - Philipps Respironics Dreamstation Auto CPAP - Philipps Respironics SystemOne Auto 7 year device lifetime 3.5% discount rate 18% of patients require retitration	NA	Mask (1 year lifetime) Humidifier and chamber (required by 40% of patients; humidifier 3.5 years lifetime) Hose Pollen filters (2 per year) Ultrafine filters (12 per year)
CPAP with telemonitoring		20 minutes of band 6 physiologist for autotitration with telemonitoring As above	Annual cost of telemonitoring access (over 5 years)	
CPAP with telemonitoring (year 1 only)			One-year cost of telemonitoring access	
Auto-CPAP only	Average cost (excluding VAT) for: - ResMed AirSense AutoSet	NA	NA	

Type of CPAP	Cost components			
	Annuitised device cost	**Retitration**	**Telemonitoring**	**Annuitised consumable costs**
Auto-CPAP with telemonitoring	- Philipps Respironics Dreamstation Auto CPAP - Philipps Respironics SystemOne Auto	NA	Annual cost of telemonitoring access (over 5 years)	
Abbreviations: APAP auto-continuous positive airway pressure; CPAP continuous positive airway pressure; NA not applicable; VAT value added tax				

Appendix 9 Additional cost-effectiveness results

9a Scenario analyses

Table 77 INMB estimates at £20,000 per QALY gained for each novel device vs respiratory polygraphy and oximetry

	Compared to respiratory polygraphy						Compared to oximetry					
Setting of evaluation	Clinic						Clinic					
	AcuPebble	Brizzy	NightOwl	Sunrise	Watch PAT 300	Watch PAT ONE	AcuPebble	Brizzy	NightOwl	SunRise	Watch PAT 300	Watch PAT ONE
Base case	£141	-£15	-£7	£127	-£13	-£36	£1,165	£1,009	£1,017	£1,152	£1,012	£989
Lower risk cohort	£153	£0	-£6	£129	-£26	-£49	£1,134	£980	£975	£1,110	£955	£931
Higher risk cohort	£131	-£24	-£6	£126	£1	-£23	£1,173	£1,018	£1,036	£1,168	£1,043	£1,020
20% correlation	£134	-£18	-£10	£124	-£14	-£37	£1,403	£1,250	£1,258	£1,393	£1,255	£1,232
40% correlation	£127	-£22	-£14	£120	-£15	-£38	£1,663	£1,515	£1,522	£1,657	£1,521	£1,498
Oximetry devices posted	NA	NA	NA	NA	NA	NA	£1,166	£1,010	£1,018	£1,152	£1,012	£989
Higher volume of sleep studies	£145	NA	NA	NA	-£7	NA	£1,169	NA	NA	NA	£1,017	NA
Lower volume of sleep studies	£135	NA	NA	NA	-£18	NA	£1,160	NA	NA	NA	£1,006	NA
Company stated time for data review	£150	-£10	£5	£137	-£3	-£26	£1,175	£1,014	£1,029	£1,161	£1,021	£998

	Compared to respiratory polygraphy						Compared to oximetry					
Setting of evaluation	Clinic						Clinic					
	AcuPebble	**Brizzy**	**NightOwl**	**Sunrise**	**Watch PAT 300**	**Watch PAT ONE**	**AcuPebble**	**Brizzy**	**NightOwl**	**SunRise**	**Watch PAT 300**	**Watch PAT ONE**
Base case	£141	-£15	-£7	£127	-£13	-£36	£1,165	£1,009	£1,017	£1,152	£1,012	£989
Lower cost for home RP	£116	-£40	-£32	£102	-£38	-£61	£1,160	£1,004	£1,013	£1,147	£1,007	£984
List price for Sunrise	NA	NA	NA	£113	NA	NA	NA	NA	NA	£1,138	NA	NA
RTA costs include police costs	£137	-£21	-£9	£126	-£10	-£33	£1,186	£1,028	£1,040	£1,175	£1,039	£1,016
RTAs involve >1 casualty	£138	-£20	-£8	£127	-£11	-£34	£1,180	£1,022	£1,034	£1,168	£1,031	£1,008
Conservative management is cost of letter	£129	-£26	-£11	£117	£2	-£22	£1,180	£1,025	£1,041	£1,168	£1,053	£1,030
CPAP with autotitration	£135	-£24	-£10	£126	-£9	-£32	£1,196	£1,036	£1,051	£1,187	£1,052	£1,029
CPAP with telemonitoring	£144	-£10	-£6	£128	-£15	-£38	£1,149	£995	£1,000	£1,133	£991	£968
CPAP with telemonitoring (1st year only)	£136	-£24	-£10	£126	-£11	-£34	£1,196	£1,037	£1,051	£1,187	£1,050	£1,026
Auto-PAP with telemonitoring	£149	-£1	-£3	£129	-£19	-£42	£1,118	£968	£966	£1,098	£950	£927
Lower CPAP cost	£130	-£33	-£11	£123	**£1**	-£22	£1,213	£1,050	£1,071	£1,206	£1,084	£1,061

	Compared to respiratory polygraphy						Compared to oximetry					
Setting of evaluation	Clinic						Clinic					
	AcuPebble	**Brizzy**	**NightOwl**	**Sunrise**	**Watch PAT 300**	**Watch PAT ONE**	**AcuPebble**	**Brizzy**	**NightOwl**	**SunRise**	**Watch PAT 300**	**Watch PAT ONE**
Base case	£141	-£15	-£7	£127	-£13	-£36	£1,165	£1,009	£1,017	£1,152	£1,012	£989
No treatment impacts on CV events	£140	-£11	-£3	£128	-£7	-£30	£1,131	£980	£989	£1,119	£985	£962
Higher drop-out for oral devices	£141	-£15	-£7	£127	-£13	-£36	£1,166	£1,010	£1,019	£1,153	£1,013	£990
PSG as third sleep study	£141	-£15	-£7	£127	-£13	-£36	£1,166	£1,010	£1,018	£1,153	£1,013	£990
Devani 2021 accuracy data for AcuPebble	£138	NA	NA	NA	NA	NA	£1,162	NA	NA	NA	NA	NA
Novel devices collected in person	£147	-£9	-£4	£131	-£6	-£24	£1,172	£1,016	£1,020	£1,155	£1,019	£1,001
Kelly 2022 accuracy data for Sunrise	NA	NA	NA	£451	NA	NA	NA	NA	NA	£1,475	NA	NA
Disable functionality of app for AcuPebble	£141	NA	NA	NA	NA	NA	£1,165	NA	NA	NA	NA	NA
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Failure rate for Sunrise	NA	NA	NA	[redacted]	NA	NA	NA	NA	NA	[redacted]	NA	NA

	Compared to respiratory polygraphy						Compared to oximetry					
Setting of evaluation	Clinic						Clinic					
	AcuPebble	**Brizzy**	**NightOwl**	**Sunrise**	**Watch PAT 300**	**Watch PAT ONE**	**AcuPebble**	**Brizzy**	**NightOwl**	**SunRise**	**Watch PAT 300**	**Watch PAT ONE**
Base case	£141	-£15	-£7	£127	-£13	-£36	£1,165	£1,009	£1,017	£1,152	£1,012	£989
[redacted]	NA	NA	NA	[redacted]	NA	NA	NA	NA	NA	[redacted]	NA	NA
Abbreviations: FN false negative; CPAP continuous positive airway pressure; CV cardiovascular; ICER incremental cost-effectiveness ratio; Incr incremental; INMB incremental net monetary benefit; PSG polysomnography; QALYs quality-adjusted life-years; RTA road traffic accident												

9b One-way sensitivity analyses

Table 78 Parameter values used in the base case, probabilistic analysis and one-way sensitivity analyses

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
Prevalence							
OSAHS in cohort	0.82	NG202	Beta	Alpha 11.93 Beta 2.64	0.63	1.00	NG202 95% CI limits
Moderate-severe in OSAHS cohort	0.68	NG202	Calculated as 1 - prevalence of mild OSAHS				
Severe in moderate-severe cohort	0.60	NG202	Beta	Alpha 31.59 Beta 21.01	0.47	0.73	NG202 95% CI limits
Mild in OSAHS cohort	0.32	NG202	Beta	Alpha 4.39 Beta 9.16	0.08	0.56	NG202 95% CI limits
Diagnostic accuracy							
Oximetry: low sensitivity	0.51	NG202	Drawn from WinBUGS meta-analysis output		0.08	0.93	NG202 diagnostic evidence report 95% CI limits
Oximetry: low specificity	0.89				0.15	1.00	
Oximetry: high sensitivity	0.36				0.13	0.65	
Oximetry: high specificity	0.99				0.95	1.00	
Respiratory polygraphy: low sensitivity	0.95	Xu 2017	Dirichlet on re-created 2x2 table		0.87	0.99	Xu 2017 95%CI limits

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
Respiratory polygraphy: low specificity	0.69				0.39	0.91	
Respiratory polygraphy: high sensitivity	0.93				0.81	0.99	
Respiratory polygraphy: high specificity	0.86				0.70	0.95	
AcuPebble: low sensitivity	[redacted]	[15] Sanchez Gomez (2024) [20]	Dirichlet on 2x2 table		[redacted]	[redacted]	[15] Sanchez Gomez (2024) [20] 95% CI limits
table AcuPebble: low specificity	[redacted]				[redacted]	[redacted]	
AcuPebble: high sensitivity	0.93				0.77	0.99	
AcuPebble: high specificity	0.97				0.85	1.00	
Brizzy: low sensitivity	0.93	Martinot 2017	Dirichlet on re-created 2x2 table		0.87	0.97	Martinot 2017 95%CI limits
Brizzy: low specificity	1.00				0.51	1.00	
Brizzy: high sensitivity	0.89				0.80	0.94	
Brizzy: high specificity	1.00				0.83	1.00	
NightOwl: low sensitivity	0.93	Lyne 2023	Dirichlet on 2x2 table		0.85	0.98	Lyne 2023 95%CI limits calculated from 4x4 contingency table
NightOwl: low specificity	0.77				0.55	0.92	
NightOwl: high sensitivity	0.89				0.76	0.96	
NightOwl: high specificity	0.82				0.68	0.91	
Sunrise: low sensitivity	0.91	Pepin 2020	Dirichlet on re-created 2x2 contingency tables		0.89	0.92	Pepin 2020 95%CI limits
Sunrise: low specificity	0.94				0.91	0.97	

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
Sunrise: high sensitivity	0.92				0.90	0.94	
Sunrise: high specificity	0.84				0.81	0.87	
WatchPAT 300: low sensitivity	0.96	Tauman 2020 for WatchPAT 200U	Dirichlet on re-created 2x2 contingency tables		0.90	0.99	Tauman 2020 re-created 2x2 table
WatchPAT 300: low specificity	0.25				0.006	0.81	
WatchPAT 300: high sensitivity	0.88				0.79	0.94	Tauman 2020 reported 95CI limits
WatchPAT 300: high specificity	0.63				0.38	0.84	
WatchPAT ONE: low sensitivity	0.96	Tauman 2020 for WatchPAT 200U	Dirichlet on re-created 2x2 contingency tables		0.90	0.99	Tauman 2020 re-created 2x2 table
WatchPAT ONE: low specificity	0.25				0.006	0.81	
WatchPAT ONE: high sensitivity	0.88				0.79	0.94	Tauman 2020 reported 95CI limits
WatchPAT ONE: high specificity	0.63				0.38	0.84	
Correlation between multiple sleep studies	0	Assumption	Fixed				
Failure rate							

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
Oximetry	4.64%	Newcastle Regional Sleep Service	Beta	Alpha 8.95 Beta 184.05	1.68%	8.10%	Lower value from Newcastle Regional Sleep Service lower 95%CI; Upper value from Rofail 2010
Respiratory polygraphy	5.40%		Beta	Alpha 53.95 Beta 945.05	4.00%	9.00%	Lower value from Newcastle Regional Sleep Service lower 95%CI; Upper value from Devani 2021
AcuPebble	0.6%	Devani 2021	Beta	Alpha 0.99 Beta 165.01	0%	1.77%	Devani 2021 95%CI limits
Brizzy	4.00%	Martinot 2017	Beta	Alpha 3.96 Beta 95.04	0.16%	7.84%	Martinot 2017 95%CI limits

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
NightOwl	11.50%	Lyne 2023	Beta	Alpha 12.88 Beta 99.12	5.62%	17.39%	Lyne 2023 95%CI limits
Sunrise	10.53%	Kelly 2022	Beta	Alpha 3.89 Beta 33.11	0.77%	20.28%	Kelly 2022 95%CI limits
WatchPAT 300	3.28%	Mueller 2018	Beta	Alpha 1.97 Beta 58.03	0%	7.75%%	Mueller 201895% CI limits
WatchPAT ONE	3.28%	Mueller 2018	Beta	Alpha 1.97 Beta 58.03	0%	7.75%	
Time to diagnosis							
Oximetry	3	Assumption	Gamma	Alpha 25 Beta 0.12	1.5	6	Assumption
Respiratory polygraphy	3				1.5	6	
AcuPebble	3				1.5	6	
Brizzy	3				1.5	6	
NightOwl	3				1.5	6	

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
Sunrise	3				1.5	6	
WatchPAT 300	3				1.5	6	
WatchPAT ONE	3				1.5	6	
Time to treatment							
Oximetry	3	Assumption	Gamma	Alpha 25 Beta 0.12	1.5	6	Assumption
Respiratory polygraphy	3				1.5	6	
AcuPebble	3				1.5	6	
Brizzy	3				1.5	6	
NightOwl	3				1.5	6	
Sunrise	3				1.5	6	
WatchPAT 300	3				1.5	6	
WatchPAT ONE	3				1.5	6	
Treatment delay for false negative	1	Assumption	Gamma	Alpha 25 Beta 0.04	0	3	Assumption
Utilities – severity of OSAHS							
Multiplier for mild OSAHS	0.81	NG202	Cholesky decomposition of covariance regression coefficients		0.60	0.97	Lower value from Skirko; Upper value +20% of base case value
Multiplier for moderate OSAHS	0.77	NG202			0.61	0.92	
Multiplier for severe OSAHS	0.74	NG202			0.60	0.88	
Utilities – improvements with treatment							
CPAP-treated mild OSAHS	0.02	Calculated from change in ESS – see Treatment impacts of ESS below					

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
CPAP-treated moderate OSAHS	0.02						
CPAP-treated severe OSAHS	0.03						
Bespoke MAD-treated mild OSAHS	0.023						
Bespoke MAD-treated moderate OSAHS	0.021						
Bespoke MAD-treated severe OSAHS	0	Assumption	Fixed				
Utilities - CV events (males)							
Stable angina new event	0.83	Assumption	Beta	Alpha 3.32 Beta 0.66	0.67	1.00	+/- 20% base case value
Stable angina post event	0.83			Alpha 3.32 Beta 0.66	0.67	1.00	+/- 20% base case value
Unstable angina new event	0.72	Pockett 2018		Alpha 6.18 Beta 2.35	0.58	0.87	+/- 20% base case value
Unstable angina post event	0.83	Pockett 2018		Alpha 3.32	0.67	1.00	+/- 20% base case value

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
				Beta 0.66			
MI new event	0.77	Pockett 2018		Alpha 5.06 Beta 1.54	0.61	0.92	+/- 20% base case value
MI post event	0.87	Pockett 2018		Alpha 2.30 Beta 0.33	0.70	1.00	+/- 20% base case value
TIA new event	0.78	Luengo-Fernandez 2013		Alpha 672 Beta 189	0.75	0.90	Lower value is lower 95%CI limit; upper value is estimate used in NG202
TIA post event	0.78	Luengo-Fernandez 2013		Alpha 603 Beta 170	0.75	0.90	
Stroke new event	0.63	Pockett 2018		Alpha 8.59 Beta 5.03	0.50	0.76	+/- 20% base case value
Stroke post event	0.96	Pockett 2018		Alpha 7.42 Beta 3.56	0.54	0.81	+/- 20% base case value
Utilities – CV events (females)							
Stable angina new event	0.78	Assumption	Beta	Alpha 4.82	0.62	0.93	+/- 20% base case value

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
				Beta 1.39			
Stable angina post event	0.78			Alpha 4.82 Beta 1.39	0.62	0.93	+/- 20% base case value
Unstable angina new event	0.65	Pockett 2018		Alpha 8.13 Beta 4.39	0.52	0.78	+/- 20% base case value
Unstable angina post event	0.78	Pockett 2018		Alpha 4.82 Beta 1.39	0.62	0.93	+/- 20% base case value
MI new event	0.69	Pockett 2018		Alpha 7.01 Beta 3.12	0.55	0.83	+/- 20% base case value
MI post event	0.82	Pockett 2018		Alpha 3.81 Beta 0.86	0.65	0.98	+/- 20% base case value
TIA new event	0.78	Luengo-Fernandez 2013		Alpha 672 Beta 189	0.75	0.90	Lower value is lower 95%CI limit; upper value is estimate used in NG202
TIA post event	0.78	Luengo-Fernandez 2013		Alpha 603 Beta 170	0.75	0.90	

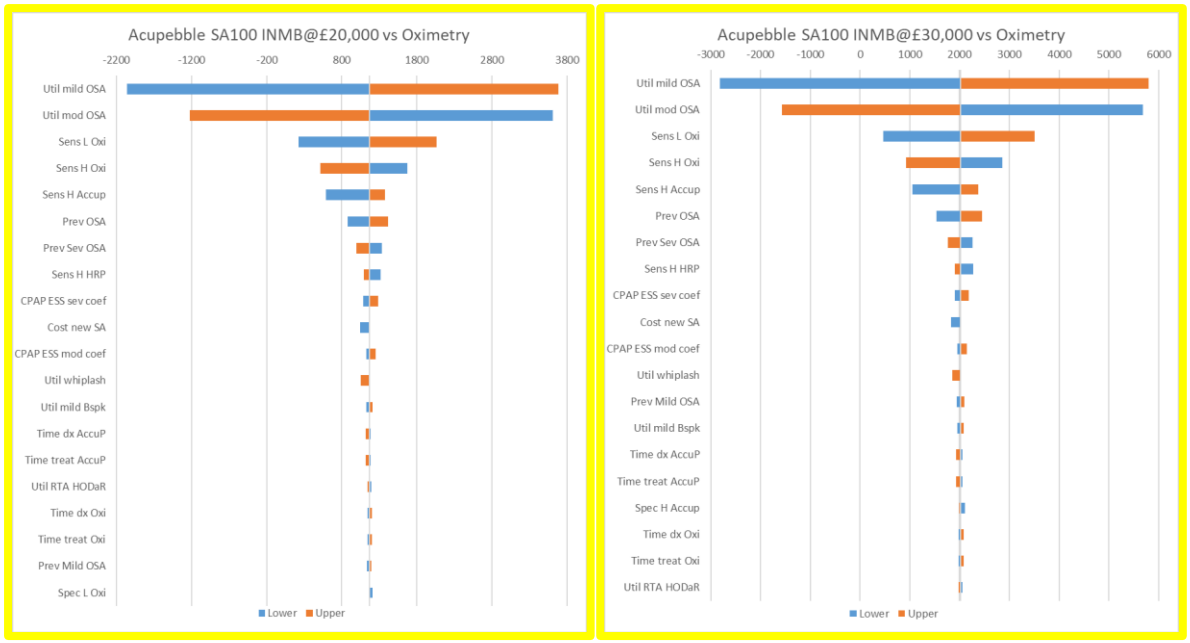
Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
Stroke new event	0.56	Pockett 2018		Alpha 10.54 Beta 8.42	0.44	0.67	+/- 20% base case value
Stroke post event	0.62	Pockett 2018		Alpha 8.93 Beta 5.52	0.49	0.74	+/- 20% base case value
Utilities – RTAs							
Serious	0.62	NG202; Pink 2014	Beta	Alpha 1.38 Beta 0.85	0.09	1.00	95%CI limits
Slight	-0.085	NG202	Normal	Mean -0.085 SE 0.021	-0.13	0.04	95%CI limits
Treatment impacts on ESS							
CPAP-treated mild OSAHS	-1.91	Feltner 2022	Normal	Mean -1.91 SE 0.360	-2.87	-1.2	Lower value from NG202; Upper value is upper 95%CI limit
CPAP-treated moderate OSAHS	-2.21	Feltner 2022		Mean -2.21 SE 0.360	-2.92	-1.51	Feltner 2022 95%CI limits

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
CPAP-treated severe OSAHS	-3.08	Feltner 2022		Mean -3.08 SE 0.321	-3.71	-2.45	
Risk of RTA with untreated OSAS	0.168	NG202	Lognormal	Mean -1.783 SE 0.033	0.158	0.179	NG202
Costs of sleep study							
Oximetry	£17.86	NG202	Fixed cost, but resource use variable (see below)				
Cost to review data and prepare report for novel device	£17.67	20 minutes for band 6 (assumption)	Gamma	Alpha 25 Beta 0.8	£3.42	£37.67	Lower value is 5 minutes for band 5; upper value is 20 minutes for consultant
Cost to review data and prepare report for oximetry	£8.83	10 minutes for band 6 (assumption)		Alpha 25 Beta 0.4	£3.42	£28.25	Lower value is 5 minutes for band 5; Upper limits is 15 minutes for consultant
Costs – CV events							
Stable angina new event	£1,059.50	National Schedule	Gamma	Alpha 25 Beta 42.4	£708	£1,582	Lost value with CC score 0-3; Upper

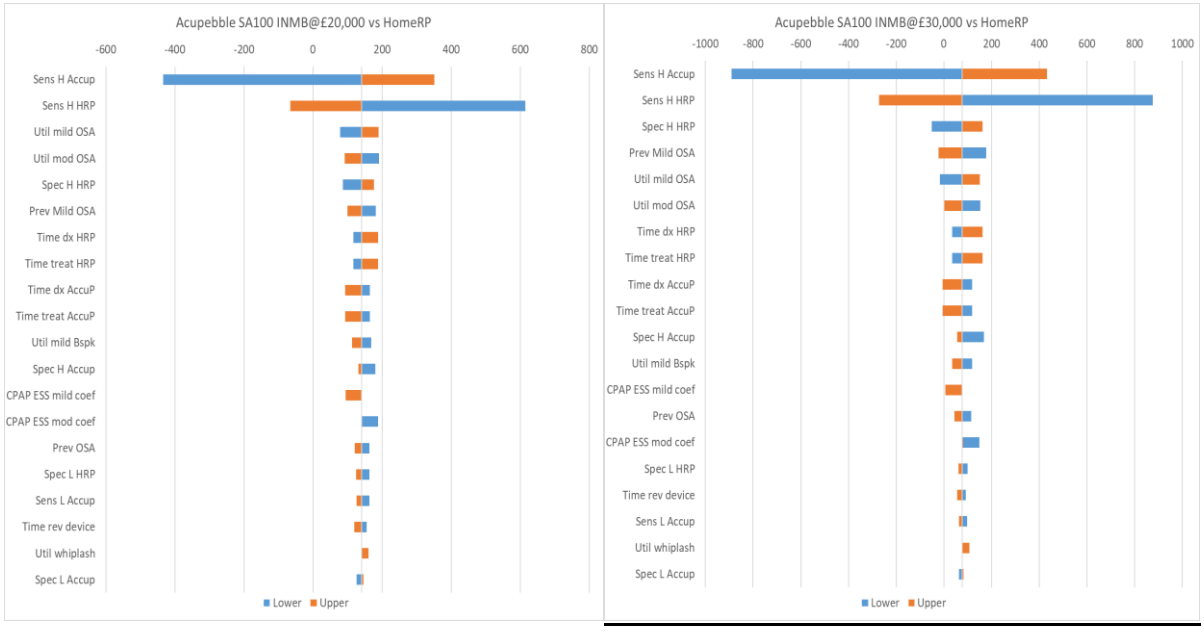
Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
		Reference Costs					value with CC score 12+
Stable angina post event	£297.92			Alpha 25 Beta 11.9	£708	£1,582	Lost value with CC score 0-3; Upper value with CC score 12+
Unstable angina new event	£2,545.31			Alpha 25 Beta 101.8	£2036	£3054	+/- 20% base case value
Unstable angina post event	£297.92			Alpha 25 Beta 11.9	£238	£358	+/- 20% base case value
MI new event	£5,056.65			Alpha 25 Beta 202.3	£4,045	£12,655	Lower value is lower 95%CI limit; Upper value from Zhou 2023 calculator
MI post event	£836.68			Alpha 25 Beta 33.5	£669	£6,929	
TIA new event	£1,902.33			Alpha 25 Beta 76.1	£1,522	£2,283	+/- 20% base case value
TIA post event	£639.25			Alpha 25 Beta 25.6	£511	£767	+/- 20% base case value
Stroke new event	£19,169,30			Alpha 25	£14,144	£23,003	Lower value from Zhou 2023

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
				Beta 766.8			calculator; Upper value is upper 95%CI limit
Stroke post event	£7,277.11			Alpha 25 Beta 291.1	£5,822	£8,733	+/- 20% base case value (note value from Zhou very similar to base case value)
Standardised mortality rates (SMRs) for CV events							
Stable angina	1.95	Rosengren 1998	Lognormal	Mean 0.668 SE 0.086	1.65	2.31	95%CI limits
Unstable angina	2.19	NG24	Lognormal	Mean 0.784 SE 0.033	1.30	2.33	Lower value from Ellis 2019; Upper value is upper 95%CI limit
MI	2.68	Bronnum-Hansen 2001	Lognormal	Mean 0.986 SE 0.041	1.30	2.91	
TIA	1.40	Dennis 1990	Lognormal	Mean 0.336 SE 0.126	1.05	2.60	Lower value from Clarke 2003; Upper value from

Parameter	Base case		Probabilistic analysis		One-way sensitivity analyses[a]		
	Value	Source	Distribution	Parameters	Lower value	Upper value	Source/ Note
							Rutten-Jacobs 2013
Stroke	2.72	Bronnum-Hansen 2001	Lognormal	Mean 1.001 SE 0.024	2.30	2.85	Lower value from Dennis 1993; Upper value from Rutten-Jacobs 2013



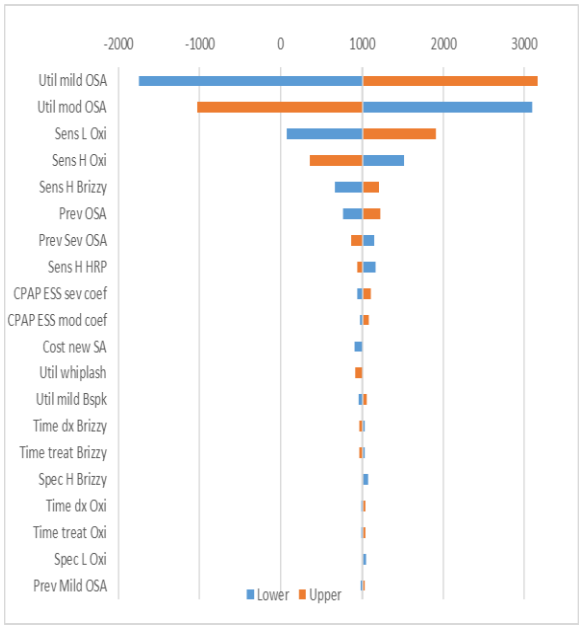
a) vs oximetry at £20,000 per QALY gained b) vs Oximetry at £30,000 per QALY gained



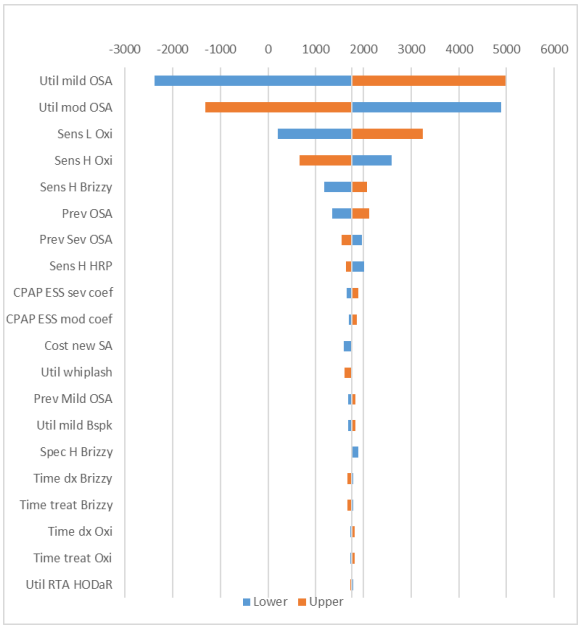
c) vs RP at £20,000 per QALY gained d) vs RP at £30,000 per QALY

Figure 12 INMB for AcuPebble vs oximetry and respiratory polygraphy at £20,000 and £30,000 per QALY gained

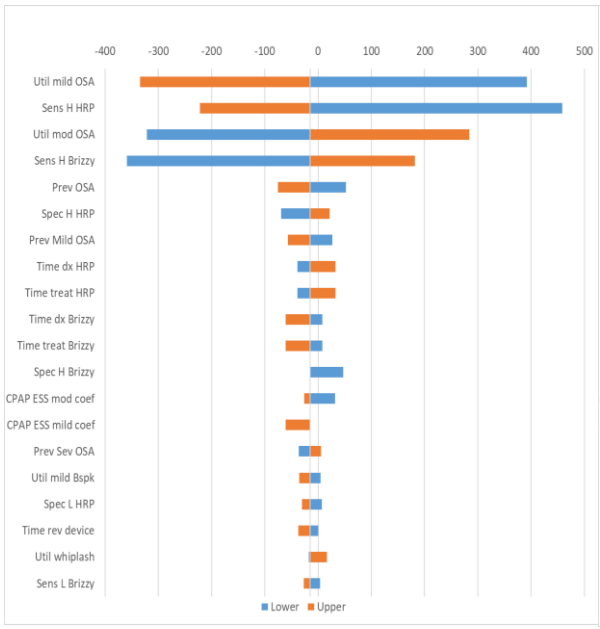
Abbreviations: coef coefficient, CPAP continuous positive airway pressure, dx diagnosis, ESS Epworth Sleepiness Scale, INMB incremental net monetary benefit, MAD mandibular advancement device, Mod moderate, RP respiratory polygraphy, Sens sensitivity, Sev severe, Spec specificity, Util utility



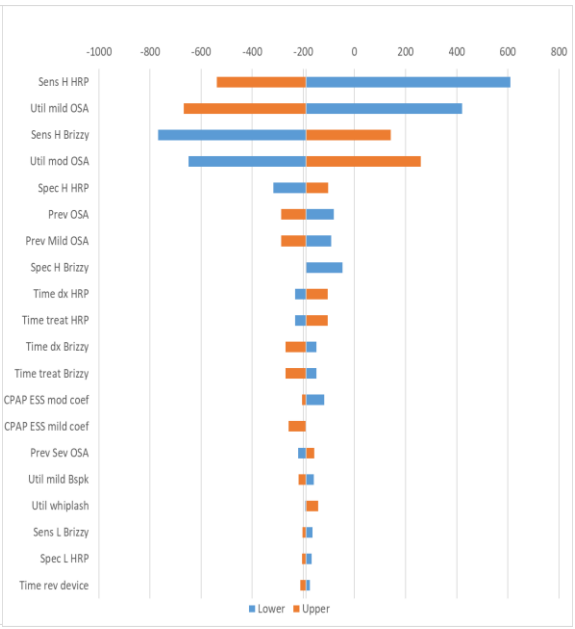
a) vs oximetry at £20,000 per QALY gained



b) vs Oximetry at £30,000 per QALY gained



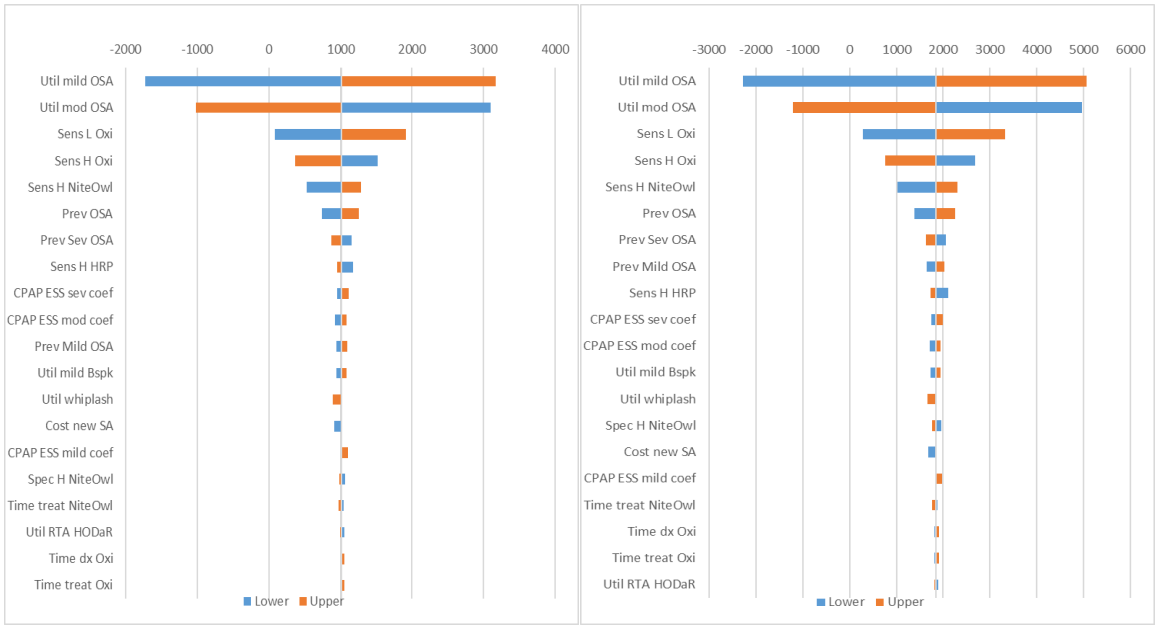
c) vs RP at £20,000 per QALY gained



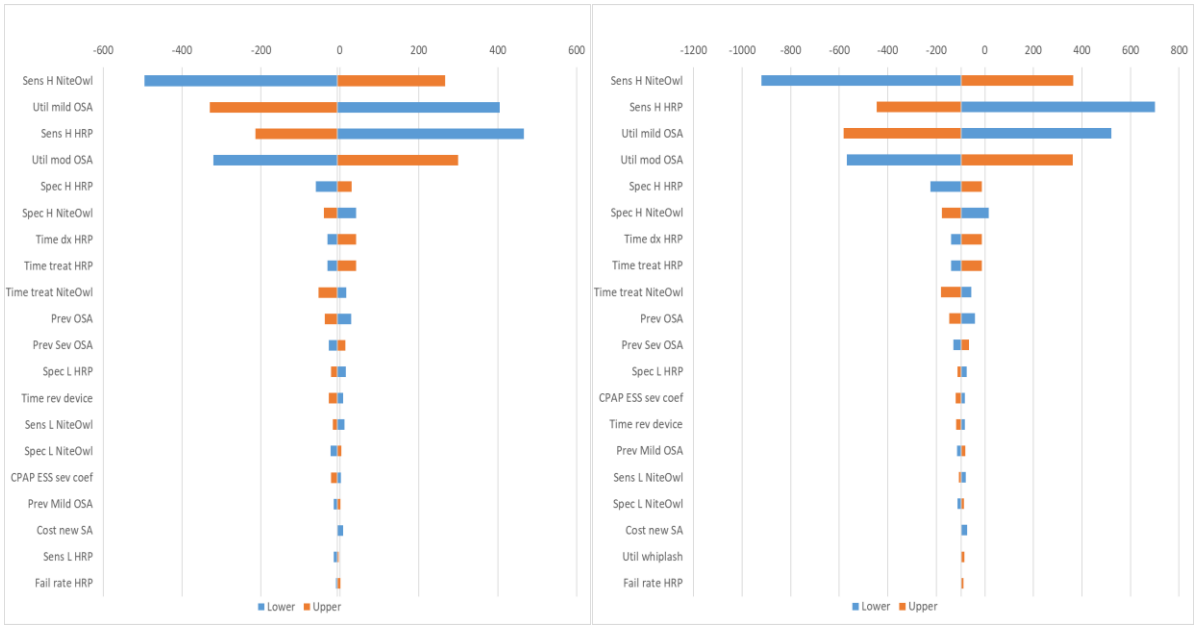
d) vs RP at £30,000 per QALY

Figure 13 INMB for Brizzy vs oximetry and respiratory polygraphy at £20,000 and £30,000 per QALY gained

Abbreviations: coef coefficient, CPAP continuous positive airway pressure, dx diagnosis, ESS Epworth Sleepiness Scale, INMB incremental net monetary benefit, MAD mandibular advancement device, mod moderate, ox oximetry, RP respiratory polygraphy, RTA road traffic accident, sens sensitivity, sev severe, spec specificity, Util utility



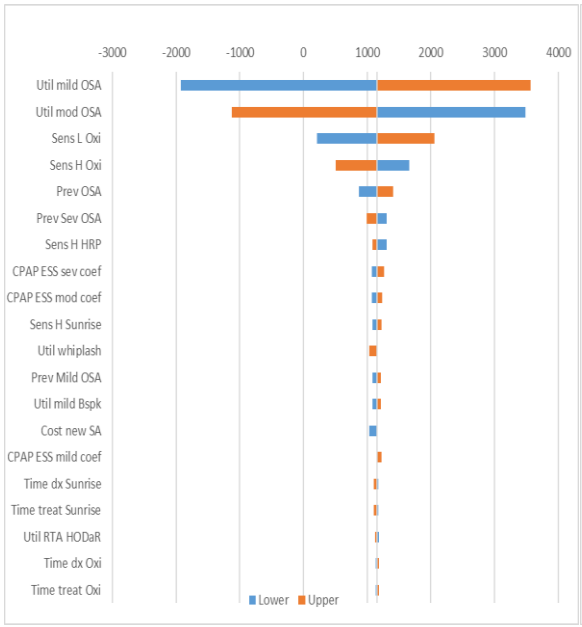
a) vs oximetry at £20,000 per QALY gained b) vs Oximetry at £30,000 per QALY gained



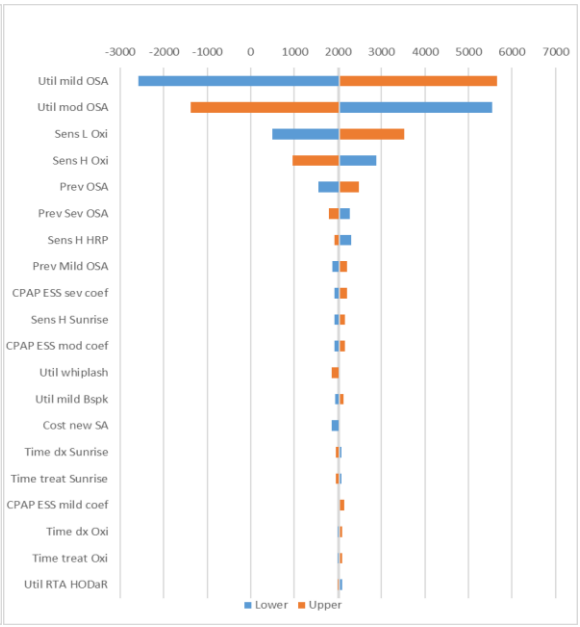
c) vs RP at £20,000 per QALY gained d) vs RP at £30,000 per QALY

Figure 14 INMB for NightOwl vs oximetry and respiratory polygraphy at £20,000 and £30,000 per QALY gained

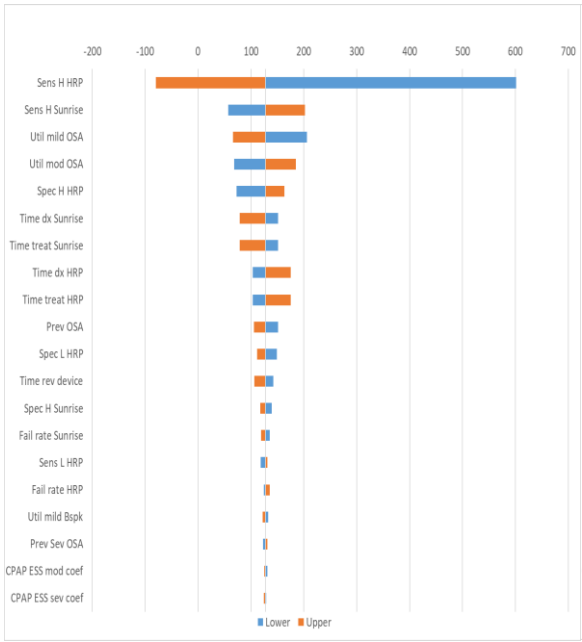
Abbreviations: coef coefficient, CPAP continuous positive airway pressure, dx diagnosis, ESS Epworth Sleepiness Scale, IMNB incremental net monetary benefit, MAD mandibular advancement device, mod moderate, ox oximetry, RP respiratory polygraphy, RTA road traffic accident, sens sensitivity, sev severe, spec specificity, Util utility



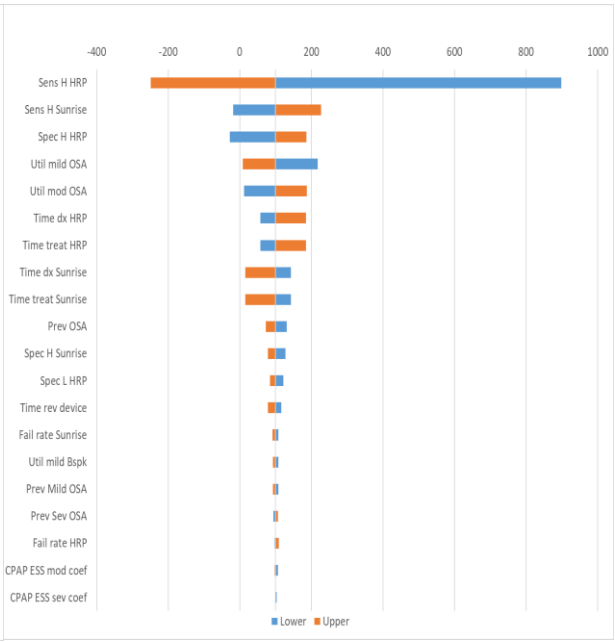
a) vs oximetry at £20,000 per QALY gained



b) vs Oximetry at £30,000 per QALY gained



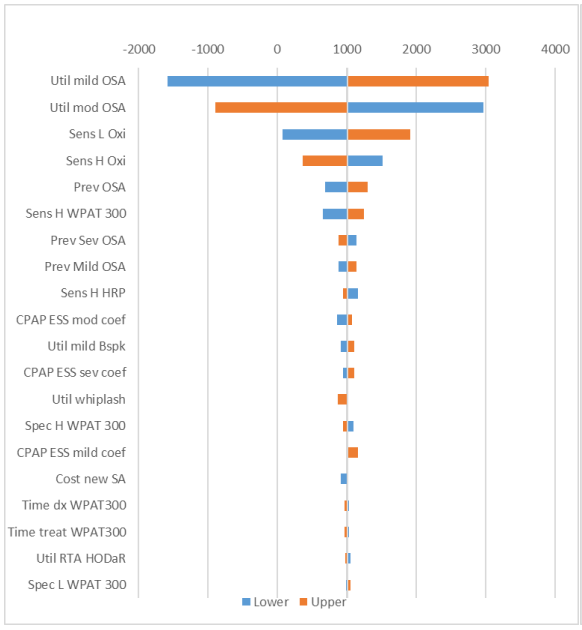
c) vs RP at £20,000 per QALY gained



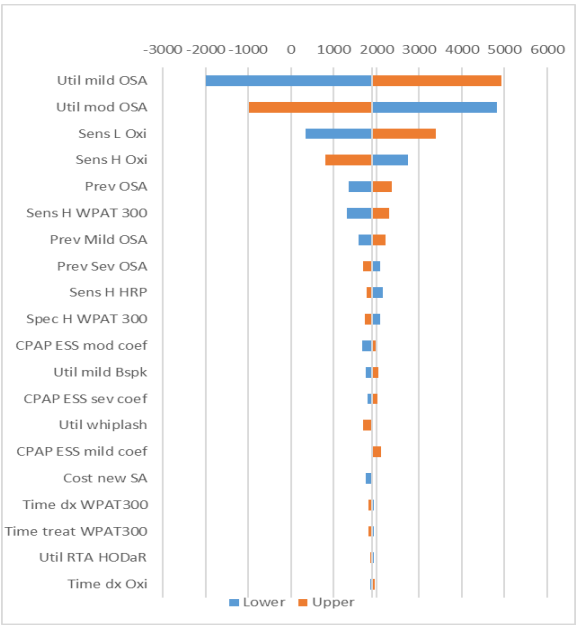
d) vs RP at £30,000 per QALY

Figure 15 INMB for Sunrise vs oximetry and respiratory polygraphy at £20,000 and £30,000 per QALY gained

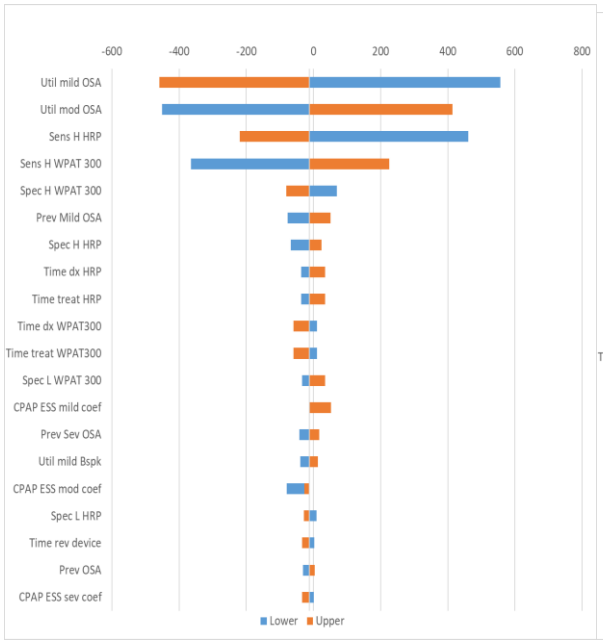
Abbreviations: coef coefficient, CPAP continuous positive airway pressure, dx diagnosis, ESS Epworth Sleepiness Scale, INMB incremental monetary benefit, MAD mandibular advancement device, mod moderate, ox oximetry, RP respiratory polygraphy, RTA road traffic accident, sens sensitivity, sev severe, spec specificity, Util utility



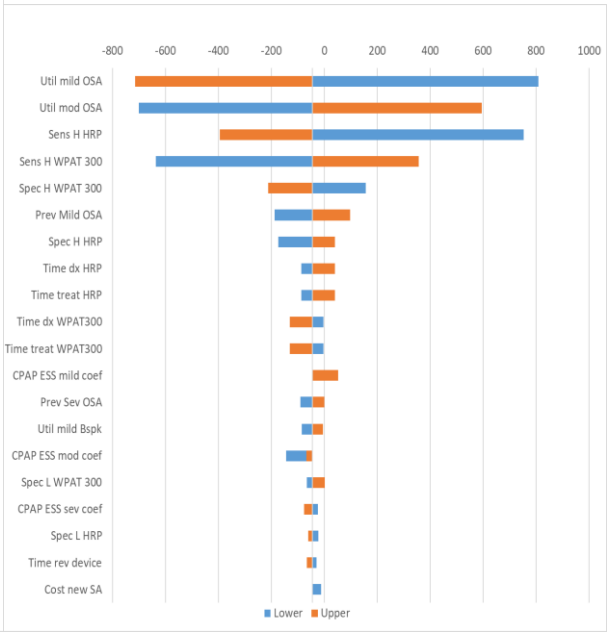
a) vs oximetry at £20,000 per QALY gained



b) vs Oximetry at £30,000 per QALY gained



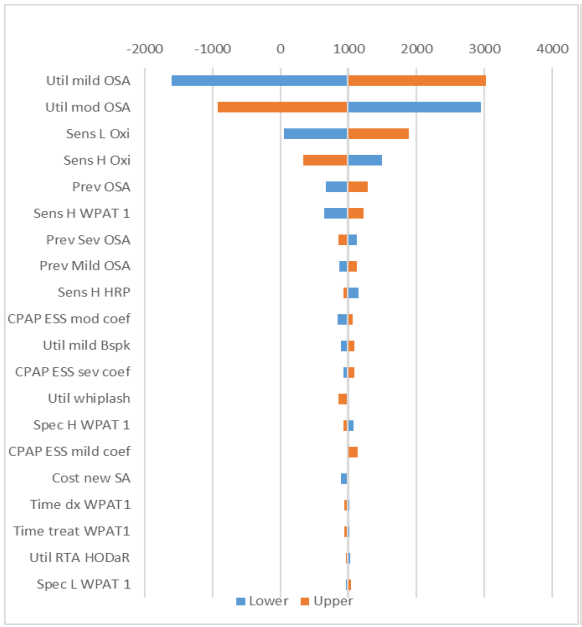
c) vs RP at £20,000 per QALY gained



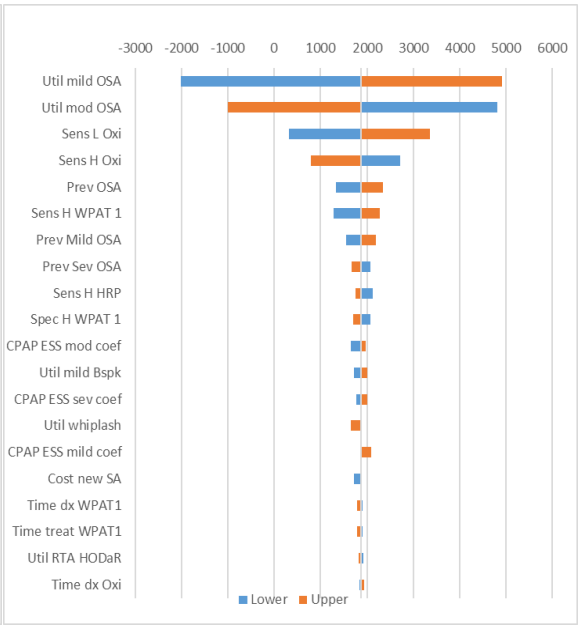
d) vs RP at £30,000 per QALY

Figure 16 INMB for WatchPAT 300 vs oximetry and respiratory polygraphy at £20,000 and £30,000 per QALY gained

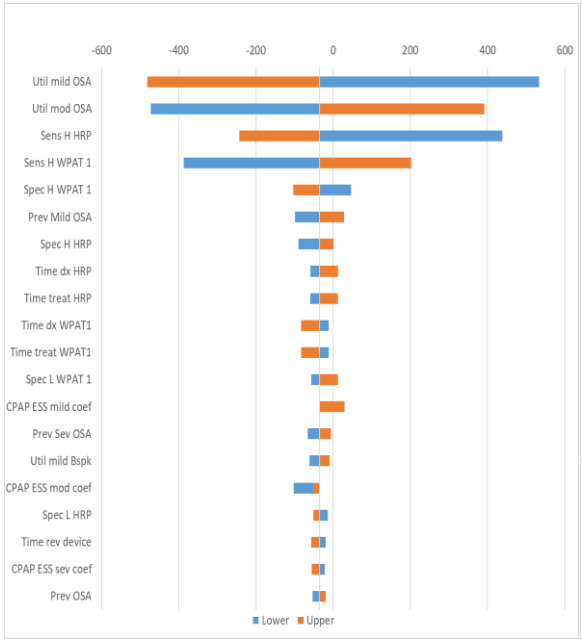
Abbreviations: coef coefficient, CPAP continuous positive airway pressure, dx diagnosis, ESS Epworth Sleepiness Scale, MAD mandibular advancement device, mod moderate, ox oximetry, RP respiratory polygraphy, RTA road traffic accident, sens sensitivity, sev severe, spec specificity, Util utility



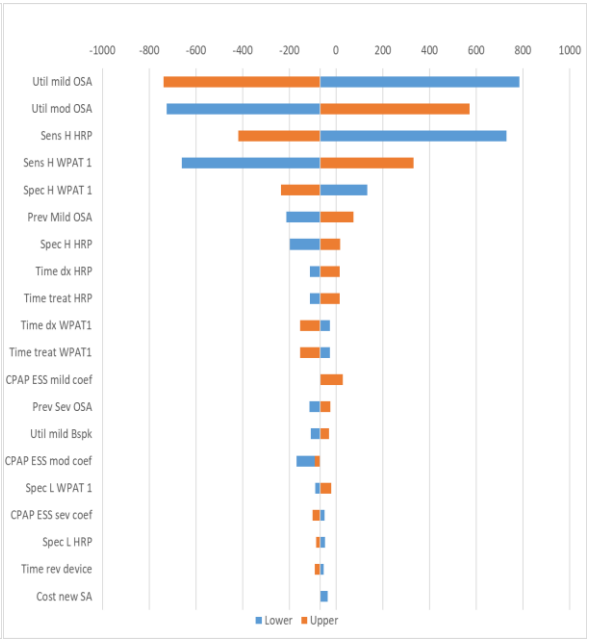
a) vs oximetry at £20,000 per QALY gained



b) vs Oximetry at £30,000 per QALY gained



c) vs RP at £20,000 per QALY gained



d) vs RP at £30,000 per QALY

Figure 17 INMB for WatchPAT ONE vs oximetry and respiratory polygraphy at £20,000 and £30,000 per QALY gained

Abbreviations: coef coefficient, CPAP continuous positive airway pressure, dx diagnosis, ESS Epworth Sleepiness Scale, MAD mandibular advancement device, mod moderate, ox oximetry, RP respiratory polygraphy, RTA road traffic accident, sens sensitivity, sev severe, spec specificity, Util utility

NICE National Institute for Health and Care Excellence

Diagnostics Assessment Programme

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Professional organisation submission

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

NICE National Institute for Health and Care Excellence

About you

1. Your name	Sarah Kearney
2. Name of organisation	Association of Respiratory Nurses
3. Job title or position	Respiratory Nurse Specialist
4. Are you (please select Yes or No):	An employee or representative of a healthcare professional organisation that represents clinicians? Yes A specialist in the treatment of people with this condition? Yes A specialist in the clinical evidence base for this condition or technology? Yes Other (please specify):
5a. Brief description of the organisation (including who funds it).	Association of respiratory Nurses Corporate sponsorship
5b. Has the organisation received any funding from any company with a technology included in the evaluation in the last 12 months? [Please refer to the final scope for a full list of technologies included. The final scope is here] If so, please state the name of company, amount, and purpose of funding.	No
5c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No

Unmet need

6. In your view, is there an unmet need for people with suspected obstructive sleep apnoea/hypopnoea syndrome and healthcare professionals involved in diagnosing the condition?	Yes, there is need for smaller integrated devices that have a multichannel diagnostic approach but with less intrusive or sleep disturbing equipment. Easier to use for patients and potentially with automated plus manual validated data.

The technologies and condition

7. How is the condition currently treated in the NHS?	Multi approaches Gold standard CPAP therapy whether fixed or auto pressure Mandibular advancement splints (in patients with failed CPAP tolerance, mild OSA), Positional therapy (if sleep study reveals a low Ahi in a position) Hypoglossal nerve stimulation (newer in the UK one centre currently) Surgery for some if crowded airway to improve CPAP compliance Maxillomandibular advancement
8. Are any relevant clinical guidelines we should be aware of, and if so, which?	NICE NG 202 TA139
9. Is the pathway of care well defined? Does it vary or are there differences of	Pathway is fairly straight forwards but where the patient is tested for OSA may differ depending on if a GP surgery provides sleep studys or patients go to secondary care. This may depend on commissioning or specialist

opinion between professionals across the NHS? (Please state if your experience is from outside England.)	GP who can provide a service. It is easier and more straight forward that a GP refers into a service that provides both the sleep study testing and subsequent treatment. This reduces delays and aids patient prioritisation.
10. How does healthcare resource use differ between the technology and current care?	A band 2/3 could potentially send out the testing equipment so a qualified physiologist/nurse/ahp doesn't need to do this.
11. In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	Its more efficient to provide novel testing, and then subsequent analysis and treatments via a specialist service. This helps provide a uniformed service, which can be easier and quicker for a patient to navigate. For simple straight forward sleep apnoea referrals, a dedicated team who test and provide the service is better co-ordinated.
12. What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	Depending on the type of novel device, IT, Software or web based infrastructure. Initial purchase of devices or purchase and testing via an external company depending on the device.
13. Do you expect the technology to provide clinically meaningful benefits compared with current care?	The technology may assist in an increase in testing capability if easier to use, less clinical tie interpreting data however on the flip side if a test is sensitive but not specific then further tests maybe required.
14. Do you expect the technology to increase health-related quality of life more than current care?	Not particularly. If the device is easier to use, less repeating of tests then that may improve time to treat which then could raise QOL

The use of the technology

15. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	Depending on the novel approach of the technology this should be easier to use for the patient. Practical professional limitation: storage of devices, postage, courier of devices. Loss of devices or patients not returning them can cause problems. If devices are not returned then less to issue to future patients. Easier if patients do not need to attend a hospital/GP. Patients may need to have their own mobile technology to link to a device which may be a limitation.
16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	Yes the technology innovative. Whether some of the measures have enough evidence is yet to be seen. The main impact will be the ease of use, quick results, accurate results (?), minimal health related benefits over and above current testing
17. Does the use of the technology address any particular unmet need of the patient population?	Potentially for patients who may struggle to attend appointments then posting out some of the smaller novel technology testing equipment may provide for this population. Post therapy checking in patient with high AH and low SPO2.

Sources of evidence

18. Are you aware of any relevant evidence for this assessment? Please provide details	Paediatric sleep-disordered breathing | British Thoracic Society | Better lung health for all (brit-thoracic.org.uk) New BTS paediatric sleep disordered breathing guidline (2023).

Equality

19. Are there any potential equality issues that should be taken into account when considering this treatment?	The testing could be used on any person. Evidence of testing on a variety of people different ethnicity, disability would be useful to identify limitations of testing equipment.
20. Consider whether these issues are different from issues with current care and why.	No different to current testing.

Key messages

21. In up to 5 bullet points, please summarise the key messages of your submission.	• New innovative and less cumbersome methods of providing sleep studys will benefit patients • It is important to understand the limitations of all the devices and where they can be used and where another devcie would be more appropriate. • The IT/Software is an important consideration for clinicians as well as patients if expected to use their own mobile/computers to download the devcie data. • The level of recyclability/reusability is important to consider. •

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

Please select YES if you would like to receive information about other NICE topics - YES or NO

For more information about how we process your personal data please see our privacy notice.

NICE National Institute for Health and Care Excellence

Diagnostics Assessment Programme

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Professional organisation submission

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

About you

1. Your name	Dr Tim Quinnell
2. Name of organisation	British Thoracic Society
3. Job title or position	Consultant Physician
4. Are you (please select Yes or No):	An employee or representative of a healthcare professional organisation that represents clinicians? Yes A specialist in the treatment of people with this condition? Yes A specialist in the clinical evidence base for this condition or technology? Yes or No Other (please specify):
5a. Brief description of the organisation (including who funds it).	The British Thoracic Society is a professional membership organisation, and a registered charity. Funding sources include membership income, income from ownership of academic journals and conferences/short courses.
5b. Has the organisation received any funding from any company with a technology included in the evaluation in the last 12 months? [Please refer to the final scope for a full list of technologies included. The final scope is here] If so, please state the name of company, amount, and purpose of funding.	No
5c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	no

Unmet need

6. In your view, is there an unmet need for people with suspected obstructive sleep apnoea/hypopnoea syndrome and healthcare professionals involved in diagnosing the condition?	Yes. There are a large number of referrals for suspected sleep disordered breathing to sleep services across the UK. Demand currently outstrips capacity. Furthermore, whenever there is a need for patients to attend in person to collect and return home monitoring equipment this puts those living further away and/or with limited means of transport, at a disadvantage in terms of access to sleep diagnostics. To be able to have rapid and easy access to sleep study results would be beneficial to HCPs and patients. This exists in some cases already, particularly with oximetry which is straightforward and relatively quick for someone with the appropriate expertise to interpret. The role of oximetry is however contentious among HCPs and recent NICE guidance recommends respiratory polygraphy as entry level test instead. Those services using polygraphy as entry level test often require manual analysis to be undertaken on all studies, as automated analysis isn't sufficiently reliable. This either requires a trained sleep scientist or the clinician themself to interpret it. Either of these are time consuming and contribute to service pressures and delays. Polygraphy equipment is expensive, cumbersome and not always easy to send to patients (expensive if courier required). Furthermore some kit is not straightforward to apply and patients must be taught how to do so. There is a pressing need to formally evaluate these newer devices (some sleep centres already use them) most importantly to understand validity and reliability and how they compare with existing devices as well as each other. They need to be user friendly, cost effective and have favourable environmental impact profile. The findings will help inform service providers and commissioners when developing sleep services. If the devices satisfy these criteria then they could potentially help improve efficiency of and accessibility to sleep diagnostic services, and potentially reduce inappropriate variation in practice and access to care. Sustainability should be a key consideration in the evaluation. Can environmental impact be included in outcome measures?

The technologies and condition

7. How is the condition currently treated in the NHS?	See scoping document and NICE guidance.
8. Are any relevant clinical guidelines we should be aware of, and if so, which?	NICE guidance. Awareness of AASM (American) guidance would also be helpful.
9. Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	My comments are limited to OSAHS in 16yrs+. Reasonably well defined. There are differences in practice of CPAP initiation. NICE guidance covers this area but recommendations are contentious and may already be out of date. Diagnostics are also contentious, with many services using oximetry rather than polygraphy. NICE recommends polygraphy first line but concedes oximetry can be used with caveats. The case for first line polygraphy is not supported by evidence because as the NICE committee identified this is lacking. Health economics analysis which found oximetry to be less cost effective than polygraphy is considered by some to have been based on flawed assumptions. Oximetry is also arguably significantly more accessible and more efficient than polygraphy.
10. How does healthcare resource use differ between the technology and current care?	Difficult to say. Where newer devices are used there may be nil or reduced time required for trained analysis of data compared to existing kit. If they provide accurate polygraphy level results as efficiently and conveniently as oximetry then they could potentially be a gamechanger provided criteria discussed in (6) are met. This will depend on device. Where deliverable this could vary significantly from services that require patients to attend in person to collect /return equipment.
11. In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	Potentially controversial. Many would argue that their use should remain limited to dedicated sleep apnoea clinics, or respiratory clinics which have a sufficient OSA diagnostics case load to enable development and maintenance of expertise (i.e. secondary/tertiary care). It is essential for OSA diagnostic results to be interpreted by experienced clinicians and applied within the clinical context. Others would argue that while they should first be used in secondary/tertiary care they might eventually have a role in primary care where clinical suspicion of OSA is high. However experience of some members of the group is that primary care led screening can lead to over-referral.
12. What investment is needed to introduce the	Will depend on device and info likely to be available from commercial providers. Investment in training of staff to use devices and interpret/apply results.

technology? (For example, for facilities, equipment, or training.)	
13. Do you expect the technology to provide clinically meaningful benefits compared with current care?	Yes IF significantly improves efficiency and access that allows more patients to be treated quickly and appropriately, and through this waiting list benefits. One risk of making it easier to diagnose OSA through better access to more sensitive kit, is that there is potential to lower threshold for initiating CPAP therapy. This may theoretically lead to lowering of adherence rates which could adversely impact cost effectiveness. This would need to be monitored.
14. Do you expect the technology to increase health-related quality of life more than current care?	Yes, potentially, bearing in mind all above.

The use of the technology

15. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	Experience is that much easier to use than polygraphy overall (accupebble, WatchPAT etc all very feasible to use, apply), possibly fewer repeated tests. Just need to understand limitations of use. Need efficient process for return of devices (e.g.courier better than post but more expensive) and dedicated staff managing process/ available for troubleshooting. Useful to have other parameters (such as AHI, RDI, snore etc) compared to oximetry alone but ref point in (13) about potential impact of more sensitive kit on CPAP treatment threshold. Is there a need to evaluate how well patients feel connected with their care with tests being organised prior to first consultation for example, if say from primary care, or before secondary care OPA?

16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	Yes, potentially, if more accessible and provides reliable, accurate results with less/no need for expert manual analysis.
17. Does the use of the technology address any particular unmet need of the patient population?	Those who have difficulty accessing sleep diagnostics that require healthcare venue attendance.

Sources of evidence

18. Are you aware of any relevant evidence for this assessment? Please provide details	

Equality

19. Are there any potential equality issues that should be taken into account when considering this treatment?	There are concerns about oximetry being less accurate in non-Caucasian populations. Evidence is insufficient to provide certainty about the nature or scale of the issue. If any of the technologies being examined don't use oximetry (remembering standard polygraphy also requires it for scoring AHI) and is found to be accurate and reliable, then that could potentially be advantageous and provide reassurance about test accuracy in defined populations. New devices may favour those patients able to use technology better. Therefore overall there is a need to ensure use of the most appropriate test for the individual, and for this requirement to be incorporated into an efficient pathway capable of dealing with large patient numbers.
20. Consider whether these issues are different from issues with current care and why.	These issues are already present in current care.

Key messages

21. In up to 5 bullet points, please summarise the key messages of your submission.	• User friendliness, accessibility and efficiency of new tests over existing ones needs to be better to justify using them. • Costs of new tech – need to be on par with those in use or sleep services won't be able to afford them. • Accuracy of new tech is key and should be rigorously examined given many of the devices use novel surrogate markers of respiratory events. • Sustainability/eco impact particularly important for single-use.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

Please select YES if you would like to receive information about other NICE topics - YES or NO

For more information about how we process your personal data please see our privacy notice.

Patient Group Submission Template for Diagnostic Technologies

NICE
Health Technology Assessment (HTA) on
Automated home testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome (provisional title)

Please read the accompanying guide fully before completing this submission template.

Information about your organisation	
Organisation name	Sleep Apnoea Trust
Contact person's name	Chris Rogers
Role or job title	Managing Secretary & Trustee
Email	chris.rogers@sleep-apnoea-trust.org
Telephone	07798 588838
Postal address	PO Box 60, Chinnor, Oxon, OX39 4XE
Organisation type	Patient/carer organisation (e.g. a registered charity) ☒ Informal self-help group ☐ Unincorporated organisation ☐ Other, please state:
Organisation purpose (tick all that apply)	Advocacy ☒ Education ☒ Campaigning ☒ Service provider ☐ Research ☒ Other, please specify:
What is the membership of your organisation (number and type of members, region that your group represents, demographics, etc)? 1500 sleep apnoea patients in the UK who pay an annual subscription to support the charity	

Patient Group Submission Template for Diagnostic Technologies

Declarations	
Do you have any conflicts of interest? None	
Did anyone outside your organisation help you prepare this submission?	Yes ☐ No ☒
If yes – who helped you and in what way? Please tell us if the people helping you were paid and if they have any conflicts of interest.	
Are you willing for this submission to be shared on our website?	Yes ☒ No ☐
We may invite you to a scoping meeting where this technology is to be discussed. Would a member of your group be willing to join such a meeting (this may be in person or virtually)?	Yes ☒ No ☐

Patient Group Submission Template for Diagnostic Technologies

Impact of the symptoms, condition or disease on patients

1. How do symptoms and/or the condition or disease affect patients' lives or experiences?

The condition of Sleep Apnoea is unlike any other. There is no obvious sign of injury or disability, just an increasingly debilitating tiredness leading to an overwhelming and uncontrollable desire to sleep during normal waking hours. Its main symptoms, snoring, unrefreshing sleep, waking headaches, excessive sleepiness, tiredness or fatigue, nocturia, choking during sleep, fragmented sleep or insomnia, cognitive dysfunction or memory impairment, can have a devastating effect on a person's health, their relationships and their employment. It is a notifiable medical condition to the DVLA and can, if not treated, stop a person from driving. It also is a risk factor if a person is involved in a safety critical job.

Impact of the symptoms, condition or disease on family and carers

2. How do symptoms and/or the condition or disease affect carers/unpaid care-givers and family?

Sleep Apnoea can have a devastating effect upon a person with a family. It can place a strain on relationships, lead to sleeping in separate bedrooms, stop bed partners from sleeping, cause anger and irritation, and possible loss of income or family breakdown. The sleep apnoea sufferer cannot fully contribute to family life or share the challenge of bringing up children.

If people develop this condition combined with dementia or Alzheimer's, as their condition deteriorates, the challenge for carers can be significant. Using CPAP therapy does rely on short and medium term memory and it is sometimes the case that CPAP therapy may not be able to continue.

Patient Group Submission Template for Diagnostic Technologies

Experiences and availability of current diagnostic technologies

3. What role do currently available diagnostic technologies play in helping patients manage their symptoms and/or the condition or disease?

Diagnosis, which normally requires taking a person's sleep history and requiring them to carry out a sleep test, can be onerous. It used to mean two or three visits to a Sleep Clinic following a GP referral. However, the new diagnostic equipment that is being slowly introduced has the capability to transform this process from GP referral. It can mean a phone call from a sleep specialist clinician to get the sleep history and the postal delivery of portable advanced diagnostic device that can be worn like a wrist watch for one or two nights. It is then returned by post and the diagnosis is made. If being treated using CPAP therapy, telemonitoring by the Sleep Clinic can assist the person in managing their equipment and getting settled on using it regularly and successfully. However, at present this is a random process, based on hospital trust's priorities, rather than a national NICE style standard that delivers excellence.

4. What unmet information needs do people currently have due to the lack of an available diagnostic technology for their symptom or condition?

NICE NG202, the new OSAHS guideline is an outstanding manual on the process of referral, diagnosis and treatment. It is easily understandable for prospective patients but it is not backed up by reality.

Primary care referral rates varies significantly across the UK and can be a barrier to successful progress in treating the millions as yet undiagnosed. Then the diagnostic pathway varies considerably, as does the use of advanced diagnostic equipment. Therefore, at present it is a postcode lottery.

It is accepted that, while there is progress in management of the sleep apnoea via developments in the diagnostic pathway as part of the NHS Future Collaboration project, this needs to be matched by a similar process to identify the best and most affordable new diagnostic technology.

It will enhance, simplify and improve the process to a nationally recognised standard that gives people the confidence that the process of referral, diagnosis and treatment of sleep apnoea, matches the high standards laid down in the new NICE OSAHS Guideline, NG202.

Patient Group Submission Template for Diagnostic Technologies

About the diagnostic technology being assessed

5. What are the most important things people would like to gain from the information provided by, and/or the use of, the diagnostic technology being assessed?

More rapid and accurate diagnosis and treatment is vital. Any delay can make a difficult situation worse, especially if their employment is a vocational driving job or where vigilance is critical for safety. Additionally, in the early stages of CPAP therapy, more support is needed from Sleep Clinics to help them through what can be a challenging process, of going to sleep every night with a small electric pump and a facial, or nasal mask or nasal pillows on their head. Again, this is already identified in the Sleep Disorder diagnostic pathway improvement that is under development.

6. For those people with experience of this diagnostic technology, what difference did the information provided by, and/or the use of, the technology make in their lives or the lives of family and carers?

We would need to do a new patient survey to answer such a specific question. CPAP treatment can be so life changing for most people with sleep apnoea. Anything that can speed up the process dealing with the 2.5 million plus undiagnosed moderate of severe OSAHS sufferers in the UK (NICE) and the many with mild symptomatic OSAHS, would not only improve the lives of so moany people, it would save the NHS millions in not having to deal with the comorbidities that proliferate with undiagnosed OSA (NICE TA139). However, in order for new diagnostic technologies to be used, they must be proven to be of sufficient accuracy to help rather than hinder progress in identifying people with undiagnosed sleep apnoea

7. For those without experience of this diagnostic technology, but who are aware of studies or other sources of evidence of value, what are the expectations/limitations of having the information provided by the diagnostic technology and/or using the diagnostic technology?

That they would be referred, diagnosed and treated within the 18-week NHS target. It is clear from our observations of the social media forums (for a) that information shared on them is wildly wrong and self-diagnosis or Dr.Google can only be replaced by early and accurate referral to medical expertise.

Patient Group Submission Template for Diagnostic Technologies

Additional information

8. Please include any additional information you believe would be helpful in assessing the value of the diagnostic technology (e.g. equality issues, ethical or social issues and/or socio-economic considerations).

There is sufficient information in the community to create awareness that diagnosis in the NHS is currently lacking and the provision of CPAP support so delayed, that there is a growing pressure towards private care should a persona jib be at risk because of the delay.

Key messages

9. In up to five statements please list the most important points of your submission.

- A NICE standard, applied to the diagnostic technology pathway would drastically improve the service people receive.
- NICE would bring the very best affordable technology to a person's home and avoid un-necessary hospital visits.
- The efficiency gained would allow much better support of newly diagnosed patients in the early days of their therapy from Sleep Clinic staff
- By improving the diagnostics technology pathway, it would elevate the whole treatment process to achieve the standards set by the new NICE OSAHS Guideline, NG202.
- It would, through is success in reducing comorbidities, save the NHS millions of pounds.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Contents

Section A: External Assessment Report - Comments and EAG response

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Association of Respiratory Nurses	1	5	Results	Sunrise: you mention that Sunrise has small but positive incremental benefits. Is this an appropriate part of the document to put this information as there is no elaboration on the device. This particular device has limited channels and would also be limited in how it could be used and is therefore could be argued is not of importance in this section. Later on page 78 there are comments on the failure rate of such devices which is greater than RPG.	This is the abstract which summarises key aspects of the report. Due to abstract's restrictive word limit it is not possible to include background information and context for each result. However, further elaboration can be found in the main body of the report. See section 1.3 for a description of the novel devices, including key limitations; section 4.5.6 for a summary of evidence on test failure rates; and section 5.7.5 for information on how test failure is included in the economic model.
Association of Respiratory Nurses	2	6	Conclusion	The paragraph states: It is difficult to assess the cost-effectiveness of the novel devices compared with respiratory polygraphy, and the clinical and economic effects of the different novel devices are unclear. And therefore from this a recommendation would be not to use novel devices? It is not clear what the message is here or how this helps.	We refrain from making explicit recommendations for practice at this stage because our role is to inform the NICE diagnostic advisory committee decision making and guidance to the NHS. When NICE guidance is published, there will be a link to the EAG report on the NICE website, as one of the evidence sources informing the guidance. Thus, the reader will be able to cross refer between the report and the recommendations in the guidance.
Association of Respiratory Nurses	3	75 And 85	4.5.7	Number of repeat sleep studies done (people over 16 years of age). It is not clear how when the novel devices do not report on how many repeat studies were needed how the economic modelling factored this in or whether a percentage was taken e.g. 10% failure rate, The number of repeated test could have significant impact on cost and test accuracy. This may have been calculated and I have missed it.	For devices where failure rates have not been available from the included clinical effectiveness studies, we have used alternative sources to inform this in the model. This is described in Section 5.7.5 of the original Report, and in the Erratum. For further information, please also see responses to comments 26, 27, 128, 162 and 174 below.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Sleep Apnoea Trust Association	4	General	General	Sleep Apnoea Trust Association (SATA) acts in the interests of patients , carers, and families to increases awareness , support people in the diagnosis and effective treatment of OSAHS . Novel devices must be simple and easy for people to use at home, but we note that trials in a home setting have been limited if occurring at all which is cause for concern. For example, the Chinese home study(Xu et.al. 2017) with 80 participants is not necessarily translatable into a UK context. There is a reference to the design of the devices involving patients but appeared to be no reference to this or feedback to assure how each had met " user testing ".Since there was uncertainty from the work about the different specifications as these applied to models of devices it was not possible to track if or how change management based on feedback was used to improve the usability of the design of each model of device. There seems no carer involvement or reference and extremely limited evidence of patient reported outcomes submitted for this work . In looking at failures in tests there are references to " user related factors " which was concerning and yet at the same time limited or no " patient training " stated to be required. In the modelling there has been some assessment of NHS staff time for training in correct use and support but the effectiveness of this would in turn rely on what device and how it was being supplied to the patient at home . We assume that although there is reference to trying to improve patient convenience and reduce their costs this is in fact excluded from the modelling apart from some assessment and indication about who may meet assumed postage costs for supply and return . On a practical point given the changes to logistics infrastructures and services it is unclear what assumptions are made for patient access and convenience of returns of devices (home collection , drop off points etc as all would influence and facilitate timely returns). We note that for a range of reasons overall the work was unable to confirm potential utilisation of devices in children aged 2-16 years although there were research recommendations proposed that may help further this activity . SATA will rely on the clinical and specialist expertise to assess the clinical and cost effectiveness of the devices to the NHS, but the following points arose from our review so are provided for information and consideration as or where relevant. There is variation in practice between sleep centres and also variation in what the terms oximetry and respiratory polygraph refer to. We assume this has been understood and taken into account when considering the existing processes and pathways to arrive at cost comparisons between each of the devices . It is also assumed for example that as the comorbidities of some patients excluded them from the use of certain devices there was an assessment as to how this impacted on the profile of the population being used in the modelling work. The wording being used concerning the functionality of a device appeared at times to apply to what "could be offered ". This seemed to indicate different specifications presumably at different costs (standard versus add ons or upgrades) and left uncertainty as to how or if this had been considered. This seems an important factor for any commissioning body to consider and to be able to account for in assessing the business case for any overall NHS changed pathway costs for service provision .	*Xu et al. 2017* We acknowledge the limitation of the study by Xu (please see section 5.7.4 of the Erratum) and have added further detail on this study. *Patient experience* Any potential for greater acceptability of the novel devices over the comparators is not directly captured in the model. As the sleep studies occur over just a few nights at most, any attempts to incorporate acceptability in terms of QALYs would have a negligible impact on the model results. Acceptability may be captured indirectly via the failure rates for the novel devices (should they be lower than those for the comparators). Given the difficulties and likely small impacts on the model results of trying to capture these potential benefits, such claims are best dealt with as part of the deliberative process and depend on the evidence for greater acceptability over the comparators. *Variation in current practice* We are aware of variation in practice and have tried to explore this in the scenario analyses we have undertaken, for instance in the assumptions on how devices get to/from the patient, the types of treatments offered to patients. Please see Table 38 in Section 5.10.2, which details all scenario analyses undertaken. Tables 47 and 77 show results for scenario analyses related to variation in practice. *Comorbidities* Due to the variability of comorbidities across devices, and a lack of evidence on the likely size of populations having to be excluded

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				It was somewhat surprising that the team were unable to access NHS supply chain costs to inform this work. We note the extensive work that has been undertaken in carrying out this assessment in particular the effort made in trying to consider available evidence and make adjustments to draw reasonable comparisons . We trust that there will be many points of learning in addition to the research recommendations that will to enable the work and evidence on the devices to be further developed to overcome the current variables and uncertainties identified	from use of the difference devices, we did not include the impact of comorbidities in the model.
British Paediatric Respiratory Society	5	5	Abstract	"Sensitivity was generally high, in the range 80 to 100%, and fell below 80% in just two instances." It is not clear whether an instance refers to a study or to one of the devices. Please could this be clarified.	In theory an instance could be sensitivity / specificity at one or more cut-off values within a study, or between studies. For the latter this could be the same device or different devices. We alluded to this in the preceding sentence: *"Sensitivity and specificity estimates vary across the studies and also within studies at different severity cut-offs".* The quoted text refers to two studies of two different devices, i.e. in one study which evaluated the Brizzy device, and in another study which evaluated Sunrise. In each study, sensitivity was below 80% at only the highest cut-off value tested. To avoid confusion we have replaced instances with studies.
British Paediatric Respiratory Society	6	26		Please could the abbreviations INMB and ROC be added to the abbreviations list	These have been added.
British Paediatric Respiratory Society	7	29	1	The definition of apnoeas and hypopnoeas is different between adults and children. Please could the definition for children be added as per the AASM criteria	The definition of apnoeas and hypopnoeas for children has been added in-line with the AASM criteria.
British Paediatric Respiratory Society	8	29 and 181	1	Please could the spelling of Down Sydrome (not Down's syndrome) be corrected. This is a common mistake which is a particularly senstive topic for families of children with Down Syndrome. This error also appears later in the document	All spellings, with the exception of journal article titles in the reference section of the report, have been corrected to "Down Syndrome")
British Paediatric Respiratory Society	9	30	1	Please could consideration be given to rephrasing the following sentence in the respiratory polygraphy section " Straps are fastened around the torso and chest". The torso includes the chest. Perhaps wording along the lines of "fastened around the chest and abdomen" would be better	This has been rephrased with the suggested wording
British Paediatric Respiratory Society	10	throughout		Spelling of the words hypopnoea and apnoea and interchanged with hypopnea and apnea (usual American spellings). It would make sense to be consistent and use the UK spelling throughout	Spellings have been changed to UK variant throughout the report with the exception of journal article titles in the reference section, literature search strategies and direct quotes from other sources.
British Paediatric Respiratory Society	11	31	1.2.1	In people <16 years of age the corresponding AHI and ODI scoring criteria are scaled down to reflect the fact that children breath faster than adults. Is this really the case? I thought it was related to normative AHI values for children being lower and morbidity seen at lower AHIs in children.	This section has been revised and the discussion of adult and childhood AHI and ODI criteria has been moved to section 1.3.7.
British Paediatric Respiratory Society	12	37	1.3.5	If the Watchpat is suitable for adults and children aged 12 and above I don't understand why there is no data available on children 12 and above. The company website talks about how they validated the watchpat for diagnosing OSA in adolescents, but gives no details. Would it be worth asking for the data from the company? Doing a very quick pubmed search, there is a paper :Choi	The paper by Choi et al., 2018 (Choi JH, Lee B, Lee JY, Kim HJ. Validating the Watch-PAT for Diagnosing Obstructive Sleep Apnea in Adolescents. J Clin Sleep Med. 2018 Oct 15;14(10):1741-1747. doi: 10.5664/jcsm.7386. PMID: 30353803; PMCID: PMC6175781.) was identified in the EAG's searches. The full paper of this study

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				JH, Lee B, Lee JY, Kim HJ. Validating the Watch-PAT for Diagnosing Obstructive Sleep Apnea in Adolescents. J Clin Sleep Med. 2018 Oct 15;14(10):1741-1747. doi: 10.5664/jcsm.7386. PMID: 30353803; PMCID: PMC6175781. It was likely identified in the search, so would be worth mentioning why it did not mean inclusion criteria in the text. In contrast, Nightowl is for children aged 13 and over and you have clearly stated that "One completed study comparing NightOwl to PSG, NCT0476473 2021, 10 with an age-related inclusion criterion of "13 years and older", meets our inclusion criteria. However, it does not report any results. Attempts to obtain clarification or data from the study investigators have so far proved unsuccessful. For the sake of clarity, we will not refer to this study in the remainder of this chapter". This makes it very clear to the reader why there is no data, so maybe something similar could be done for Watchpat?	was screened for eligibility and did not meet the inclusion criteria for the review as the device used was WatchPAT 200 (only WatchPAT 300 and WatchPAT ONE were eligible for inclusion). This study together with the reason for exclusion is listed in Table 56 ("Full text publications excluded from systematic review of clinical effectiveness") of the report.
British Paediatric Respiratory Society	13	37	1.3	It would be helpful to describe the size of the devices for those where there are studies in children in particular Sunrise and Acupebble and whether any safety data in children has been published. My understanding is that they are both relatively small devices that could pose potential choking hazards in young children. Children particularly those with OSAHS are restless sleepers. This restlessness with frequent arousals could easily result in removal of these devices with children then placing them in their mouths. This is a concern in an unsupervised home environment. These risks might by exacerbated in children with co-morbidities and cognitive impairment. This limitation should be acknowedged (potentially in the limitations section)	Device sizes have been added for Sunrise and AcuPebble. As stated in section 1.3.1 of the report, AcuPebble SA100 is only intended for use in adults. For devices already intended for use in young children (Brizzy and Sunrise) relevant safety-related warnings have been added to the text.
British Paediatric Respiratory Society	14	67	4.3.2	[redacted]	[redacted]
British Paediatric Respiratory Society	15	83	4.8.1	Table 10 is difficult to interpret in the absence of the Acupebble data which I assume has been redacted by the company. It would obviously be helpful if this could be shared	The company has provided this information as academic in confidence. Any decision to lift the redaction would need to be agreed between the company and NICE.
British Paediatric Respiratory Society	16	84	4.8.2	It would be helpful to clarify further the methodology behind the 2015 Martinot study. The detail suggests that mandibular movements are being referenced against pulse transit time but on review of the paper they are referenced against obstructive and central events as determined by PSG. I think rewording this text to reflect this would be helpful	We have added a note that the study included examination of the timing of MM and PTT during episodes of OAH and central sleep apnoea.
British Paediatric Respiratory Society	17	85	4.8.2	It would be helpful to understand whether the 95% limits of agreement are available for the Martinot PSG vs Sunrise study as this would give a better idea of variability from the presumed gold standard test (PSG). It would also be helpful to understand if there was any change in the 95% limits of agreement across different severities of obstructive sleep apnoea	The 95% CI is provided for the intraclass correlation coefficient from the Martinot 2022 study for PSG vs Sunrise (see final paragraph in 4.8.2).
British Paediatric Respiratory Society	18	181 and 185	5.11.1 and 5.11.2	The statement "Thus, the guidelines indicate that for children with no comorbidities, home sleep studies are only relevant where mild sleep disordered breathing is suspected, or where there is inconsistency between symptoms and findings from sleep questionnaires and clinical examination." Consideration should be given to expanding this sentence to include children without co-morbidities referred for tonsillectomy with suspected severe OSA. Whilst this isn't in the BTS OSA pathway referred to, it is a good practice point in the BTS guideline recommendations hence this should be considered.	Thank you for your comment, we agree that consideration of good practice points is also important. However, it is not clear which good practice point is being referred to in this comment. There is a recommendation on use of pulse oximetry from BTS guideline question 11 on oxygen saturation monitoring before tonsillectomy: "*A pre-operative pulse oximetry sleep study before tonsillectomy (with or without adenoidectomy) may be considered for children without comorbidities with suspected severe OSA*" (BTS Online Appendix 11). But as we understand it, this point relates to the use

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					of pulse oximetry to predict peri-operative complications for surgery planning, so it is not directly relevant to the use of home sleep studies for diagnosis. We have edited the sentence in section 5.11.1 our report to be clear that it relates to the use of diagnostic sleep studies.
British Paediatric Respiratory Society	19	182 and 185	5.11.1 and 5.11.2	The document states that in children experts on the panel state that only oximetry is undertaken at home and RP would be conducted in hospital. There is a variation across the UK and there are a number of centres conducting RP in the home with success rates in the order of 85% within their clinical services. This should be acknowledged and might affect analysis and interpretation of the paediatric data	We have noted this point in the report.
British Paediatric Respiratory Society	20	256	5.11.1	"The simplest pathway would be in identification of OSAHS in children without comorbidities, where the use of novel devices as an alternative to pulse oximetry at home could be explored. However, such children appear to be a minority of those being investigated for OSAHS." I am not sure these children are a minority of those being investigated for OSAHS?	The statement that children without a co-morbidity are a minority of children being tested for OSA was based on estimates from two of the clinical experts advising the EAG. However, as the experts cited a wide range of estimates, and we have not been able to find a good published source, we have deleted this point from the text.
British Paediatric Respiratory Society	21	256	5.11.1	"For children with mild OSA, due to uncertainty in effectiveness of treatments, it is often unclear how to proceed." Maybe this should be rephrased as being unclear how to proceed sounds like no one knows what to do.	We have rephrased this sentence as: 'For children with mild OSA, there is uncertainty over the effectiveness of treatments.'
British Paediatric Respiratory Society	22	296	Acupebble SA100	[redacted]	[redacted]
ZOLL Itamar	23	37	1.3.5/6	Changes required in the description of both WatchPAT devices: 1. The description of both devices is missing a key feature in the usage of the WatchPAT technology – the option to view the study raw data (signals), manually edit the analysis to adjust the score of the test. To promote the utilization of manual editing, we supported the development of scientifically validated guidelines for manual editing of WatchPAT tests. As we will demonstrate in the next comment, this feature and its utilization proved to be an essential factor behind the wide acceptance of WatchPAT devices amongst sleep specialists in the UK, and around the world. 2. The description also mentions that the manufacturer states that WatchPAT should not be used in patients on medication that include alpha blockers, or people with sustained non-sinus cardiac arrhythmia. This is not accurate. In page 2 of the WatchPAT Operational Manual (https://www.itamar-medical.com/wp-content/uploads/2022/02/Operation-Manual-WP300-Europe-OM2196380.pdf) these two topics are referred to under the title **Precautions.** With respect to medication, the manual states that three hours washout before taking the sleep study is sufficient.	1. The capability to access raw data and manual editing appeared in section 1.3.6 (WatchPAT ONE). This has now also been added to section 1.3.5 (WatchPAT 300). 2. What we have written in the report corresponds with section 1.3 (Precautions) on page 2 of the WatchPAT manual. Whilst the sub-heading is indeed 'Precautions', directly below this heading it says "should not be used with…" The text from the manual is reproduced below (we have added bold type for emphasis). " The WatchPAT™300 should **not be used** in the following cases: 1. Use of one of the following medications: alpha blockers, short acting nitrates (less than 3 hours before the study). 2. Permanent pacemaker: atrial pacing or VVI without sinus rhythm. 3. Sustained* non-sinus cardiac arrhythmias. * In the setting of sustained arrhythmia the WatchPAT's automated algorithm might exclude some periods of time, resulting in a reduced valid sleep time. A minimum valid sleep time of 90 minutes is required for an automated report generation.
ZOLL Itamar	24	117	5.7.3	WatchPAT accuracy: This section assesses the accuracy of the WatchPAT 300 and WatchPAT One, based on data form the Tauman et al study that was published in 2020. The Tauman study was performed with and older version of the WatchPAT, the WatchPAT 200U. We noticed that the authors repeatedly mentioned the lack of WatchPAT 300 and WatchPAT One validation studies. We wish to advise that both WatchPAT 300 and WatchPAT One are using identical algorithm to the one	Thank you for confirming the identical features of WatchPAT 300 and WatchPAT One and WatchPAT 200U. This supports our inclusion of WatchPAT 200U as supporting evidence where data for WatchPAT 300 and WatchPAT were lacking.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				used in WatchPAT 200U and in the Tauman study. In addition, both devices produce identical signals to that of WatchPAT 200U. This similarity allowed us to get FDA and CE for the new devices based on technological continuity.	
ZOLL Itamar	25	117	5.7.3	WatchPAT Specificity: Based on the Tauman study the authors of the assessment calculated the specificity of WatchPAT in AHI≥5 at 0.25. We wish to challenge this analysis and suggest that in practice, the specificity of WatchPAT in AHI≥5 is much higher and should be estimated at around 0.60. Our claim is based on the Zahng et. al JCSM 2020 study (https://jcsm.aasm.org/doi/10.5664/jcsm.8278), where a team of researchers from Johns Hopkins University developed scientifically validated guidelines to perform manual editing of the WatchPAT automated scoring, when necessary. They found that by applying the manual editing, especially at low AHI, the accuracy of the result is improved significantly. In the supplementary information for the study, in table S2 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7161441/bin/jcsm.16.4.563.SD1.pdf) it is shown that by performing the manual editing, the specificity at AHI≥5 is improved from 0.29 to 0.60. Following the completion of that study and the introduction of the guidelines, the Johns Hopkins research team developed a training program designed to teach WatchPAT users to manually review and edit the WatchPAT automated sleep reports. ZOLL Itamar adopted this program and started promoting it amongst WatchPAT users all over the world. We can confirm that in the past four years all WatchPAT users in the UK have been trained to perform manual editing, and to the best of our knowledge, most WatchPAT users in the UK routinely perform manual editing when necessary.	Thank you for suggesting the Zhang study as an alternative source of evidence for specificity of WatchPAT and also for the information about the manual editing training progamme. The Zhang study was identified by our literature searches and screened for possible inclusion in the systematic review. However, the version of WatchPAT (200) is outside the scope of this NICE diagnostic assessment. As we explain later, we only permitted evidence for older versions of WatchPAT in cases where there was a lack of evidence for WatchPAT 300/ONE. In such cases we used supporting evidence from WatchPAT 200U studies, but did not consider that versions older than 200U (i.e. 200, 100) to be as comparable with the current models. Furthermore, we do not believe it is appropriate to use accuracy estimates from different studies in this way. Due to uncertainty in the estimation of the specificity value at the cut-off of AHI ≥5 from Tauman, the upper 95%CI value is 0.806. In our one-way sensitivity analyses (Section 5.10.3 and Appendix 9b), the value of 0.806 is used (which we note is higher than the 0.6 noted here). At this high value, the results indicate that WatchPAT 300 and WatchPAT ONE would be considered cost-effective at both WTP thresholds.
ZOLL Itamar	26	121	5.7.5	WatchPAT failure rate: This section estimates the potential failure rate of the WatchPAT 300 and WatchPAT One. It does so based on the Ioachimescu et al, JCSM 2020 study, which was performed with the WatchPAT200U devices. The study states that 5.84% (31/531) of the studies were excluded (failed). However, it attributes the cause of the failed studies to PAT/Oximetry signals. While the origin of the PAT signal could only be related to the WatchPAT, a failed oximetry signal could also be related to the PSG. Not only that the study doesn't provide details of the number of fails that are associated with each signal, it also doesn't say if the oximetry issues are related to the WatchPAT. In addition, this study was performed simultaneously with a full PSG at a sleep laboratory and not at the patients' homes. Therefore, we believe that this study cannot serve as a good reference to the expected failure rate of the WatchPAT devices. We would like to suggest an alternative study as reference for failure rate. In our opinion this study represents much accurately the true failure rate of WatchPAT200U: The Nobuaki Tanaka et al, Circulation Journal 2021 study (https://www.jstage.jst.go.jp/article/circj/85/3/85_CJ-20-0782/_pdf/-char/en), included 776 patients that were tested with WatchPAT200U at home. Out of the 776 tests, only 2 tests failed, bringing the success rate of the device to 99.7%. Considering that the Tanaka study was performed at the patients' homes, we	We have received confirmation from the study lead author that the version of WatchPAT used was **200** and **not 200U**. They also confirmed that all of the test failures were related to WatchPAT (31 studies with poor PAT or SpO2 signals out of 531 total studies) and there were no test failures related to PSG. We have updated the evidence informing the failure rates for WatchPAT 300 and WatchPAT ONE, using the Mueller 2022 study (please see Table 21 of the Erratum). This study was overlooked as a source of evidence on failure rates for WatchPAT in our original report. Mueller et al's evaluation of WatchPAT 300 is a more recent study than Ioachimescu, albeit with a smaller sample size. The failure rate from Mueller is 3.28% (2/61). Thank you for recommending Tanaka 2021, however due to it being a retrospective study it is not clear how many sleep studies individuals had before they had a successful sleep study.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				believe that it should be considered as much better evidence of the true success rate of the WatchPAT devices.	
Sunrise	27	5 10 78 120 121	Abstract Scientific summary 4.5.6 5.7.5 5.7.5	We recommend using the 5.88% failure rate (1/17) reported in Alsaif (2022) for the Sunrise device, as detailed in your literature review focusing on a home environment. This is in preference to the rate reported in Kelly (2022). Our suggestion stems from the fact that the data in Alsaif (2022) is more recent, collected in 2022 and 2023, compared to the 2020 data from Kelly (2022). The rationale for this recommendation is twofold. Firstly, it takes into account the advancements and updates made to the Sunrise device and its user instructions since 2020, including improvements in Bluetooth connection stability. Secondly, our recommendation aligns with the observations in the report: *"It should be acknowledged that some of the factors contributing to failed tests were not anticipated by the study investigators and, with the benefit of hindsight, were preventable. The expectation is that the learning from these instances will have prompted necessary changes to testing protocols, device features and user instructions to avoid similar failures occurring again. If this is the case then novel device failure rates in clinical practice would be lower than those reported in the studies, all other factors being equal."* Given these developments, we believe that the failure rate from Alsaif (2022) more accurately reflects the current performance of the Sunrise device.	Since we submitted our report we have received a confidential manuscript submitted for publication reporting the final results of the Alsaif et al study. We have updated our report with the final results, superseding the interim results previously presented in a conference abstract. The failure rate for Sunrise reported from the full analysis is [redacted] Both studies (Alsaif and Kelly 2022) were conducted in the home [redacted] Due to a preference from NICE for keeping AIC data out of the base case if possible, and due to [redacted] we have decided to use the failure rate from Kelly in our base case. However, we have conducted an additional scenario analysis using the failure rate for Sunrise from Alsaif, please see Section 5.10.2.
Sunrise	28	33	1.3	Could you kindly make the following update for accuracy and consistency? ● Replace 'Sunrise (Hello Sunrise)' with 'Sunrise (Sunrise)' to correct the manufacturer's name in parentheses Thank you for your attention to these details.	All instances of 'Sunrise (Hello Sunrise)' within the report have been replaced with 'Sunrise (Sunrise)'
Sunrise	29	36	1.3.4	Could you kindly make the following updates for accuracy and consistency? ● Replace 'Sunrise (Hello Sunrise)' with 'Sunrise (Sunrise)' to correct the manufacturer's name in parentheses ● Replace 'central AHI (cAHI)' with 'central AHI (CAHI)' for consistency ● Replace 'respiratory effort related arousals (RERA) index' with 'respiratory effort related arousal (RERA) index' for consistency ● Replace 'awakenings and arousal index' with 'awakening and arousal index' for consistency ● Remove the statement 'The manufacturer advises against using the device in patients with implantable devices, e.g. pacemakers.' as this information is no longer up to date	All updates listed have been changed In regard to the final bullet point of the comment (i.e. "Remove the statement 'The manufacturer advises against using the device in patients with implantable devices, e.g. pacemakers.' as this information is no longer up to date") we have removed this sentence from the report as requested. However, we do note that page 4 of the Sunrise HCP user manual and page 5 of the Sunrise patient user manual, which were supplied by Sunrise to NICE for the purpose of this assessment, do state "Do not use the device with a pacemaker or similar implanted device since it could impair its functioning"

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Thank you for your attention to these details.	
Sunrise	30	41	2.2	Could you kindly make the following updates for accuracy and consistency? ● Replace 'Sunrise (Hello Sunrise)' with 'Sunrise (Sunrise)' to correct the manufacturer's name in parentheses ● Replace 'Sunrise (from three to 18 years old)' with 'Sunrise (for patients over three years of age) to correct an error in the age specification Thank you for your attention to these details.	All instances of 'Sunrise (Hello Sunrise)' within the report have been replaced with 'Sunrise (Sunrise)' We have amended the age specification as requested (and also in section 1.3.4 for consistency).
Sunrise	31	46	3.2.2	Could you kindly make the following update for accuracy and consistency? ● Replace 'Sunrise (Hello Sunrise)' with 'Sunrise (Sunrise)' to correct the manufacturer's name in parentheses Thank you for your attention to these details.	All instances of 'Sunrise (Hello Sunrise)' within the report have been replaced with 'Sunrise (Sunrise)'
Sunrise	32	69 70	4.4 4.4	Would it be possible to review Table 8 on page 70 and the associated discussion on page 69 in light of the details provided in comment no. 18? Your attention to this matter and consideration of these points would be highly appreciated. Thank you.	Table 8 has been reviewed and revised. Please see comment 44 for further details.
Sunrise	33	74	4.5.1.1	It has been observed that in Table 9, the study 'Pepin et al 2020' seems to be missing. It is suggested that this may be an unintended omission.	Thank you for bringing this to our attention. Pepin et al 2020 has been added to Table 9.
Sunrise	34	117	5.7.3	Could you kindly make the following updates for accuracy and consistency? ● Replace Sunrise specificity (95% CI) for AHI > or = 5 by 0.94 (0.91, 0.97) to correct typo error ● Replace Sunrise sensitivity (95% CI) for AHI > or = 15 by 0.92 (0.90, 0.94) to correct typo error Thank you for your attention to these details.	Thank you for highlighting these minor errors in Table 19 of the EAR (now table 18). This has been corrected in the Erratum and in the model.
Sunrise	35	115 116 117	5.7.3 5.7.3 5.7.3	Another potential approach for the base case analysis is to combine data from the Pepin 2020 and Kelly 2022 studies. By aggregating the 2x2 confusion matrices (with the number of patients) for both cut-offs, a more comprehensive assessment of performance can be achieved (even if this combination is primarily influenced by the Pepin 2020 study, due to its significantly larger patient cohort). This method presents the following diagnostic accuracy:	Given the differences between the studies, including that Pepin was conducted in the clinic, while Kelly was conducted in the home, we prefer to continue to treat these data separately, and to report the studies separately.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

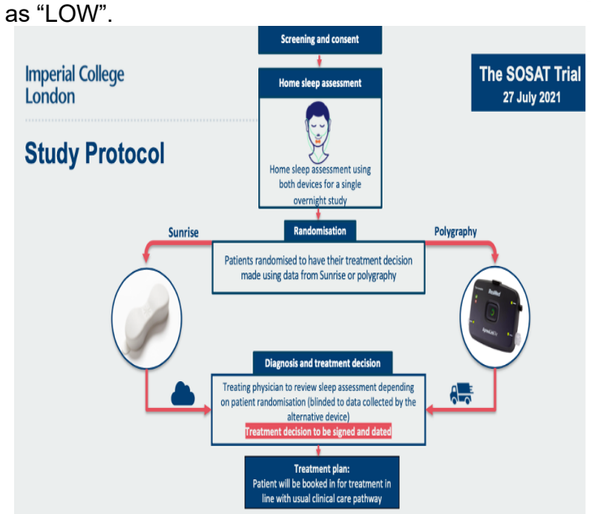
Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				• Sensitivity at low cut-off: 0.907 (95% CI: 0.881 - 0.929) • Specificity at low cut-off: 0.942 (95% CI: 0.895 - 0.979) • Sensitivity at high cut-off: 0.925 (95% CI: 0.899 - 0.950) • Specificity at high cut-off: 0.833 (95% CI: 0.797 - 0.869) This analysis encompasses a total of 407 patients, combining 376 patients from Pepin 2020 and 31 patients from Kelly 2022. The aggregation of data from these two studies not only enhances the statistical power through a larger sample size but also provides a broader perspective on the diagnostic performance across different patient populations.	In our model we use the larger study, conducted in the clinic (Pepin 2020) in the base case, and use data from Sunrise in the home (Kelly 2022) in a scenario analysis. Moreover, the aggregation approach described to obtain these results is not ideal as it assumes all participants as if they were from the same study. Bi-variate meta-analysis methods are preferred, yet require more than 2 studies.
Sunrise	36	120	5.7.5	Could you kindly make the following update for accuracy and consistency? • Replace ‘Kelly et al 2020 study’ with ‘Kelly et al 2022 study’ to correct a typo error Thank you for your attention to these details.	This has been updated throughout the report.
Sunrise	37	142 143 148 318	5.7.12 5.7.12 5.7.12 Appendix 8	Could you kindly make the following updates for accuracy and consistency? • Replace ‘However, the company advised that the cost per sleep study would depend on the volume of devices ordered: 1-9 = £73, 10-49 = £68, 50-99 = £65, 100+ = £62.’ with ‘However, the company advised that the cost per sleep study would depend on the volume of devices ordered: 5-9 = £73, 10-49 = £68, 50-99 = £65, 100+ = £62.’ to correct a typo error (specifically changing '1-9' to '5-9') • Replace ‘In base case analyses, we assume that should a sleep study fail, the full cost of a new device would be incurred to undertake a second sleep study.’ with ‘In base case analyses, we assume that there is no additional sleep study or device cost to the NHS for a failed sleep study with the Sunrise device.’ to correct erroneous information. Indeed, in the event of a device failure, the manufacturer replaces the device at no additional cost to the NHS. Thank you for your attention to these details.	This typo has been corrected in the report. In response to an EAG request for information, the company reported that “[redacted]” Due to a preference from NICE to avoid CIC data from the base case analysis, where possible, we assume that costs of devices for any repeat tests are included. However, we report on a scenario analysis [redacted]
Sunrise	38	168	5.10.1	Regarding Figure 8, could you provide some insight into why the Sunrise and Brizzy devices show a bell curve pattern, whereas the other devices display a sigmoid curve? What factors contribute to this difference in curve shapes? Thank you.	At each level of willingness to pay, the CEAC shows the proportion of PSA iterations in which each device is the most cost-effective option. As the willingness to pay increases, devices that are expected to have a higher projected QALY gain become increasingly likely to be the most cost-effective option.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					Since the model results indicated that the WatchPAT devices were associated with the greater QALY gains, these devices had increasing probabilities of the being the most cost-effective as WTP increased. Hence, Brizzy and Sunrise had lower probabilities of being cost-effective at this higher WTP thresholds. Note that all analyses have been re-run for the Erratum.
Sunrise	39	169	5.10.2	It has been observed that in Table 39, the scenario analysis 'Kelly 2022 accuracy data for Sunrise' seems to be missing. It is suggested that this may be an unintended omission. In the section “Scenario analyses for diagnostic accuracy estimates,” discussing the results of this scenario analysis would be beneficial. Specifically, it could be highlighted that, under this scenario, Sunrise is estimated to be more costly but also to provide more QALYs compared to oximetry, with a pairwise ICER of £6,869. Furthermore, when compared to respiratory polygraphy, Sunrise is estimated to be dominant, offering better outcomes at a lower cost. It is also noteworthy that Sunrise demonstrates positive incremental net benefits at both £20,000 and £30,000 per QALY gained thresholds in comparison to oximetry and respiratory polygraphy.	Table 39 (now table 38) has been updated. In the updated analyses, consideration of accuracy data from Kelly 2022 for Sunrise, as opposed to Pepin 2020, leads to a INMB of £451 (vs base case of £127). As the overall conclusions for Sunrise do not change, we keep these results in the Appendix (please see Appendix 9a). However, we note the limitations of Kelly 2022 as a source of accuracy data, especially given the relatively small sample of participants (please see Section 5.7.3).
Sunrise	40	191 196	Discussion Conclusions	We have noticed that the contraindications of different devices were not considered in the economic model. We think it is important that the discussion and conclusions sections bring up this limitation. Specifically, they should highlight that some of the evaluated devices may not be suitable for certain patient groups. This is especially relevant for people with OSA, who are typically at a higher risk of cardiovascular events compared to the general population. And it's worth noting that costs related to cardiovascular events account for over 60% of the total costs for all devices. For instance, devices based on Peripheral Arterial Tonometry (PAT) are not recommended for use in individuals on alpha-blocker medications. This is a notable consideration since many patients with cardiovascular diseases are prescribed these medications. Relevant details can be found on pages 36-37 of the report. Including this information in the discussion and in the conclusions would provide a more comprehensive understanding of the economic model's context and limitations.	Due to the variability of comorbidities across devices, and a lack of evidence on the likely size of populations having to be excluded from use of the difference devices, we did not include the impact of comorbidities in the model. However, note that we have conducted scenario analyses for higher/lower CVD risk cohorts (EAR 5.4.1, 5.7.1 and Table 77). These give similar relative and absolute INMB values for the different devices.
Sunrise	41	191 196	Discussions Conclusions	The report acknowledges the emergence of portable testing devices with novel features for diagnosing sleep disordered breathing, noting their advancements in	Thank you for suggesting aspects of portable testing which should be considered. There is limited available evidence for many of these to inform modelling, but we do discuss their potential

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				technology aimed at enhancing performance, convenience, and acceptability for home testing of OSAHS. While we understand the challenges in quantifying all these aspects, and their exclusion from the model, it is important that the discussion and conclusions sections address this limitation. These aspects are very likely to influence both cost and QALYs. Examples of some key aspects not fully accounted for in the model include: ● Time to diagnosis or treatment initiation ● Advantages of disposable devices over reusable ones (e.g. faster diagnosis and elimination of bottlenecks due to device retrieval) ● Healthcare resource use and costs, including expenses for device investment and replacement, patient backlog management, staff capacity allocation, and volume of managed patients ● Patient-reported outcomes and care experience encompassing ease of use, acceptability, comfort, satisfaction, data representativeness, anxiety, travel costs and time, need for time off work, and impact on earnings ● Accessibility to home testing and number of patients lost to follow-up or unwilling to undergo prescribed tests, influenced by device usability ● Treatment compliance and effectiveness, potentially enhanced by user-friendly devices and faster time to diagnosis or treatment initiation Addressing these factors in the report would provide a more comprehensive understanding of the full spectrum of implications these novel devices have on healthcare delivery and patient outcomes.	implications where relevant (see sections 6.2.2, 6.3 and 6.4). When making their recommendations for guidance, the diagnostic advisory committee will take into consideration any factors affecting costs and QALYs which haven't been incorporated into the cost effectiveness calculations.
Sunrise	42	191 196	Discussions Conclusions	While the report concludes that WatchPAT devices produce more QALYs due to their higher sensitivity compared to home respiratory polygraphy and other devices, we believe that a closer examination reveals nuances worth considering. For instance, the sensitivity for the low cut-off between home respiratory polygraphy (RP) and WatchPAT is very similar (0.958 for WatchPAT vs. 0.953 for home RP), and for the high cut-off, the sensitivity is actually lower for WatchPAT (0.877 vs. 0.930 for home RP). This suggests that the increased QALYs from WatchPAT might be attributed to its lower specificity at both cut-offs, potentially leading more OSA patients, particularly those in the mild severity, to effective treatment and thereby generating more QALYs. This aspect could indicate a limitation in the model or the assumptions made, as it seems counterintuitive that devices with lower diagnostic accuracy yield higher QALYs and cost-effectiveness. This point warrants further investigation and, at	Thank you for highlighting these points in your comments. The impacts of the specificities at the low and high diagnostic cut-offs should be thought about separately: <u>Specificity at low (AHI = 5) diagnostic cut-off (distinguishing no OSA from OSA)</u> As this comment highlights, changes to these specificities should have no impact on the QALYs (but they were). Looking at this more closely, we identified an error in the Markov model for people without OSA who receive CPAP (Model 3a). The utilities were incorrectly specified (females were given the male utility values). This meant that those misdiagnosed as having OSA and receiving CPAP had increased QALYs. Therefore, misdiagnosing people with no OSA was beneficial, i.e. a poor specificity at AHI of 5 was leading to more people receiving CPAP therefore having more QALYs, than good specificity at AHI = 5. After correcting this error, changing the lower cut-off specificities has no impact on the QALYs.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				the very least, should be mentioned in the discussion and conclusions if no resolution is found. This observation aligns with the scenario analysis in Table 45, where it is assumed that 75% of people diagnosed with mild OSAHS are treated with CPAP. In this scenario, all devices are associated with greater total QALYs due to more people receiving effective treatment, and compared to home RP, all devices are estimated to be cost-effective at a WTP threshold of £20,000 per QALY gained. Further support for this observation comes from the sensitivity analyses, where the most impactful parameters were utilities for mild and moderate OSAHS and the sensitivity and specificity estimates for the novel devices and comparators. The sensitivity analysis reveals that specificity at the high diagnostic cut-off for novel devices and home RP also significantly influences the results (as shown in Figures 15 to 17). The report may benefit from more clearly highlighting, perhaps initially, the very small, highly uncertain, and statistically non-significant estimated QALY difference between home respiratory polygraphy and novel devices. Interpreting the results requires a careful approach. This perspective is supported by multiple points: ● The probabilistic scenario analysis, where broad and overlapping confidence intervals indicate significant uncertainty in incremental costs and QALYs for each novel device compared to oximetry, and particularly when compared to home respiratory polygraphy. ● This is further highlighted in the scenario analyses where the sensitivity and specificity estimates from Xu et al. are replaced with those from Pereira et al., 2013, and those used in the NG202 economic analysis. ● The sensitivity analyses also clearly illustrate the uncertainty related to the parameters and methodological choices used in the model.	<u>Specificity at high (AHI = 15) diagnostic cut-off (distinguishing moderate-severe OSA)</u> Changing these specificity estimates does impact on QALYs, but this is not an error. The impact relates to those with mild OSA. Those correctly identified as having mild OSA may receive conservative management, CPAP or MAD. Those with mild OSA misdiagnosed with moderate or severe are only treated with CPAP and MAD. Note that only CPAP and MAD impact on utility in the truly mild group. Please see Table 29 in Section 5.7.11 for details of treatment impacts. With a higher specificity, the number of people with mild OSA misdiagnosed as having mod/severe OSA will decrease. Which means that fewer people with mild OSA have CPAP or MAD, which reduces their QALYs (and costs). As this comment highlights, this pattern is also reflected in the scenario analysis where 75% of those with mild receive CPAP (as opposed to 20% in the base case analysis). We have already emphasised uncertainty over the cost-effectiveness results in multiple places through the report - Abstract, Scientific summary, Economic results (5.10.1), Discussion (6.2, 6.3, 6.4) and Conclusions (7.1). We have further highlighted this point in discussion of the updated results of the scenario analyses (5.10.2) and one-way sensitivity analyses (Section 5.10.3).
Sunrise	43	248	Appendix 2	Could you kindly make the following update for accuracy and consistency? ● Replace 'Sunrise (Hello Sunrise)' with 'Sunrise (Sunrise)' to correct the manufacturer’s name in parentheses Thank you for your attention to these details.	All instances of 'Sunrise (Hello Sunrise)' within the report have been replaced with 'Sunrise (Sunrise)'

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Sunrise	44	273	Appendix 5	Could you kindly make the following update for accuracy and consistency? ● Replace 'Alsaif et al 2020' with 'Alsaif et al 2022' to correct a typo error Thank you for your attention to these details. Alsaif et al., 2022 Regarding index test - risk of bias - signalling question 2 and judgement: The data collection and interpretation in the SOSAT study accurately reflect real-life conditions. For scoring OSA severity, the Sunrise device utilises pre-specified thresholds established in Pepin et al., 2020. These thresholds are clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included. Based on this, we suggest changing the assessment for question 2 to "Yes" and the associated risk being categorised as "LOW". Regarding reference standard - signalling question 2 and judgement: We confirm that in the study, the treating physician was blinded to the data collected by the alternative device, as indicated in the screenshot below from the study protocol presentation in July 2021. Based on this, we suggest changing the assessment for question 2 to "Yes" and the associated risk being categorised as "LOW".  Kelly et al., 2022 Regarding patient selection - concerns regarding applicability - judgement: Based on the positive assessment performed, we think that the concern should be "LOW" instead of "UNCLEAR". Please note that the related Table 8 page 70	We have replaced the study name with "Alsaif et al., 2023" as the primary paper for this study is now an unpublished manuscript first authored by Dr Alsaif and dated 2023. Signalling question 2 asks "If a threshold was used, was it pre-specified? Alsaif et al (2023) does not explicitly state which diagnostic thresholds were used. Whilst it is recommended that the thresholds established by Pepin et al 2020 are followed when using the Sunrise device in practice, it doesn't necessarily mean these thresholds were used by Alsaif. For a LOW risk of bias judgement to be given the study would have stated in advance (e.g. in the study protocol/clinical trials record) which thresholds were to be used, and in the study final report whether these were used. As this information is not available for this study we have judged the risk of bias to be UNCLEAR on this domain. Regarding reference standard - signalling question 2 and judgement we have amended this to "Yes" and the judgement of this domain to "LOW" Regarding the patient selection - concerns regarding applicability – judgement, we have changed this from "UNCLEAR" to "LOW" and updated Table 8.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				shows as expected a green emoji under applicability concerns - patient selection. Regarding index test - risk of bias - judgement: The data collection and interpretation in this study accurately reflect real-life conditions. For scoring OSA severity, the Sunrise device utilises pre-specified thresholds established in Pepin et al., 2020. These thresholds are clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included. Based on this, we suggest changing the assessment for the associated risk being categorised as “LOW”. Regarding index test - concerns regarding applicability - judgement: The information presented in Table 8 on page 70 appears inconsistent with the assessment indicating that the concern is’"LOW'. Could you please update the table by replacing the question mark with a green emoji to reflect this assessment? Thank you Martinot et al., 2022 (child) Regarding index test - risk of bias - signalling question 2 - judgment: This study established the pre-specified thresholds for OSA severity scoring, which are included in the Sunrise report (for child). These thresholds are clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included. Based on this, we suggest changing the assessment for question 2 to “Yes” and the associated risk being categorised as “LOW”. Pepin et al., 2020 Regarding index test - risk of bias - judgment: This study established the pre-specified thresholds for OSA severity scoring, which are included in the Sunrise report (for adult). These thresholds are clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included. Based on this, we suggest changing the assessment for the associated risk being categorised as “LOW”. Regarding reference standard - risk of bias - signalling question 2 - judgment: We do confirm that PSG scorers were totally blind to Sunrise results to avoid any bias in the interpretation and comparison of results. This information can be found in the “Methods” section of the paper: “In this prospective, diagnostic study of adult patients who were referred for a single overnight in-laboratory PSG, the PSG was used as the reference method and, with blinding, was compared with simultaneous MM recordings using the Sunrise system“. The first sentence of the discussion also states the following: “In a large, prospective cohort of patients with and without OSA, we evaluated the agreement between MM-derived Sr-RDI and blindly scored PSG-RDI (Figure 2B).” Based on this, we suggest changing the assessment for question 2 to “Yes” and the associated risk being categorised as “LOW”.	Regarding the index test - risk of bias – judgement, the same argument above is made in relation to Kelly et al, 2022 study. It is implied by the company that because the Sunrise device utilises pre-specified thresholds established in Pepin et al., 2020 then these would have been applied by Kelly et al 2022. In actuality, Kelly et al did a post hoc analysis to optimise the cut-off points of MM-ORDI for diagnostic decisions, compared with the criterion standard cut-off values of obstructive PSG-ORDI. For this reason we judge both the risk of bias and concerns regarding applicability for this domain as ” HIGH” (applicability changed from ‘unclear’ to high). Martinot et al., 2022 also utilised post-hoc analysis to optimise cut-off points, therefore the QUADAS index test - risk of bias - signalling question 2 – judgment is “No” and the associated risk for this domain is “HIGH” (changed from ‘unclear). Pepin et al., 2020 utilised post-hoc analysis to optimise cut-off points therefore the index test - risk of bias - signalling question 2 – judgment is “No” and the associated risk for this domain is “HIGH” (changed from ‘unclear’). Regarding the reference standard - risk of bias - signalling question 2, we have changed the assessment to “Yes” and the associated risk to “LOW” (changed from ‘unclear’).

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Based on the above comments, could you please update the related table 8 on page 70 accordingly? Your attention to this matter would be greatly appreciated. Thank you. Table 8 Overview of QUADAS-2 assessments Study | RISK OF BIAS: PATIENT SELECTION, INDEX TEST, REFERENCE STANDARD, FLOW AND TIMING | APPLICABILITY CONCERNS: PATIENT SELECTION, INDEX TEST, REFERENCE STANDARD Alsaif 2022 ? ☹ ☹ ? ☺ ☺ ☺ Devani 2021 ☺ ☺ ☺ ☺ ☺ ☺ ☺ Kelly 2022 ☺ ☹ ☺ ☺ ☺ ☹ ☺ Lyne 2023 ? ☺ ☺ ☺ ☺ ? ☺ Martinot 2015 (child) ☺ ? ☺ ☺ ☺ ☺ ? Martinot 2017 ☺ ☺ ☺ ☺ ☺ ? ☺ Martinot 2022 (child) ☺ ☹ ☺ ☺ ☺ ? ☺ Massie 2021 ☺ ☺ ☺ ☺ ☺ ? ☺ Mueller 2022 ☺ ? ? ☺ ☺ ☺ ☺ NCT04031950 (child) ☺ ☺ ☺ ☺ ☺ ? ☺ Pepin 2020 ☺ ☹ ☺ ☹ ☺ ☺ ? ☺ Pillar 2020 ☹ ☺ ☺ ☺ ? ☺ ☺ Tauman 2020 ? ☺ ☺ ☺ ☺ ☺ ☺ Van Pee 2022 ☺ ☺ ☺ ☺ ☺ ☺ ☺ ☺ Low Risk ☹ High Risk ? Unclear Risk	Table 8 has been updated to reflect the changes we have described above.
Sunrise	45	309	Appendix 6	Could you kindly make the following update for accuracy and consistency? • Replace 'Sunrise (Hello Sunrise)' with 'Sunrise (Sunrise)' to correct the manufacturer's name in parentheses Thank you for your attention to these details.	All instances of 'Sunrise (Hello Sunrise)' within the report have been replaced with 'Sunrise (Sunrise)'
Sunrise	46	322	Appendix 9	Could you kindly make the following update for accuracy and consistency? • Replace 'Kelly 2020 accuracy data for Sunrise' by 'Kelly 2022 accuracy data for Sunrise' to correct typo error Thank you for your attention to these details.	This typo has been corrected.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	47	5	Abstract, results	**The following colour coding has also been used, as much as possible, to help with the review of comments: blue relates to comments which are statistical or scientific, green relates to comments on health inequalities, red represents information commented on which is factually incorrect.** Second paragraph: Summarised statements of performance based on sensitivity and specificity are made. Sensitivity and specificity on their own should not be given to generalise performance. Their limitations as statistical metrics in isolation to validate medical devices, and specifically in OSA have been widely established (this can be elaborated upon with references). Hence, this paragraph is vague and serves no purpose. Please give more statistical metrics (PPV, NPV, LRs), confidence intervals, and refer to specific devices. They are different, have different technological characteristics and hence cannot be lumped together.	This is the abstract which summarises key aspects of the report. Due to the abstract's restrictive word limit it is not possible to go into detail on every aspect. However, further elaboration can be found in the main body of the report. Please see sections 4.5.1 and 4.8.1 for details on the diagnostic accuracy data identified in our review.
Acurable Limited	48	5	Abstract, results	Third paragraph: It is stated that limited data is available on the time to interpret device outputs and patients using novel devices. This is a simplistic generalisation. Devices have been given regulatory approvals for different outputs and intended uses. Note that regulatory approval for different intended use is only granted on the basis of evidence provided to regulatory bodies, which goes well beyond what is in the public domain. Hence, for example, it would be possible to infer the time it can take in the best case scenario for a device that has been granted regulatory approval for fully automated diagnosis, and this would be very different to a product in which a clinician has to go through raw signals or partly score. It is also different if the outputs are immediately available to clinicians after the night's test, or if this depends on patients going to the clinic. The time it takes patients to return systems could easily be researched since there is a large amount of data from other devices. Some devices have this functionality whereas others do not. However, trying to determine the time to diagnosis on the basis of how long it takes clinicians to provide patients with results after they have the data (ie what is implied to not be available) is not the correct approach, since this depends on factors that are in many cases transitory, heavily healthcare centre dependent, and ultimately have nothing to do with the technologies. It is also not the same if the technologies already generate an immediately available clinician/patient report that has been approved by regulators (and hence clinicians only have to glance at it), than if that report has to be created by clinicians. Please be more specific about what is meant by "limited data" here.	Please see section 4.5.4 for data on time to interpret device outputs and section 5.7.12 for information that we used to cost related healthcare professional time.
Acurable Limited	49	5	Abstract, results	Third paragraph: In the same paragraph it is also stated that there is limited patient experience of using these novel devices. Again, this is a generalisation. There is evidence for some devices, and in some cases this evidence is more than there has ever been with other devices in the market (for other conditions), since usability studies have been carried out. There has also been evidence provided on their use in vulnerable populations for at least one device that has	Please see sections 4.7.2, 4.7.3 and 4.10 for available data on ease of use, acceptability and patient and carer experience.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				not been acknowledged. Furthermore, and regardless of this, it would be very clear to anyone that any of these devices is significantly easier for patients to use than RP. Hence making the same assumption in terms of diagnostic time and usability for both is simply the wrong assumption which will unavoidably bias the model in favour of RP. Please be more specific about the evidence there is.	
Acurable Limited	50	5	Abstract, results	Last paragraph: The authors present the overall results of their analysis in terms of QALYs. However these cannot be correct because the assumptions and the quantitative data used for the models were incorrect. **Hence this analysis and all the text that results from it needs to be corrected.** Further explanations can be found in subsequent comments.	The model has been updated as is explained in subsequent responses, in particular responses to comments 57 (regarding updated accuracy estimates for AcuPebble) and 167 (regarding utility estimates).
Acurable Limited	51	5	Limitations	The authors state that there is a high uncertainty over the cost effectiveness results, and the results are sensitive to different sources of data on the accuracy of the novel devices and respiratory polygraphy. This paragraph is very vague and hence serves no purpose beyond stating an opinion. This opinion may also lead to incorrect understanding, in the absence of the main reason for uncertainty: the methodology followed made up assumptions (which as a note were not correct in many cases, so those need to be corrected before addressing the descriptive part of the limitations). Any other uncertainty can be quantified with statistical metrics on the basis of real-world evidence that exists. In the cases where it couldn't, that could be highlighted point by point, but not with a general statement since the technologies are all different and "one rule does not apply to all". **The authors need to state that there is uncertainty caused by their model having limitations since the latter is based on simplifications and the very basic assumption of all devices being similar in their outputs, regulated intended use and inherent limitations. Hence, a model should be created that accommodates for the differences between the devices, establishing assumptions that are device-specific and not extrapolating from one device to another. Alternatively, and preferably, the EAG should treat devices differently and not try to simplify by making assumptions for devices on the basis of what others do, that go against the information given by the manufacturers.**	Please see sections 5.10.2, 5.10.3 and Appendix 9 for details on scenario and sensitivity analyses that we conducted to explore the impact of uncertainties of model parameters, data sources and parameters. We also discuss key limitations and uncertainties in sections 6.2 and 6.3.
Acurable Limited	52	6	Conclusions	These conclusions, which imply that oximetry is an option, **are potentially indirectly promoting health inequalities** in terms of patients' safety (see all later comments related to implying oximetry is an alternative). Also, the authors say that it is difficult to assess cost-effectiveness. The main reason why this is, is because the assumptions are simplistic and their methodology flawed. See the rest of the comments, but just an illustrative example (more elaborated upon in the review of the rest of the document) is that it is assumed that novel technologies would not enable faster diagnosis, with the argument that diagnosis is ultimately dependent on the time clinicians take to talk to the patients. This ignores the fact that with some of the new technologies scoring and creating reports would not be necessary and results are instantaneous (for some), including automated diagnosis etc. If one assumes that none of these are advantages, and one doesn't take into account the patient experience, the	*Potential impacts on inequalities* Thank you for raising this important issue. We appreciate that use of oximetry and other light-based methods of assessment does have a potential impact on health inequalities. We understand that NICE included home oximetry in the scope as an alternative comparator to home respiratory polygraphy because of its widespread use in practice. The scope does however note a potential equality issue because technologies that use a light based assessment (PPG sensors and/or oximetry) may overestimate oxygen in the blood for people with darker skin tones. For this reason, the NICE scope requested subgroup analysis for people from black, Asian and minority ethnic

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				outputs of the model will clearly show there is very little advantage to using them. **However, this is the wrong assumption. This assumption must be corrected in the model, as well as all the descriptive text resulting from it.** Also, **if this report continues as is, the conclusions must say that the outcomes are questionable for women and individuals with dark skin. Anything else implies bias** (this can be elaborated upon, although it is addressed more in detail in other comments).	backgrounds. In addition to oximetry, the WatchPAT and NightOwl devices use a light based (PPG) technology. Other included novel devices can be used with compatible oximeter equipment. The inclusion of oximetry as a comparator means that the committee can consider if any of the novel technologies being assessed could reduce health inequalities compared to oximetry. NICE is committed to promoting equality of opportunity and eliminating unlawful discrimination, and it is usual practice for NICE and the advisory committee to consider any health inequalities issues during decision making and guidance preparation. We therefore anticipate that there will be discussion of the limitations of oximetry/PPG for people from black, Asian and ethnic backgrounds, as for other equality issues, at the forthcoming committee meeting. In line with the scope, we have included oximetry as a comparator in addition to respiratory polygraphy in our systematic review and economic evaluation. We assessed whether the studies included in the review reported on race or ethnicity for study participants. For adults, only two studies provided this information: Devani et al. 2021 for AcuPebble, and Van Pee et al. 2022 for NightOwl (US participants only). Massie et al. 2022 (NightOwl) also stated "Persons of diverse racial and ethnic backgrounds were included" but did not report the statistics. One study for children reported ethnicity: NCT04021950 (AcuPebble). We have added further detail to EAR sections 4.2.2 and 4.3.2 and Tables 60 and 61. None of the studies included in our review reported results for these (or other) subgroups. If data were available to adjust our economic results for any differences in diagnostic accuracy between subgroups of patients defined by ethnicity or other characteristics, we would do so. In the absence of such an adjustment, we emphasise that our cost-effectiveness results do not favour oximetry: the novel home-based devices are likely to provide a cost-effective alternative to oximetry (see EAR Section 5.10.1 and Tables 35 and 36). This conclusion is driven by the poor sensitivity of oximetry, which would be exacerbated for people with darker skin. Interpretation of the cost-effectiveness results for the novel devices relative to home respiratory polygraphy is harder because the estimated differences in costs and QALYs are very small and highly uncertain (as emphasised throughout our report). In this context, we suggest that health inequality impacts of alternative devices may be an important consideration. We have added further information about these issues at appropriate points in our updated report, including the Abstract, Scientific summary, characteristics of included studies (4.2 and

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					4.3), economic methods and results (5.4.1, 5.7.1, 5.10.1) and in the Discussion and Conclusions chapters (6.4 and 7.1). To emphasise the position of oximetry as an 'alternative comparator', when home respiratory polygraphy is not available, we have also changed the order of reporting of results in our report, with results for the novel devices compared to respiratory polygraphy reported first, followed by the results for novel device compared to oximetry. *Time to diagnosis or treatment* Due to a lack of evidence on differences in the time to diagnosis or treatment between use of the novel devices and current practice, we assume no difference in the model base case analysis. As part of our one-way sensitivity analyses, we investigate the impact of reducing time to diagnosis or treatment (by 1.5 months from 3 months). The tornado plots in Appendix 9b show that when time to diagnosis or treatment is assumed to be 1.5 months for the novel devices and 3 months for home RP, the INMBs for the novel devices increase by approximately £23 (at WTP of £20,000 per QALY) and £40 (at WTP of £30,000 per QALY). At the WTP threshold of £20,000 per QALY gained, this leads to Brizzy, NightOwl and WacthPAT300 having positive IMNBs (i.e. cost-effective compared to respiratory polygraphy). . A scenario analysis using unpublished time to treatment data is also conducted to assess the likely impact of this on the model (please see Section 5.10.2 and Table 77). This scenario analysis had a small impact on the model results and did not change the cost-effectiveness conclusions. Should more evidence on the impact to diagnosis and treatment time become available in the future, the model can be updated to include such evidence. Please see our response to comment 4 (from the Sleep Apnoea Trust Association) regarding the limitation of our model not including aspects of patient convenience/acceptability
Acurable Limited	53	7	Scientific summary	**This summary should also cover the recently discovered limitations of systems working on the basis of PPG** (which is the signal informing the extraction of the SPO2 values, amongst others) on those with dark skin. It should possibly also cover the fact that regulatory bodies are currently evaluating how to address the safety (in terms of wrong diagnosis) implications this has for individuals with dark skin.	Please see above comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	54	7	Scientific summary	Second paragraph: the authors say that they used an adjustment method to compare all novel devices against a common reference standard (PSG). Note	With regard to the adjustment to compare AcuPebble to PSG please see our response to comment 138 (Acurable).

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				that: t**he reference method chosen has been misunderstood, and has consequently been misused** (the wrong formula has been taken from the cited paper, and the wrong justification for using it has also been given). This is elaborated upon in the comments relevant to this in the remainder of the document.	
Acurable Limited	55	9	Methods	**In the subsection, "External Assessment Group (EAG) independent economic assessment":** The authors take oximetry as a comparator. This is **not an acceptable choice**, since it can potentially lead clinicians to decisions that could put patients with dark skin (for example black ethnic minorities) at a non-negligible unacceptable risk, based on the recent evidence that has been brought to the public light as a result of the increased mortality of dark skin people during COVID: the fact that regulations around oximeters are heavily biased towards those with white skin.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	56	9	Methods	**In the subsection, "External Assessment Group (EAG) independent economic assessment":** The model only considers cardiovascular events and road traffic accidents. However there are many other well known adverse consequences of OSA and those are not even mentioned in passing. This should have been at least discussed, referenced and also described as limitations of the model.	Please note that consequences such loss of productivity at work or school are beyond the remit of the NICE reference case. With respect to consequences such as fatigue and depression please see response to comment 167.
Acurable Limited	57	9	Methods	**In the subsection, "External Assessment Group (EAG) independent economic assessment", second paragraph:** This states that due to the fact that other devices had not been tested in the home environment (i.e. which is how they should have been tested, considering that was ultimately their intended environment of use, which is known to be harder than in hospitals), the performance quantification of the only device that had (i.e. AcuPebble) was **artificially modified to apply the same baseline.** This is not only **scientifically flawed** (for a long list of reasons that will be further discussed in other comments corresponding to the sections of the report where this was covered), but it **is also untrue to imply that the EAG did not have PSG data** for AcuPebble. That data exists, was provided to NICE (within the stipulated deadline), and appears to not be acknowledged throughout the entirety of the report. It is only in the Appendix, in the middle of a long list that most readers would miss, that the report is explained to have been excluded because the trial was ongoing. This should not be a reason, given: 1- The relevant arm of the trial, for what was needed in terms of evidence, had finished; 2- If there was any doubt about point 1 the manufacturers could have been asked, which did not happen; 3-	The reason for adjusting accuracy data from the study by Devani was due to the comparator in Devani being home RP, rather than it being PSG. It was not because all other studies were conducted in the clinic. In the Erratum we use data from the **Phase 1 Virgen Macarena trial**, to inform the base case analysis, rather than the adjusted data from Devani et al study. Therefore, all analyses for AcuPebble have been updated. For further detail on the rationale for the adjustment to compare AcuPebble to PSG please see our response to comment 138 (Acurable). Regarding the **Phase I Virgen Macarena trial** we have included this study in the Erratum, informed by the January 2024 journal publication of this study. (Sanchez Gomez et al (2024) Sanchez Gomez J, Pramono RXA, Imtiaz SA, Rodriguez-Villegas E, Valido Morales A. Validation of a Wearable Medical Device for Automatic Diagnosis of OSA against Standard PSG. J Clin Med. 2024 Jan 19;13(2):571. doi: 10.3390/jcm13020571. PMID: 38276077; PMCID: PMC10816319. The following points explain the process followed previously:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Regardless of 1 and 2, even if the trial was not complete, the CI of the results account for the patients that had been evaluated and the CI are taken into account in the model, so there would have been no scientific excuse not to include these results; 4- Even if this had not been available, the whole economic analysis could have been run with the data available, without adjustment, and then the limitations explained, rather than making up results that would bias the model. **All of the analysis for AcuPebble needs to be redone without the artificial adjustment, and every single section of the document (including this one) that makes reference to the adjustment and results from it needs to be modified once this is eliminated.** **Note that, critically, the artificial adjustment itself is made using the incorrect equation, so even if the methodology for adjustment was fair, which it is not, as explained in more detail in later comments, the numbers for AcuPebble would be incorrect.**	• A short unpublished report of this two-phase study was provided by the company to NICE in August 2023. • As per our standard process for all company supplied study reports, we screened the Phase I Virgen Macarena trial report for potential inclusion in the systematic review. However, it was unclear whether the study met the inclusion criteria due to limited detail on the study design, methods, participants and outcomes. • We also noted apparent contradictions between the clinicaltrials.gov record for this study and the company supplied report. For example, whether the study was restricted to adults or a paediatric population (NB. the company later confirmed there had been a factual inaccuracy in their report, and the study is on an adult not a paediatric population). • There was no mention of completion date for either phase of the study, (The title of the company's report states "Preliminary Analysis Report".) • • No description of phase 2 other than brief details on the clinicaltrials.gov site. • No mention of whether the results of phase 1 and phase 2 were independent of each other or whether the intention was for them to be combined in a single data set. • We also noted ambiguities in the results included in the company's report which would also need clarification if the study was to be included (e.g. no explicit definition of the diagnostic thresholds used; non standardised patient characteristics data in Spanish, lacking any commentary). In situations when it is unclear whether a study meets the inclusion criteria for a systematic review, we would aim to contact the study investigators for clarification. For the reasons stated above it took additional time for us to screen this study to determine which aspects of the study we would need to seek clarification on before we could make an informed judgment about inclusion status. We thank the company for responding to our request for information (November 2023). The study is included in the systematic review and economic evaluation in the Erratum based on the results currently available from Phase I. Finally, in response to point 4 made by the company above, a scenario analysis using the published data available from Devani, without adjustment, was conducted in the original report. This has also been done in the Erratum (please see Section 5.10.2 and Appendix 9a).

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	58	9	Scientific summary, Results	**This should be redone** since something has failed in the methodology used, considering that they have missed for the performance metrics the landmark independent publication of WatchPAT (500 people) that, amongst other things, shows a massive drop in performance with respect to what has been used. This is particularly important considering that trial turned out to have a 72% Black or African American population (i.e dark skin) which does highlight in the context of sleep the problem of diagnosing using devices that heavily rely on PPG as the sensing modality (PAT in this case), and do not have any direct indicator of airflow. Note: After writing this paragraph it was observed the EAG explained why they had excluded this paper in the Appendix in the middle of a long list that most readers would miss. This reason was that it was done with the WatchPAT200, despite it being an intervention considered eligible in the EAG's methodology. This is also not a strong argument/excuse, considering that results have been “mixed and matched” from different WatchPATs for their analysis, so even their methodology allows this. Furthermore, independently of the “mix and match” methodology (which is not correct for reasons that will be further explained in subsequent comments) from the point of view of diagnosis of OSA, and even more specifically in what relates to the performance metrics that are being used for the health economics analysis, there would need to be justification of why one of them is different/same to one another (for example, having an extra channel might make no difference if that output has no relevance to the evaluation of what is being compared. Or even if it did, if that channel increases the confidence and still the performance is poorer, a case contemplating this needs to be accommodated for in the health economics analysis). And even if this is not done **these results need to be discussed in the text properly,** considering the recent evidence that has come out about medical devices that generate outputs from PPG potentially having been designed with bias against black people, this device being based on PAT which is a signal resulting from PPG, and this being the only trial in which a majority of dark skinned people took part.	With regard to the exclusion of the Ioachimescu 2020 paper, see our response to comments 253, 24 and 26.
Acurable Limited	59	9-10	Scientific summary, results	The second and third paragraphs of the "Results" subsection, spanning pages 9-10, fail to acknowledge the fact that the diagnostic accuracy data for another study with AcuPebble was available. **This needs to be corrected.**	Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (2023)
Acurable Limited	60	10	Scientific summary, results	In the third paragraph text is copied from the abstract referring to the accuracy data of devices. See comments related to this in the abstract above (001 - 004) since they are the same.	The comment numbers have changed and comments 001 – 004 are now comments 47-50. Please see our response to comment 47, since it is the same.
Acurable Limited	61	10	Scientific summary, results	The fifth paragraph is also copied from the abstract (it refers to their assessment of limited data being available). See comments on the abstract in relation to this (001 - 004).	See response to comment 60 directly above

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	62	10	Scientific summary, results	Children and young people aged 2-16 years subsection: The authors fail to mention the fact that some manufacturers are claiming children can use their devices without ever having done a clinical validation in children. **They should identify those devices and warn against their use.**	It is not within our role to issue such directives. This is more appropriately addressed at a policy level (e.g. NICE).
Acurable Limited	63	11	Scientific summary, results	**External assessment Group (EAG) independent economic assessment subsection:** This subsection is identical to the one in the abstract. See comments there. In a nutshell **this section needs to be fully rewritten because the results need to be recalculated**. It also needs to **bring up important points, such as the fact that accuracy data that led to the number of Sunrise had been obtained with post-hoc determination of diagnostic thresholds that are not the conventionally established ones.** What this means is that based on the ROC curve, a diagnostic threshold was decided as the cut off for e.g. moderate sleep apnoea (for example 13, as opposed to 15)where the results would be best, and that is what has been reported and used.	The limitations of the results from the Sunrise studies (where post-hoc optimisation has been used), are highlighted in discussion of the accuracy evidence, and in the Discussion/Limitations (Section 6.3.2 now section 6.2.2).
Acurable Limited	64	11	Scientific summary, Results	**Last paragraph of the Results section (Scientific Summary)**: **If oximeters are highlighted there needs to be a warning of the limitations of oximetry and the fact that these conclusions do not apply to those with dark skin** (for example black ethnic minorities). The numbers here are also incorrectly obtained. **This needs to change once the correct inputs are applied to the model, and hence after the correct numbers have been calculated**.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities. Regarding model inputs, please see our response to comments 128 and 138
Acurable Limited	65	12	Scientific summary, Conclusions	First bullet point: **This paragraph needs to be removed**, unless it clearly states the dangers of using oximeters for those with dark skin (with full explanation).	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	66	12	Scientific summary, Conclusions	The second bullet point, which states "oximetry is estimated to misclassify a high proportion of people with mild OSAHS as not having OSAHS...", has to be more assertive. It is “estimated” underplays the problem.	The results discussed here are estimated from the model, therefore it is appropriate to refer to them as being estimated.
Acurable Limited	67	12	Scientific summary, Conclusions	Third bullet point: See previous comments (ref. 012) about mixing results from different WatchPAT products. This can put patients at risk, unless proper research is done (the data exists) to fully justify this is ok. Also, there should be a comment about sustainability here. Disposing of a WatchPAT has a massive environmental cost. This needs to be commented upon.	Comment 012 is now renumbered as comment 58. We repeat our response to comment 58 - with regard to the exclusion of the Ioachimescu 2020 paper, please see our response to comments 24, 26 and 253. With respect for the comment on the environmental impact of disposable devices, we agree that this is an important issue for consideration, although assessment of this impact is beyond the scope of the EAG report.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	68	12	Scientific summary, Conclusions	Fourth bullet point: **A comment (with explanation of what it means) of Sunrise doing post-hoc adjustment of thresholds, so that performance values are given at an optimum point that does not correspond to the diagnostic thresholds and that is what leads to better results, must be made.**	Please see response to comment 106.
Acurable Limited	69	12	Scientific summary, Conclusions	Fifth bullet point: The investigators are putting all devices "in the same bucket" and that is **both scientifically unjustifiable and also misleading from the point of view of this exercise**:1- The uncertainty is already statistically determined in whatever metrics have been used to report their results. Hence it is not for the EAG to make a comment on this, mostly considering that the choice of statistical metrics is already questionable. Sensitivity and specificity have been long deemed not to be the most representative to determine the performance of diagnostic devices. So the EAG themselves are introducing "uncertainty" in the devices evidence by developing a model based on poorly informed and in most cases barely justifiable scientific assumptions; 2- They are also implicitly conveying the message that all the validation trials done by the devices have the same value in terms of evidence, when in some of the trials there hadn't even been a power calculation to start with (and if there was this was not presented). This paragraph has to change to be unequivocal and correct for all of this.	Due to the amount of uncertainty surrounding the results from the model analysis (which includes the probabilistic analysis, the one-way sensitivity analyses and scenario analyses) for all devices, we believe this caution should be highlighted in our report conclusions. This uncertainty is not just parameter uncertainty that can be captured via the PSA, but also uncertainty as to the most appropriate data sources to use (e.g. accuracy estimates for the comparator of RP). Please see response to comment 94 on the use of sensitivity and specificity in the model.
Acurable Limited	70	13	Research recommendations	Research recommendations subsection: This section is completely wrong. The composition of the EAG is missing regulatory, medical devices, ethics and OSA experts, which clearly shows in these conclusions. However, this being in a report endorsed by NICE can be taken by others as "dogma", which can have a negative impact on the research community overall. On the basis of this, it should be removed. A brief analysis of why is given point by point below (this can be elaborated upon on request): 1- The EAG recommends doing more trials to compare the accuracy of PSG versus home PG. This is questionable considering that there are many factors that will affect the so-called "accuracy" that have nothing to do with the technology of RP with respect to PSG. Amongst others, night to night variability. And there are situations in which RP will give a better representation of the state of the disease in a particular patient than PSG would, even leading to a different diagnosis, since RP at home might allow a more natural sleep and consequently better represent what truly happens to the patient. Hence, what would be the value of posing this scientific question, beyond what has been done and published already? 2- The EAG recommends diagnostic accuracy trials if new versions of devices are launched. This might be completely unnecessary, and hence this recommendation could be unethical (there needs to be a reason to subject patients to clinical trials). It depends on what changes a certain device has with respect to others. And if there are changes that would affect the diagnostic performance, the evidence is already there (albeit it might not be in the public	Please see responses to comment 250 (Acurable) and comment 255 (BTS) below.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				domain) because the devices have undergone a regulatory process which requires this evidence (or alternatively the device would have a different regulatory intended use, which is something critical that the EAG hasn't even posed as a question in this exercise). This question should have been asked by the EAG to the manufacturers, not just because of this point but because if there are relevant differences already affecting accuracy the new numbers should have been used for the whole evaluation. 3- The answer to the broader question they pose (whether diagnostic accuracy results are transferable between different settings for sleep studies) as a research priority is already known: the answer is that it is not generalisable since it depends on the technological characteristics of the devices, so it does not make sense to try to infer how the results would convert on the basis of what happened with devices that are technologically different. 4- The investigators propose to do validation in children at home. This is something that has already been widely considered. There is a fundamental reason why comparison studies in children are done in the clinic, mostly in those with comorbidities: someone needs to be making sure the children keep the RP-PG sensors on. Carrying out trials with children at home poses ethical issues, which is the reason why ethics committees generally request them to be done in hospitals. 5- The authors suggest finding ways to have indirect comparisons between novel and conventional devices with appropriate adjustment for heterogeneity. This makes no scientific sense because one cannot extrapolate to infer performance a comparison that is evaluating a technology that is completely different. As for the conclusions in children, the EAG hasn't even based these comments on proper literature review. They base them on conversations with others, without saying how many others, what questions were they asked, how was any quantifiable data from those conversations obtained (for example, was it anecdotal, based on a specific centre, based on publications, etc…). This is very unscientific.	
Acurable Limited	71	29	1.1	Second paragraph: The authors are referring to the effects of untreated OSAHS in adults without giving references. Since this is important information the reader should be able to trace the source.	We have added a reference to the NICE Guideline on OSAHS in adults (NG202).
Acurable Limited	72	29	1.1	Third paragraph: The authors are referring to the effects of untreated OSAHS in children without giving references. Since this is important information the reader should be able to trace the source.	We have added a reference to Bitners et al. 2020.
Acurable Limited	73	29	1.1.1	First paragraph: The authors are referring to a prevalence of the disease in the UK population without citing the sources where they obtained this prevalence for.	A reference has been added (Bitners et al. 2020).

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	74	29	1.1	**Second paragraph:** The authors are referring to increased risks of OSAHS in certain conditions but they are not giving quantification for those risks. Some of them are certainly more significant than others. Hence this should be given, together with references to support their statements.	This is the list of conditions that NICE guideline NG202 advises clinicians to be aware when assessing people for OSAHS since people with these conditions have a higher prevalence of OSAHS. Quantification of risks for each of these conditions is beyond the scope of the background section of this report .
Acurable Limited	75	29-30	1.1.1	The authors switch from risk of having OSASH in the adult population in paragraph 2, to the potential causes of OSAHS in children. This is lacking quantification or references to the sources. Hence this fails to inform the reader nor sets up the context in a scientific manner. Furthermore, the whole argument in this section is very messy since it passes from risks of certain comorbidities in adults to leading causes of OSAHS in children. We suggest that it should be: 1- Risks of co-morbidities in adults; 2- Risks of co-morbidities in children; 3- Potential causes in adults; 4- Potential causes in children. All of it should be properly quantified and references given.	Please see our response to comment 71. Thank you for the suggested format. We think the current approach is clear enough for its purpose. Please note also that we are required by the NIHR Journals Library to keep within an overall word limit for the main body of the report.
Acurable Limited	76	30	1.2	This paragraph massively downplays the issues associated to pulse oximetry diagnosis. The word “regarded” means “*consider or think of (someone or something) in a specified way*”. The authors refer to the problem of sensitivity of the pulse oximeter being thought of as being less than in other tests, and gloss over it without a reference for quantification. All current pulse oximeters in the market suffer from a massive inaccuracy problem when used with people with dark skin. This is the result of the signal that the SPO2 value has been extracted from (PPG) massively varying with the colour of skin because the light transmission and absorption properties are different. Because of this in people with dark skin the interfacing electronics fails to work properly. Also, and most importantly, the calibration algorithms from which SPO2 is obtained have been fitted to those with fair skin so they fail for those with dark skin. Note that this has recently been proven following the disproportional amount of deaths of patients with dark skin during COVID, which has led to the conclusion that any diagnostic metric that is obtained using the PPG signal has to be questioned in individuals with dark skin, and the use and testing of oximeters overall is currently under evaluation by regulatory bodies around the world because of this. The reason why oximeters have passed regulatory processes is that the ISO standard the latter rely on only requires 67% of pooled data (i.e. no patient distinction) from 10 patients (20 samples per patient), only 2 of whom has go dark skin, to be within 4% of the absolute truth. On the basis of this, all the samples from patients with dark skin can be off and the oximeters still pass. Furthermore, the ISO standard that regulatory bodies require oximeters to pass, only requires patients to be static. The way the EAG downplays the sensitivity issues, considering how influential NICE is is potentially dangerous and promotes health inequalities. It is not acceptable to assume that just because they might have followed obsolete references they are not aware of what is happening with oximeters right now (i.e. the fact the evidence resulting from COVID is currently being evaluated, and regulatory bodies are working to see how they can update the standards and the validation to avoid the bias against those with dark skin in all clinical applications in which oximeters are being used). **Oximeters should not be considered as an alternative put forward in this report and if they are, every time they are mentioned the reader needs to be reminded of the**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				dangers of using oximeters with patients with dark skin, and the potential health inequalities.	
Acurable Limited	77	30	1.2	Respiratory polygraphy is scientifically poorly described. The first sentence describes the physiological signals that are measured, but that sentence is missing airflow. Then the second sentence is giving a layman explanation of the type of sensors, but that is also not rightly phrased because it implies that it is only the tubes of the nasal cannula that are attached to a monitor, whereas all the other sensors are too.	The description of respiratory polygraphy is based on that used in NICE guideline (NG202) Obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s (page 50). We have amended the text by adding “airflow” and making it clear that all sensors are attached to the monitor.
Acurable Limited	78	30	1.2	The description of PSG is scientifically weak and not unequivocal. It does not make reference to the sensing technique used for the additional channels (i.e. EEG, EMG, etc). The exact additional clinical/physiological information is also too vague. For example, in the way it is described the “additional assessment” of sleep quality could just be via a consumer device with no EEG. However that is not PSG.	The description of PSG is that used in NICE guideline (NG202) Obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s (page 50). The description will therefore not be changed
Acurable Limited	79	30	1.2	Penultimate paragraph on the page: This states that pulse oximeters can be sufficient to inform of a diagnosis with a high pre-test probability. This is not correct and it is potentially promoting health inequalities. It has been recently demonstrated following the disproportionate deaths of people with dark skin during COVID (although it would have been known if one was to assess the regulatory framework behind the development of pulse oximeters), that pulse oximeters (as well as other sensing modalities) that operate in the principle of PPG are not sensitive enough in detecting desaturations in those with dark skin. Hence this paragraph is discriminatory and negatively biased against those with dark skin. **Either unequivocally explain the differences and risks for people with dark skin or remove references to oximeters as an alternative.**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	80	30	1.2	Also in the penultimate paragraph of the page, the statements are vague and unscientific: What are “significant comorbidities”? What comorbidities is it referring to? Where are the references? **Please correct and clarify.**	Please see our response to comment 71.
Acurable Limited	81	33	1.3	This section, describing the diagnostic technologies under assessment, is very much incomplete and it massively downplays the benefits of emerging technologies. As before it lacks references and the statements are so vague that this lacks scientific rigour. Example of benefits that have not been addressed: patients not having to travel, addressing health inequalities, enabling in some cases the diagnosis of vulnerable individuals, the increased confidence in the diagnostic output representing the “true diagnosis” as a result of the patient being able to have a natural night sleep as opposed to being forced to a position. The External Assessment Group was provided in advance with evidence and documentation to support this. However this has been ignored.	Please see our response to comment 71. This section is a general introduction to the diagnostic technologies and the claims made for their potential benefits. In turn this provides the rationale for doing this *independent* diagnostic assessment. Our role is to critically evaluate the evidence for benefits (and harms) and therefore whether the claims made are duly substantiated. At this formative section of our report we must maintain our independent objective academic position with regard to claimed benefits or any other consequences that we will be assessing.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	82	34	1.3.1	**The description of AcuPebble is factually incorrect. The incorrect facts must be corrected.** The manufacturer provided the correct description but the External Assessment Group provided their own interpretation (or the interpretation of others found in Google) as opposed to using the manufacturer's and regulatory approved documentation. More specifically: 1- In the way it is written it implies the device records only sounds generated by the patient's respiratory and cardiac functions. Whilst not being entirely wrong, the device records a mixture of acoustic signals that go beyond that. For example, the device also outputs patient activity throughout the night, not corresponding to the respiratory and cardiac function. This information was provided to NICE in advance, but it has not been incorporated. **Please be more precise.** 2- The description says that a Wifi connection is strictly needed. **This is factually incorrect.** A wi-fi connection is not strictly needed. How the device works when there is no WifI was extensively explained in documents submitted to NICE (based on real world data). **This must be corrected.** 3- **The list of outputs of the device is incomplete**, despite the fact this was provided by the manufacturer. **This must be corrected.** 4- The description says that the device is not intended to be used for people with suspected arrhythmias. **This is factually incorrect.** The EU/UK device has been approved by regulators and can be used by people with arrhythmias. NICE had been provided with the user EU/UK manual but that user manual doesn't seem to have been looked at. **The exclusion of people with implants is also obsolete. This must be corrected.** 5- The description says that the device is not for use for people with significant cardiopulmonary or neurological disorders. **This is factually incorrect.** The device can indeed be used in people with both. This has been approved by regulators (CE mark). And in fact the manufacturers had provided NICE with evidence showing the device had been used in people with both during the clinical trials.**This must be corrected.** 6- The description says that the device cannot be used in people with a known allergy to acrylate. Whilst the adhesive contains acrylate, nearly all other devices presented in this review also contain acrylate and the manufacturers don't have this exclusion in the instructions for use, and neither do other medical devices that also rely on adhesives having acrylate in their composition. Acurable provided NICE with evidence showing that the other adhesives also contain acrylate, as well as informing them that AcuPebble SA100 could now be used with patients with an allergy to acrylate at the clinician's discretion. So if this distinction is going to be made for AcuPebble it should also be made for all the others. **Please make the same comment for all the others that use acrylate**	1. We have updated the text in line with information provided in response 10 of Acurable's responses to NICE's request for information (DAP70_request for information_Acurable_v0.3 [ACIC]) 2. We have updated the text in relation to Wifi connection with information provided in response 17 of Acurable's responses to NICE's request for information (DAP70_request for information_Acurable_v0.3 [ACIC]) 3. The outputs listed were based on the bullet points in response 11 of Acurable's responses to NICE's request for information (DAP70_request for information_Acurable_v0.3 [ACIC]). We have amended the wording so that is identical to the wording of the black bullet points. The content of the white sub-bullet points have not been added as this is too granular in the context of the background section of the report. 4. Text stating the device is not intended to be used for people with suspected arrhythmias has been removed. 5. Text stating that the device is not for use for people with significant cardiopulmonary or neurological disorders has been removed. 6. Text relating to acrylate has been removed. 7. Testing stating the device might not be suitable for patients with bruxism has been removed. 8. The information regarding pregnancy was stated for Brizzy as its' manufacturer provided this information to NICE. The information regarding pregnancy has been added to the description of AcuPebble SA100 now that Acurable provided this information in the comment

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				or eliminate for AcuPebble. The details of the different adhesives for all the technologies were passed to NICE in summer 2023. 7- The description says that the device might not be suitable for patients with bruxism. **This is factually incorrect**, and in fact in some of the evidence provided to NICE of validation of the device some of the patients had bruxism. There is absolutely nothing in the design of the device or in the algorithms that make it unsuitable for people with bruxism. **Please correct.** 8- Also, it is noted that for others it is said that there is no exclusion for pregnancy whereas the same is not said for AcuPebble. **Please state that there is no exclusion for pregnancy.**	
Acurable Limited	83	36	1.3.3	NightOwl: The report states that NightOwl is suitable for children from 13. Have the manufacturers provided evidence of this having been tested in children? Also, the report is referring to a device that is not commercialized ("the device to be commercialised in the UK"). Does that particular device have the UKCA or CE mark? If not, please could it be clarified why has it been considered for this review? **Please remove from this report and work any device that is not commercially available in the UK. And clarify whether any of the manufacturer's claims for intended use in children is backed by any evaluation in children.**	With regard to regulatory approval please see our response to comment 100 (Acurable).
Acurable Limited	84	36	1.3.3	NightOwl: Nothing is mentioned about the accuracy of the output in people with dark skin. This device is almost exclusively based on PPG, and the validation has been done following the ISO 80601-2-61:2017 standard. It has recently been proven following the disproportional amount of deaths of patients with dark skin during COVID that any diagnostic metric that is obtained using the PPG signal has to be questioned in individuals with dark skin, and the use and testing of oximeters overall is currently under evaluation by regulatory bodies around the world because of this. The reason why oximeters are not reliable in patients with dark skin is because the light absorption is very different and this leads to signals that the oximeters have not been properly calibrated for, since the above mentioned standard also requires 67% of pooled data (i.e. no patient distinction) from 10 patients (20 samples per patient), only 2 of whom has go dark skin, to be within 4% of the absolute truth. On the basis of this, all the samples from patients with dark skin can be off and the oximeters still pass. Furthermore, the standard only requires patients to be static. Nightowl's validation of the oximetry signal has been done using this standard, with ten patients. There is no mention of this limitation anywhere. Please discuss in the text.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	85	36	1.3.3	**NightOwl: There is no mention of data protection**. In order for the company to be able to offer a fulfilment process (which they are unclear whether they are already offering or not) the device and the information needs to be cleared out in	Assessing compliance with data protection regulations is outside the scope of the EAGs work. Note that recent NICE guidance has stated the need for technologies to follow NHSE's Digital

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				terms of data protection by the NHS. There are devices in this review that collect, use and store patient personal information data in a way that contravenes the data protection requirements of the NHS. **Manufacturers should be asked about this (ie. whether any NHS Trust have approved that they are the ones doing fulfilment) before assuming they all can for the health economics model. For those that can't the health economic models inputs have to change (only for those technologies) to accommodate for the fact they cannot offer it in practical terms. The report should clarify this too, and mention data protection issues and requirements of personal data sharing for the different devices. Please change accordingly, if relevant (NightOwl might not have any data protection issue. This comment is a generic one)**	Technology Assessment Criteria (DTAC), which includes data protection.
Acurable Limited	86	36	1.3.3	NightOwl: There is no mention to the fact the device is attached with an adhesive that contains acrylate. **Please mention this or remove for AcuPebble.**	Text on acrylate in relation to AcuPebble has been removed.
Acurable Limited	87	36	1.3.4	This section says for Sunrise manufacturers: "The company offers a service to send the device directly to the patient". **This device collects, stores and uses sensitive patient information (this can be simply confirmed by downloading the app)**. Could it be clarified whether this is effectively possible and implemented within the NHS? Considering what the device collects this contravenes the minimum requirements imposed by the Data Protection agreement that has to be in place for this. **Manufacturers should be asked about this (ie. whether any NHS Trust have approved that they are the ones doing fulfilment) before assuming they all can for the health economics model. For those that can't the health economic models inputs have to change (only for those technologies) to accommodate for the fact they cannot offer it in practical terms. The report should clarify this too, and mention data protection issues and requirements of personal data sharing for the different devices. Please change accordingly.**	Assessing compliance with data protection regulations is outside the scope of the EAGs work. Note that recent NICE guidance has stated the need for technologies to follow NHSE's Digital Technology Assessment Criteria (DTAC), which includes data protection.
Acurable Limited	88	36-37	1.3.4	Sunrise: There is no mention to the fact the device is attached with an adhesive that contains acrylate. **Please mention it or remove for AcuPebble.**	Text on acrylate in relation to AcuPebble has been removed.
Acurable Limited	89	36-37	1.3.4	There is no mention to the fact the device cannot be used by those with a beard unless they shave it. Note that this is important because there is a health inequalities component to it since in many minority groups beards go beyond aesthetics. **Please make a note on this in the report.**	The following text has been added to the device description "device may not be suited for bearded users and advise close shaving or usage of provided adhesive bandages to ensure optimal adhesion", which is as per Sunrise's health care professional and patient user manuals. To the description of AcuPebble SA100 in section 1.3.1 we have also added the following text "if the sensor is intended to be attached on a hairy surface, this must be shaved", which is taken from AcuPebble SA100 user manual version 1.11.1,.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	90	36	1.3.4	Has the device been regulated for children over 3? If it hasn't this should not be said in the opening sentence since strictly speaking the device cannot be used in the NHS for children over 3. **Please change accordingly.**	Please see our response to comment 100
Acurable Limited	91	37-38	1.3.5 1.3.6	**This section describes WatchPAT.** There is no discussion in these sections to the potential contraindications of using PAT on those with dark skin. This is the result of PAT being obtained directly from the PPG signal, and the absorption/transmission/reflection of light being very different in those with dark skin leading to signals that either have too low signal to noise ratio, or it saturates the hardware. Note that reference (Ioachimescu, 2020 (68)) shows a trial with African/Americans where the performance of the systems gets massively reduced as a result of this. This is elaborated upon in subsequent comments, hence the requested actions will be obvious later.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	92	39	1.4	Care pathways: There is a reminder to the reader of the recommendation of oximetry. Considering the new evidence that has appeared as a result of the disproportionate amount of deaths in COVID of those with dark skin -on the health inequalities associated to the oximeters since the regulatory process does not account for the fact the physical principles oximeters are built in works differently in dark skin, and the regulatory requirements are based on those with white skin- highlighting this recommendation in a 2024 NICE report would be equivalent to ignoring this, also at a time when regulatory bodies around the world have announced a review of the principles of oximetry due to this; and potentially promoting health inequalities. **The reader needs to be made aware of the risks of PPG based systems (and this includes oximetry) on those with dark skin and the fact any statement or conclusion in this report might not apply to them.**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	93	41	2.2	This section lists the devices together with the indicated ages. But the vast majority of these devices have not been validated for children. Could the manufacturers provide their regulatory certificate and regulatory intended use to demonstrate that their claim is legal? If not, could this report avoid giving information that is "marketing claims" (looking at future products), since not every clinician will read this report in detail and having something like this in the opening sections can lead to products being used in a way that compromises the safety of the patients? (in this case children). And if the manufacturers went through an MDD regulatory loophole (or justified it by risk assessment), **could readers be reminded throughout whenever the ages appear which devices have not been validated in children?**	The list of devices and their indications (specifically by age) is as reported in the NICE scope. In turn this is based on information provided to NICE by the product manufacturers.
Acurable Limited	94	42	2.4	**Outcomes:** The authors are using metrics to evaluate diagnostic accuracy that, although commonly reported in research papers, are not reported in isolation, since statistically it is well established that sensitivity and specificity can be misleading when considered in isolation, to evaluate diagnostic methods. More	*Diagnostic accuracy statistics* As detailed in Section 5.6.1, the model uses sensitivity and specificity estimates to determine what proportion of the modelled cohort go through each pathway of the decision tree. This is a

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				specifically, in the context of OSA it is well established that they should not be used to evaluate the merits of devices, and instead likelihood ratios (preferentially) or predictive values should be used [This can be elaborated upon if helpful]. On the basis of this, the use of sensitivity and specificity can lead to the wrong recommendations made on the basis of this economic evaluation. Thus, these variables should be changed in the model to likelihood ratios or predictive values to increase its scientific value and to further validate the final conclusions and outcomes. Although in some papers those metrics are not reported, they can be calculated from the information given by the authors.	standard approach to decision modelling of diagnostic technologies. As discussed, this approach, based on sensitivity and specificity at the two diagnostic cut-offs, is a simplifying assumption. A better approach is to model the 4x4 contingency table directly. Due to these data not being available for all devices, we followed the approach used in the original NG202 model which used estimates of sensitivity and specificity. For novel devices where we have data from the 4x4 contingency tables (AcuPebble, NightOwl, WatchPAT 300 and WatchPAT ONE), we have conducted a scenario analysis using this data directly in the decision tree against the 4x4 contingency data for respiratory polygraphy (please see Section 5.10.2). We thank Acurable for providing the 4x4 contingency data for the MACARENA trial. This parameterisation leads to NightOwl and the WatchPAT devices having favourable cost-effectiveness results: INMB at £20,000 per QALY of £26 (NightOwl), £73 (WatchPAT 300) and £49 (WatchPAT ONE), please see Table 43 in Section 5.10.2. As discussed in Section 5.10.2 of the report, improved cost-effectiveness for the WatchPAT devices is driven by the reduction in performance of respiratory polygraphy to identify people with mild OSAHS as having OSAHS, and the fact that WatchPAT is still more likely to over-diagnose the severity of mild OSAHS. As discussed for the base case results, this is likely to lead to greater QALYs. The improvement in cost-effectiveness with NightOwl is also explained by it being more likely than respiratory polygraphy to over-diagnose severity for patients with mild OSAHS when the 4 x 4 contingency data are used. We also note that [redacted]. Finally, in relation to the statement on sensitivity and specificity *"in the context of OSA it is well established that they should not be used to evaluate the merits of devices"* we would like to point out that sensitivity and specificity were the primary outcomes in the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024)).
Acurable Limited	95	44-45	3.1	This search was clearly not properly done considering there is a clinical trial registered in Clinicaltrial.gov for AcuPebble/Acurable, that has been there for years and has not been found. This is even more surprising considering that the manufacturer provided NICE with the actual information about the trial in August 2023, and this has still not been included in the review, which raises the question of how many more sources of information haven't, as well as doubts about the whole validity of this review. Later on in the report other relevant	Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024)). In addition, we would like to emphasise that the literature searches were designed and extensively tested by an experienced information specialist before deployment in the relevant databases. The other references said to "missed" by the literature searches were identified by our search but they did not meet the

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				references were noticed to be missed that also seem to indicate that the search might not have been as exhaustively done as it should (which could be partly justified by the very short time given to all the parties involved to do this). **Note:** It was noted later on in the Appendices that some of the references and a document this refers to were listed but the EAG decided to exclude for unjustifiable reasons (i.e. not completed trial- this is irrelevant considering the arm that was relevant to their assumption had been completed and they had the numbers for, including also confidence intervals; and different version of WatchPAT, when they had no problem using other versions to "mix and match" results, but they did not justify properly in the body of the text why the largest trial had been excluded. **The identification of studies must be corrected, and the data for the PSG trial of AcuPebble must be included. And the results of the loachimescu 2020 trial must also form integral part of this document considering the potential health inequalities (in terms of safety) implications it highlights.**	inclusion criteria. These can be identified by examining the list of references excluded at full text screening in Appendix 2b of our report. This is a standard reporting requirement in systematic reviews (see the PRISMA 2020 Statement). With regard to the IOACHIMESCU 2020 study please see our response to comment 253 below.
Acurable Limited	96	45	3.1	In the last paragraph before section 3.2, the authors state "to identify any further relevant primary studies we also searched the manufacturer and distributor evidence submissions to NICE". **This is clearly not true** considering Acurable provided the identifier of a registered clinical trial they haven't mentioned, together with the results of that trial that have been completely ignored by the authors of this report, who have gone as far as not to cover in the text their existence to justify theoretical adjustment of results (which make no sense) for their economic model. More elaboration of this in subsequent comments.	Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024)) . See also our response to comment 95 above. In addition, we would like to strongly refute the claim that relevant evidence was deliberately ignored to justify theoretical adjustment of the results.
Acurable Limited	97	46	3.2.3	This section rightly establishes that home oximetry alone is not considered as a comparator for people with COPD. However it still considers it as a suitable comparator for people with dark skin (even when at the top of the page the authors mention that people from black minority ethnic backgrounds are considered as a subgroup), despite the overwhelming evidence during COVID , that has led to a review of regulatory standards to approve oximeters everywhere around the world, that showed that oximeters do not work well on those with dark skin. Hence oximetry SHOULD NOT have been considered as a comparator for those with dark skin since that is equivalent to NICE being potentially a promoter of health inequalities. Alternatively **this report should claim up front and in the title that the findings do not apply to those with dark skin**.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	98	48	3.6	**Last paragraph:** It is stated "we used an adjustment method to ensure all novel devices were compared on a level playing field". Note that this is too simplistic, as will be demonstrated in comments for subsequent sections, because it is not just enough to try to be "fair to all the devices", the comparison must take into account metrics that represent the final intended use, and take into account the differences between device outputs as well as the different technological characteristics. Hence extrapolation is not possible. They need to rely on	With regard to the adjustment to compare AcuPebble to PSG please see our response to comment 138 (Acurable).

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				evidence that is unique for each one and then maybe discuss them rather than trying to make up numbers for their models. The authors of the study have ignored this, just for the sake of having a "level playing field" (i.e. have decided to take metrics that do not represent the intended use, and hence neither do the conclusions of this study). This is not scientific and it is potentially dangerous for patients since decisions can be made on the basis of claims in this report. **Remove this justification since this could be grounds to other future researchers in other areas to learn a scientific methodology that is wrong.**	
Acurable Limited	99	50	4.1	**In the section where included studies are reported:** There is at least one clinical trial of Acurable that meets the inclusion criteria. This trial is registered in clinicaltrial.gov. Although the trial is officially ongoing this is because it has two independent parts. **The first part, which is the one relevant to this report, is complete and NICE had been told about it**. These results must be included considering they were provided on time and will most certainly be published by the time this report is released. Surprisingly these haven't been acknowledged or included in the main body of the review. It is worth noticing the authors of this report do however mentioned having contacted the manufacturers of NightOwl for results of a trial which was registered so that this could be included in the review and they did not reply, which suggests that Acurable's results would have met the inclusion criteria, and in case of doubt Acurable should have been contacted to provide clarification, something that never happened. Throughout the report, as proven by the comments the EAG seems to have a double standard when generating inputs for their model corresponding to different devices. **Whilst in some devices performance is overestimated, so that devices appear better than they are, for AcuPebble the EAG has gone as far as to not consider evidence and invent data to make it appear worse.** This will become more obvious in subsequent comments.	Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024)) and comment 96. This trial was included in the main text of the report, in section 4.11 (Ongoing studies). The results of the study have now been incorporated into the systematic review in the Erratum.
Acurable Limited	100	52	Table 2	It seems the authors of this report haven't sought for evidence of some of the devices and the corresponding publications being CE-marked/regulatory approved. The fact that a manufacturer has a new version and has published the results does not mean that that device can be available to the NHS. This has to undergo a regulatory process which at the moment is incredibly slow due to the lack of Notified Bodies approved to carry out MDR evaluation. Without MDR devices cannot have major changes, and this includes any change in intended use (no matter how small this is), change of intended population, or change in whether the device is reusable or not. However, **in this table the authors have decided to use as evidence publications of very novel variants of the devices that most certainly need to undergo the MDR process (two year waiting list right now)**. This needs to be clarified. Manufacturers need to present evidence of the fact that whatever version of the device was used for those publications has the CE mark already. Otherwise these should not be taken as evidence. **Please remove from the report any analysis on health economics of devices that do not have CE/UKCA mark.** Alternatively if they are going to be considered, the search and inclusion needs to be expanded to	Thank you for highlighting the issues regarding regulatory approval and CE marking. The NICE health technology evaluations manual: methods and process (2022) states that a health technology is evaluated only if it has or is expected to have regulatory approval (or appropriate regulatory signal) by the planned draft or final guidance publication date (section 2.2.5), The NICE diagnostic advisory committee does not normally make recommendations on using a technology outside the terms of its regulatory approval (section 6.1.11 in the same manual). However, evidence relating to the technology being evaluated that is outside the terms of its regulatory approval may be considered during the assessment phase of the evaluation. This may inform the committee's discussions about the use of the technology within the scope (section 6.1.12 of the same manual). Regarding NightOwl, the company have confirmed with NICE that the NightOwl device described is the same device previously

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				all academic systems, even when there is no certainty of them ever getting to patients. **It is not enough for manufacturers to say that they are close to having it. And a double standard is not acceptable either:** in the same table, **why have the interim results of a study been taken into account for Sunrise and the results of two studies of AcuPebble SA100 have not? And why have the manufacturers not been contacted about this either?** It is both wrong scientifically and from a regulatory perspective to consider results of a device that is not the device under evaluation as representative of the device under evaluation. This makes the review scientifically flawed and could potentially put patients at risk if someone was to take this table out of context. Some of these studies have no statistical significance because the sample is too small. They should not be considered scientifically valid for this type of evaluation, unless they are identical to other studies of the same device (same version of the device, same setting, etc…) so that results can be pooled. Equally studies for which the sample has not been calculated in advance should not be considered.	named "NightOwl Mini", and that this differs only from the NightOwl reusable version in terms of the battery. All sensors and software in both devices are the same and are CE marked. The company have confirmed that the NightOwl (aka NightOwl Mini) will be the device available in the UK. As a general point it isn't always clear from study publications whether or not a given novel device evaluated in a study is necessarily the same version/iteration as the one with current regulatory approval. This tends to be the case for novel devices which don't have readily identifiable model ID/version numbering as part of the device name. We have endeavoured to provide explicit details of the novel devices in each study for transparency. We highlight studies where the device may not necessarily be the version with current regulatory approval so that the NICE diagnostic advisory committee can take this into account in their deliberations.
Acurable Limited	101	54	4.1, Second paragraph on page 54	The authors state "Table 2 illustrates further disproportionality in the evidence for novel devices. The Sunrise device was evaluated in three studies, NightOwl in three studies and …". The authors are mixing studies that should not be mixed without further due diligence: Do the three studies correspond to regulatory identical devices? If they don't they should not be considered instinctively. There is a reason why regulatory bodies require a full regulatory submission when there are different variants of devices. Also, **it is not correct that AcuPebble has only had one study with adults. Evidence was provided of results of another study. The part relevant to this review had already finished and the performance metrics as well as demographic data had been provided to NICE. However, this has not been taken into account. This study was with PSG. Hence, this whole section needs to be corrected to incorporate AcuPebble's PSG study.**	Please see our response to comment 100 Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024))
Acurable Limited	102	54	4.1, last paragraph on page 54	**It is incorrect to say that the study in children of AcuPebble SA100 was at home**. This study was in hospital.	This has now been corrected.
Acurable Limited	103	56	4.2.1	The decision to use cardio-respiratory polygraphy at home as the reference test in Devani's work is overly simplified, taken out of context, and potentially misleading in the explanation given by the EAG (third paragraph "Because of its routine use in the study center for diagnosing sleep disordered breathing"). That was not the reason to choose cardio-respiratory polygraphy. That was the reason to choose that model of cardio-respiratory polygraphy system (i.e..	We have amended the paragraph and changed the text relating to the Devani study as follows: "The study was designed to represent the conditions in which AcuPebble is typically used in practice, i.e. the home environment. The company stated that this was a requirement necessary to obtain regulatory approval. Respiratory polygraphy, which is a

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Embletta). This study was a regulatory study, and a regulatory study should represent as close as possible the conditions of the intended use. The conditions of intended use for these systems is during sleep as natural as possible at home and without an expert next to them making sure they attach everything correctly. It is for this reason that from the regulatory point of view the study should be done at home without a healthcare professional attaching sensors, and the only realistic way of doing this is cardiorespiratory polygraphy, since there is not need from the point of view of regulations of these systems to have the "neurological channels" of PSG. **Please correct this paragraph accordingly. It is wrong and misleading.**	commonly used home test, was therefore an appropriate comparator."
Acurable Limited	104	56	4.2.1	The last paragraph of page 56 is too vague, seems to imply the same for all the studies and this is not true. In the last sentence it says "A limitation of single cohort designs in this context is that the novel device is evaluated in the sleep laboratory rather than in its intended settings (i.e. the patient's home). Sleep study setting is known to influence the diagnostic performance of devices and estimates from laboratory-based studies are not representative of home based studies". This is not true in the study of Devani et al. In fact it is the opposite. This study was specifically designed not to have this limitation. Patients used the AcuPebble device themselves in their own home. **It needs to be clarified that this limitation does not apply to AcuPebble SA100.**	The paragraph in question is a narrative summary of the study characteristics presented in Table 6. As such it is necessary to use general phrasing to compare and contrast the studies. Its purpose is to describe the table, not to repeat it. In this particular example we agree it could be clearer that there were exceptions and we have therefore revised the wording accordingly. "A limitation of many of the studies with this design is that the novel device and comparator / reference standard test are evaluated in the sleep laboratory, rather than the intended setting (i.e. the patient's home)." Table 6 makes it very clear that AcuPebble was evaluated in the home setting.
Acurable Limited	105	57	4.2.1	In the last paragraph of page 57, it is incorrect to say that "Mueller and al went a step further and incorporated the concept of comfort". This concept had also been taken into account in Devani's 2021. See the questions of the usability testing. **This needs to be corrected.**	This has been corrected. This paragraph now reads: "Some studies were designed to assess the efficacy of a novel device in settings typical of those in which it is intended to be used (Devani et al 2021; Alsaif et al, 2023, Mueller et al., 2022). These studies assessed patient usability of the test, comfort levels during testing, and overall acceptability to the patient, amongst other outcomes."
Acurable Limited	106	59	Table 6	**Table 6 (Overview of included studies):** There should be a column indicating whether the sample calculation was obtained prior to the trial. This is important statistically to determine the quality of the outputs. It should also say whether the studies had adjusted the diagnostic thresholds post-obtaining the ROC curves and then from there had estimated the diagnostic performance. The reason for this is that this is equivalent to a non-blind trial.	Table 6 is intended to provide an overview of the general characteristics of the studies., but there is a limit to how much information can be usefully conveyed in one table. The two particular issues raised by the company are addressed in the text of the report: • Sample size calculation – 4.2.1 and 4.3.1 • Post-hoc threshold adjustment – section 4.4, section 4.5.1 and section 4.8.1

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	107	59	Table 6	**Table 6 (overview of included studies) , row of Devani et al, comparator/reference:** It has to say "with manual scoring" and "unattended". It also has to say "patients were not trained on the use of the device under evaluation". In "outcome measures" it should include "accuracy in event classification, including central versus obstructive apnoeas". **Correct to prevent the risk of the table being subject to interpretation.**	These have been added to Table 6 as requested.
Acurable Limited	108	59	Table 6	**Table 6 (overview of included studies), row of Alsaif:** This is not statistically significant. The sample is too low. Eliminate from the table, or state this.	This table describes key study design and methods characteristics; there is nothing of a statistical nature which could conceivably be significant or not. If, however, the intended meaning of the comment is that the study should be removed from the systematic review because the study sample size would not be sufficient to be able to detect a statistically significant result, should there be one, then to do so cannot be justified. Our inclusion criteria did not specify a minimum study sample size, as this would necessarily be an arbitrary threshold. To exclude it from the review on this basis would be similarly arbitrary and unsystematic. However, this is a moot point as we have received final results of the Alsaif study, based on a larger study sample and these have been included in the Erratum. In the previous version of our report, Footnote a in table 6 made it clear that the study sample was the number of patients recruited at the time of the interim analysis.
Acurable Limited	109	60	Table 6	**Table 6 (overview of included studies), row Massie et al:** Have these studies finished? If so, what was the enrolment number in the end? This row is incomplete compared to others. The authors need to have uniformity in the way they present their results. **Correct.**	This information has been added to Table 6.
Acurable Limited	110	60	Table 6	**Table 6 (overview of included studies) Van Pee et al,:** Either this study or Lyne et al probably should not be there because it seems one of them corresponds to a variant of NightOwl that has not received regulatory approval for the EU/UK and hence it is not usable in the NHS. If we were to consider device variants that have not received regulatory approval, this report would need to extend to all academic systems. Alternatively, it does need to consider the extension of intended use of AcuPebble SA100 to children over 1. Our scientific opinion is that **everything that does not have the CE/UKCA mark should not be in this review.** Also, is this trial the same as the one above?	With regard to CE marking and regulatory approval please see our response to comment 100 (Acurable). Van Pee et al evaluated the reusable version, whist Lyne evaluates both reusable and disposable versions. The company (ResMed) report that these are identical except for the battery. The disposable version is intended for use in the UK.
Acurable Limited	111	60	Table 6	**Table 6 (overview of included studies) Lyne et al:** As above. Has it been granted the CE-mark? Note that the current wait for Notified bodies is around two years so being ready for submission does not mean a device will have the CE mark by the time the report is published. If this corresponds to a device that does not have the CE mark, this publication has to be removed. The EAG should not consider devices that have not received the CE mark. Changing from disposable to reusable is not a minor change. It requires a full regulatory assessment. Changing the age of the population is not a minor change either. And changing the diagnostic claims/outputs neither. The EAG is considering at	Please see response to comment 110 directly above.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				least one version of NightOwl which has not received the CE mark. It is also unclear whether they are doing the same with others.	
Acurable Limited	112	60	Table 6	**Table 6 (overview of included studies) Pepin et al:** This row should say that the reported results were the optimum in the ROC curve and hence did not correspond to taking into account blindly the standard diagnostic indexes. There might be other studies in this table for which the same happen, but we haven't reviewed every single one. **Please clarify in the table.**	Please see our response to comment 106 (Acurable)
Acurable Limited	113	60	Table 6	This review might be missing studies. We have not replicated it fully ourselves, but we know of at least one that was a landmark piece of evidence, which led to the recommendation of the AASM not using PAT for the diagnosis of OSA (Ioachimescu 2020). This is further addressed in subsequent comments. This study should be included according to the Inclusion/Exclusion Screening Worksheet in Appendix 2a.	With regard to IOACHIMESCU 2020 please see our response to comment 24, 26 and comment 253
Acurable Limited	114	63	4.2.2	Again the authors of this report make vague generic statements with poor scientific rigour, which can lead to the wrong conclusions for whoever reads the report ("A limited range of characteristics were reported by the studies, with a focus on age, sex and weight). Devani et al did report characteristics that the authors implied studies didn't (for example the ethnicity). Furthermore, **Acurable submitted to NICE a full breakdown of the ethnicity, comorbidities, etc. of the full population of consecutive patients that were recruited in three studies**. None of this has been mentioned in the description in this paragraph. **Correct to be more specific. Generic statements have very little scientific value.**	It is perfectly acceptable to make general summary statements when describing scientific evidence. The text in question is a synopsis of information given in an accompanying table (table 60). The sentence does not imply that Devani didn't report comorbidities, ethnicity etc, it is just saying that most studies focused on a small range of characteristics. Since this section was written, however, we have added a column to Table 60 to report ethnicity/race characteristics. Devani et al was one of a minority of studies which provided any data on this. Whilst our original statement still stands, we have highlighted Devani as being the exception.
Acurable Limited	115	64	4.3.1	**Second paragraph:** Diagnostic accuracy for a system that has to be deployed in the real world cannot be measured in terms of ROC curve, and even less so if the algorithms are based on machine learning (a full scientific explanation of why can be provided on request, but this is not elaborated upon for the sake of conciseness). Hence, whilst this study has got academic value, it should not be included in the context of a health economics evaluation. For similar reasons, none of the outcomes of the study of Martinot are relevant within the context of a diagnostic performance evaluation and hence this is irrelevant for a health economics evaluation. Still, if it is to be included, the reader needs to be continuously reminded of this limitation wherever a result appears, since this is not minor. **Please address this.**	The purpose of the systematic review is to identify relevant evidence on the clinical effectiveness of a health technology to inform NICE guidance. Whilst the review provides a source of clinical effectiveness evidence for the economic model not all of the evidence will necessarily be included in the model.
Acurable Limited	116	65	4.3.2	**Participant characteristics:** The authors state "As is the case with the over 16 population, only a limited range of characteristics were reported by the studies". This is very misleading. **The authors, for example, failed to acknowledge that the study of AcuPebble SA100 had collected comprehensive information**	The phrase "co-morbidities are likely to be ruled out by study exclusion criteria", was in reference to the two Martinot studies, not the NCT04031950. We have revised the paragraph with the wording "Demographic and comorbidity data were provided for study NCT04031950 but were not reported by study centre, except

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				about co-morbidities, and this had been provided to NICE. It is also misleading to say "co-morbidities are likely to be ruled out by study exclusion criteria", when it was known that **this was not the case for AcuPebble's study** and in fact the overwhelming majority of the children had serious comorbidities. **Please clarify and correct to be more specific. Generic statements have very little scientific value.**	for individual patient data of comorbidities. Demographic data were reported for the two studies by Martinot (2015 and 2022). As is the case with the over 16 years population (discussed above), only a limited range of characteristics were reported by these two studies", to make this clearer.
Acurable Limited	117	67	Table 7	**Table 7, Martinot row:** The outcome measures have to specify, that "this is measured with the ROC curve (i.e. the optimum point might not necessarily be the one that uses the established diagnostic threshold"). **Please correct.**	A footnote has been added to Table 7 that states: "Post-hoc analysis was performed to optimize the cut-offs"
Acurable Limited	118	69	4.4	In paragraph 2, the authors rightly identify the risk of bias in studies. This should be indicated in a column on the table. It is important to do this because it is often that clinicians and researchers only look at tables in systematic reviews and this bit of information is missing. **Please add this.**	The use of post-hoc analysis to optimise cut-offs was included in our assessment of risk of bias and concerns of applicability, with the index test domain judged as high risk of bias and high concern of applicability. These judgements appear as a red emoji in the Table 8 therefore there is no need to add an additional column to this table. Please also see our response to comment 106.
Acurable Limited	119	70	4.4, Table 8	The authors assess as low risk Massie and Van Pee's studies. However it is very unclear why in these studies there was a huge number of patients not accounted for**. Please give details about this.**	The number of patients enrolled in Massie et al., 2021 was the same as that analysed. This is clear from Table 6. We have added a statement to the critical appraisal form to further highlight this. For the Van Pee study, reasons for exclusion have been added to the critical appraisal form.
Acurable Limited	120	72	4.5.1	It says the table presents sensitivity, specificity, PPV, NPV and accuracy. Note this is missing likelihood ratios.[This can be elaborated upon. It has not been done because of lack of time due to time constraints imposed by NICE]. **Please add likelihood ratios.**	Please see response to comment 94.
Acurable Limited	121	74	4.5.1.1, Table 9	**Table 9** is missing the CI for the PPV and NPV of Kelly. Considering how small this study is this is important since the results are not very statistically significant.	The CIs reported by Kelly et al have been added, and 'NR' added where they have not been reported.
Acurable Limited	122	76	4.5.1.2	**Second paragraph on this page:** This is the first time the authors hint to the fact that **the trial reported in Van Pee is for a device that doesn't have regulatory approval in the UK**. If it doesn't have regulatory approval, it shouldn't be in the review to start with. Note that making a device reusable generally implies massive changes to hardware which can impact safety. It is also unclear whether the algorithms are the same. Notified body waiting times are in the order of two years right now. So, **there is no justification to include in this review devices that have not received regulatory approvals. Please remove this device from the analysis.**	See our response to comment 100 (Acurable). Van Pee et al evaluated the reusable version, whist Lyne evaluates both reusable and disposable versions. The company (ResMed) report that these are identical except for the battery. The disposable version is intended for use in the UK.
Acurable Limited	123	76	4.5.1.2	**Third paragraph on this page:** The authors state that sensitivity and specificity are the performance parameters used to inform the economic model. This is not scientifically rigorous. Sensitivity and specificity have been widely demonstrated not to be optimum statistical metrics to evaluate the performance of new medical devices, and specifically to evaluate technologies for OSA diagnosis (references can be provided for this). Hence these are not adequate inputs for the model. More appropriate inputs would be predictive values, or ideally likelihood ratios.	Please see response to comment 94.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	124	76	4.5.1.2	**In the last paragraph of this page, the authors say**: ”In notable cases () available confidence intervals for sensitivity and specificity are wide, indicating greater uncertainty in the estimates. This may be due to relatively small sample sizes…”. This statement should read “due” rather than “may be due”.	Text updated as suggested.
Acurable Limited	125	76, 77	4.5.1.2	In the paragraph spanning pages 76 and 77, the authors do not mention to take into account in their model differences between studies that adjusted thresholds post-hoc (based on ROC curves) or those that didn’t, in the same way they didn’t include confidence intervals in their models. Without this any conclusion is flawed. Just as an illustration, a study could be done with two positive patients and a ROC curve could be plotted to see which threshold could be chosen so that the two patients would be positive. That study would then have 100% accuracy! **The model needs to include somehow an adjustment for those devices for which the performance has been obtained on post-hoc, not clinically conventional thresholds. If this is not possible, this information together with a warning of caution needs to be everywhere that any number appears obtained from those performance parameters**.	The company is correct that the model does not account for this. The limitations of the results from the studies where post-hoc optimisation has been used, are highlighted in discussion of the accuracy evidence, and in the main Discussion/Limitations (Section 6.3.2 now section 6.2.2).
Acurable Limited	126	76, 77	4.5.1.2	It is not correct to imply there is no statistical way of comparing devices in terms of superiority. There certainly is (a different thing is whether it has not been done). The devices do not need to be used in the same study. **This statement should be removed.**	We point out that there is no statistical comparison of the devices. We do not say that this cannot be done: *“The devices have not been formally compared in the same study with the same population and there is no formal statistical analysis to confirm any differences or equivalence between them”.*
Acurable Limited	127	78	4.5.6	The second paragraph (in bullet points) compares failures of novel devices and comparators but does not mention the AcuPebble study. Why does this paragraph not describe the fact that there were no failed tests due to the novel device for AcuPebble reported in Devani et al, whereas all the failed tests corresponded to either the reference test or failure in the protocol? Also, there were situations in which the patient forgot to do the test, but that does not correspond to a failure of the device. **Please mention this.**	The paragraph being referred to compared the total proportion of test failures which were related to the novel device and those related to the comparator device. The total proportions could include technical failures, user related failures, or ‘other’ failures, as applicable. Thus, our definition of a failure was broader than that used by the company, which explains the difference in estimates between our report and the company. We have since revised the failure rates used in the economic evaluation (section 5.7.5 of the Erratum). Please also see our responses to related comments 128, 162 and 174 (Acurable).
Acurable Limited	128	78	4.5.6	**Last paragraph:** The authors count patients sleeping less than 4 hours a failure of the test. This is misrepresenting the technologies (both novel and reference) and the test itself. The patient sleeping less than four hours is not a failure that should be attributed to the devices in the context of this health economics evaluation. **Please correct, not just here but in any analysis that has been done considering this. This is scientifically flawed, and is resulting on wrong inputs fed into the model that will inevitably result in wrong outputs.**	Thank you for this comment, we have revisited the failures reported in Devani for the AcuPebble device. For the model, we are only interested in failures that could happen in practice that incur a cost to the NHS. Thus, of the 12 failures considered in the original model for AcuPebble from the Devani study, only 1 of these (participant left phone in another room) is included in the Erratum. Please see Section 5.7.5 of the Erratum report for more details. Note that we have reviewed the failure rate data for all devices (please also see response to comment 26).

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	129	79	4.5.6	First paragraph: The authors are counting as test failure the wrong device being given to the patient. How can this be a test failure that has got any relevance in the context of an economic evaluation for the assessment of new technologies? If nothing else, this is a pathway failure, and having easy-to-use devices that hospitals don't need to store as capital equipment and can be sent to patients directly will only serve to reduce the consequences of this time of "healthcare delivery error". Hence feeding this into the economic model in a quantifiable manner accounting for a test failure is a wrong assumption. **Please correct, not just here but in any analysis that has been done considering this. This is scientifically flawed, and is resulting in incorrect inputs fed into the model that will inevitably result in incorrect outputs.**	Please see response to comment 128 above.
Acurable Limited	130	79	4.5.6	The analysis regarding test failures is not correct. The authors are mixing failures that should not be mixed (more in subsequent sections), and basing the assessment of superiority or inferiority on just a number is incorrect. For example, some devices might have a higher number of test failures, but if those devices enable the patients to repeat the test the following night without any additional burden to the healthcare system, whereas PSG would not make this possible then that device would still be superior in this respect. **Please correct, not just here but in any analysis that has been done considering this. This is scientifically flawed, and is resulting in incorrect inputs fed into the model that will inevitably result in incorrect outputs.**	Please see response to comment 128 above.
Acurable Limited	131	81	4.7.2	**Ease of use and acceptability:** The authors are underplaying the outcomes of the usability study by saying "patients found the novel devices easy to use and sleep quality was good". In the study of Devani, the patients showed a very strong preference for this compared to cardio-respiratory polygraphy. **This needs to be stated in the text and ideally should be taken into account in the model.**	We have added further detail on ease of use and acceptability from Devani et al, Mueller et al and Alsaif et al.
Acurable Limited	132	84	4.8.1	**Last paragraph:** This paragraph lacks scientific rigour since it just gives two reasons why those studies should not be compared which are not the most important ones. The most important reason is that in one of the studies there is post-hoc optimization of the thresholds, whereas in the other one there isn't. **This should be added.**	The final sentence of the preceding paragraph explicitly states that caution is needed due to the post-hoc optimisation of diagnostic thresholds. It would be unnecessarily repetitious to mention it again directly below
Acurable Limited	133	84-85	4.8.2	This study is not as relevant in the context of this review and can be confusing for the reader since what is demonstrated is not diagnostic accuracy, but rather the relevance of a particular physiological signal. The authors of the report themselves acknowledge this but still keep it in the report. However, in any scientific publication there needs to be completeness and if this kind of work (i.e. not diagnostic accuracy) is going to be presented, the review should also have included all academic publications that demonstrate any kind of correlation between physiological signals and OSA. **Since this would be distracting and**	The study (Martinot 2015) met the inclusion criteria for the systematic review. Studies often vary in terms of the depth and breadth of data they report and its relevance to a systematic review, this study as we acknowledge is less relevant than others, but nonetheless it meets the inclusion criteria. Given the paucity of available evidence in the paediatric population the study will be informative.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				unnecessary, to keep the scientific thoroughness of the report it is recommended all this discussion is eliminated.	
Acurable Limited	134	85	4.8.3	Once again the authors are mixing in a number that they call “test failure rate” tests that were not performed for different reasons. These reasons should not be mixed because: 1- Some of them are not “test failure rate” (i.e. it is not the test’s fault that the patient ended up sleeping less than 4 hours); 2- Others were tests not performed due to failure following the protocol. Again, this is inherent to the protocol of the trial itself and hence will not affect or be representative of real world implementation. Furthermore, due to the characteristics of the devices, mostly for those that are reusable, having a test that one night does not work is not necessarily a problem if the test gives the option to the patients of repeating it for free and warns them the test needs to be repeated (so that patients do not return them to the hospital), something that cannot be done with RP or PSG because the cost is in capital equipment and if they are not Cloud based or have signal quality software incorporated on them, subtle reasons for failure (less than 4 hours of sleep) cannot be identified. This distinction also needs to be made quantifiably to consider this as inputs for the health economics study, since it is not the same a system that automatically identifies failures and informs the patient straightaway of the need to repeat the following night, that a system that needs to be taken back to the hospital for the failure to be repeated. **Correct the inputs corresponding to test failures. This probably applies to all studies/systems. Not every study that was not completed in the context of a clinical trial is relevant in the context of the real world implementation or this health economics analysis.**	Please see response to comment 128 above.
Acurable Limited	135	86	4.9/4.10/4.11	Why are the demographic characteristics of the children included not in this section, or at least summarised?	The demographic characteristics of the children are presented earlier in section 4.3, specifically section 4.3.2. Sections 4.8 to 4.10 report the results of the studies.
Acurable Limited	136	86	4.11	The second paragraph states “This study is comparing the simultaneous use of AcuPebble and PSG or AcuPebble and polygraphy, however, the setting for these tests is unclear”. Please could it be clarified what was unclear about it, that led to completely ignoring information that was key for the model? NICE was provided with the full information about this study as well as the results and extensive demographic/co-morbidity data. However, this has not been even mentioned in this document. Also, if there was any doubt in the EAG about any of this, why wasn’t the manufacturer approached to ask for information, as was done with others? All the relevant information needed for the model is available and has been available before the deadlines imposed by NICE, and if there had been any question they could have approached the manufacturers (as was done with others) which they never did. **The report needs to be modified to account for this.**	Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024))

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	137	87	4.12	The EAG claims that only one study was done in real-world conditions. However, “to provide fair comparison” with the other studies the lab conditions should be considered. This is very much debatable since ultimately one could also argue that some kind of input should take into account the fact that the other studies did not take place in real world intended use environment, which is widely established by regulators to be much harder (this is the reason why, for example, there are separate, tougher regulatory constraints for those devices that are intended to be used in the home environment. Also, this is not scientific. This is not about doing a “fair comparison”. It is not a competition. It is rather to inform what would be best for patients. **Please remove this rationale and statement.**	This section has been deleted from the Erratum following the decision to remove adjusted analysis of the Devani et al study. Please see our response to comment 138 (Acurable) directly below.
Acurable Limited	138	87	4.12	**The EAG has made up performance data for AcuPebble SA100** on the basis of “fairness to other devices”. This is both ethically unacceptable (considering the influence a NICE report has on clinical practice) and scientifically wrong. It is ethically unacceptable because **NICE had been provided with the data from a PSG study with AcuPebble** (which was being prepared for publication), **but this data has been completely ignored by the EAG**. In fact, the report is written (as is obvious from previous sections/comments) in a way that **hides the fact that the data had been provided, which is unacceptable, especially when the same has not been done for other studies of other manufacturers**. This is even less acceptable when it is clear that the data was needed for the economic model. The made up data (using a flawed argument for justification), which does not represent reality (as proven by the evidence provided, obtained from a powered clinical trial, that they decided to ignore) involves a ~30% performance reduction, which clearly is going to result in economic outputs that do not represent reality either. **This is shocking both scientifically and ethically, mostly coming from a group of scientists.** This needs to be modified and reviewed by the manufacturers before any further publication. Even if the EAG had not had any data comparing AcuPebble to PSG **(which they had)** the scientific argument is completely flawed. We have an extended scientific analysis as to why, but we will comment on this here briefly: Whilst we acknowledge that in-lab PSG is commonly regarded as the gold-standard, it is not acceptable to use a **novel and inapplicable correction formula**, using a single and arbitrarily chosen paper with acknowledged methodological issues (**full scientific evaluation and demonstration of this can be provided on request**) and risks of bias, to decide not only AcuPebble’s clinical efficacy, but its entire economic evaluation. On top of that: 1- **The EAG has completely taken out of context and wrongly applied the information and methodology in the paper they have used for the “adjustment”.**	Given the inclusion of the Phase 1 Virgen Macarena Trial* in the systematic review and economic base case, there is less imperative to adjust the diagnostic performance of AcuPebble SA100. Phase 1 Virgen Macarena Trial assesses the performance of AcuPebble SA100 referenced to PSG, thus providing a direct comparison to inform an incremental cost effectiveness analysis of novel devices. This is in-keeping with the NICE health technology evaluation manual’s general preference for direct comparison of health technologies (i.e. within the same study) over indirect comparisons (i.e. between two or more studies). It should be acknowledged, however, that the Phase 1 Virgen Macarena Trial is based in a hospital sleep laboratory in contrast to the Devani et al study which was home based and therefore directly relevant to the decision problem. The adjusted analysis we previously reported therefore no longer informs the assessment of cost effectiveness and has been removed from the Erratum for the reasons given above. However, we make the following points in response to the company’s critique. 1. It is appropriate to use statistical adjustment methods when there is a lack of suitable evidence to facilitate a connected evidence network. We reviewed several alternative methodologies and selected the approach published by Sherwin as our preferred approach. 2. We acknowledge that there was an error in our application of the adjustment method, and we appreciate the company bringing this to our attention. 3. Whilst the adjustment methodology tested by Sherwin (2022) was published only recently it is based, in part, on previous published methodological studies in this area extending as far back as 1966. In contrast to earlier work, Sherwin’s derivation of formulas is purposefully transparent in documentation, to permit independent replication. 4. It is unreasonable to judge the methodological soundness of a recent published study using citation rates or journal impact

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				2- They paper used has been cited 0 times which proves that is not a commonly accepted methodology; 3- The journal is a Q2 journal indicating low quality of peer review. Further notes as to why this is flawed can be found in the next comment (093). 4- Perhaps most shockingly, **even the formula used is the wrong one**; ie the EAG has used the incorrect formula from the paper chosen to "adjust" specificity and sensitivity, suggesting that the paper was not read properly or understood before applying this arbitrary methodology. It is also worth noting that whilst the EAG decided to arbitrarily bring down the results of AcuPebble SA100, it did not do the same for all those systems in which diagnostic thresholds not corresponding to the clinically followed ones (and with resolution of up to two decimals) had been determined post-hoc of seeing the data in the clinical evaluations. **The whole document and inputs to the health economics model must change to consider the right data obtained from the clinical study for AcuPebble SA100. The only thing that should be kept is the failure rate (re-calculated since that was also wrong. See the comments regarding this) corresponding to the evidence of the home study since that will be more representative of the real world scenario. Any questions manufacturers can assist.**	factors, given the inevitable time lag until citing publications themselves can be written and published. Indeed, judging publications by their citation rates in general is becoming increasingly discouraged in the sciences, in favour of more meaningful evidence of impact and value. In this case ultimately it is for the diagnostic advisory committee to judge the appropriateness of a given methodology, based on whether it gives them confidence in the results to enable them to make the right decision on the use of a health technology. * Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024))
Acurable Limited	139	87	4.12	Further notes can be provided regarding why the methodology used to adjust the performance results is scientifically flawed (although note that this is irrelevant considering no "theoretically extrapolated data" should be used for the comparison considering there is real world data), however something else to note is that: The paper chosen as reference for the comparison between RP and PSG, Xu et al., uses Nox T3 rather than Embletta, which was the reference test in Devani et al. The data used from this paper by the EAG is also for non-simultaneous RP and PSG, which allows the effects of inter-night variability to obscure any true diagnostic differences.	With regard to the adjustment to compare AcuPebble to PSG please see our response to comment 138 (Acurable).
Acurable Limited	140	89	5.1.1	The EAG is comparing the economic studies of the new systems to oximetry and RP. Oximetry should not be highlighted as a potential diagnostic method. Whilst some centres in the UK might still be using it, new overwhelming evidence has appeared showing the **health inequalities** caused by oximeters since the engineering/physical principles they are built on and the regulatory framework behind them does not take into account that the former do not work in dark skinned people, and there is a loophole in the standards that allows all oximeters to pass without having to pass performance (As a note: even for white people, the standard only requires 4% accuracy in completely still conditions of a healthy subject, in 67% of readings. Hence, using oximeters in isolation considering the	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				4% margin of error should be questionable). **NICE should not be promoting this**, and the way this document is written it implies that oximetry is a viable diagnostic approach.	
Acurable Limited	141	90, 91	5.1.3	**Bullet points:** All of the ones in which oximetry is used for diagnosis should be eliminated because this is equivalent to **promoting health inequalities.**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	142	93	Table 12	**Details of economic studies of interest:** Row of Phua et al: This is not relevant. It is based on a completely different healthcare system with different dynamics. Hence it should not be considered in any way for this. Row of Di Pumpo et al: Same as above. This is not relevant. It is based on a completely different healthcare system with different dynamics. Hence it should not be considered in any way for this. Geessinck et al: Please indicate location. Guest et al: Please indicate location.	Potentially useful studies were not limited by country, therefore we have kept all studies in Table 12, but have updated the table indicating the country in which these studies are based.
Acurable Limited	143	94	5.2.2	For scientific completeness, could the authors provide a rationale as to why the mentioned outcomes are chosen?	This rationale is provided in the Inclusion and Exclusion criteria in Appendix 7, Table 68: these outcomes are chosen as they can all be mapped to the EQ-5D (which is the preferred outcome for NICE's reference case for economic evaluations, please see Section 5.4.3).
Acurable Limited	144	99	5.3	The University of York carried out a health economic study in which severe and moderate are combined in the same way as NG202. This information was not provided to the EAG, but in view of the fact they have considered a model from NightOwl (Resmed), this should be considered too. The manufacturers (note this was independent, not sponsored) can be given access to this model to share with NICE if required.	As part of their company submission ResMed submitted a decision tree model, which the EAG reviewed. We have not received models from any other companies.
Acurable Limited	145	100	5.3	Once again, the EAG is considering a device that is not approved by the regulatory authorities to be used in the UK. This assessment should restrict itself to devices that are approved. Other manufacturers have data for devices that are also going to undergo regulations and that has not been submitted. Also, considering the regulatory landscape at the moment, it could take a couple of years until this is in the market. Regardless of the time, until the devices are regulated it is as if the did not exist for the public and hence should not be considered. **Please remove any reference or analysis to devices that have not been approved to be commercialised in the UK yet.**	With regard to CE marking and regulatory approval please see our response to comment 100 (Acurable).
Acurable Limited	146	101,102	5.4.1	In the last few sentences of 101 and first sentences of 102: The authors state that there is lack of evidence over the long-term impacts of OSAHS in children. However, they only give one reference. How could this have been judged with just one reference? There is plenty of literature that should have been reviewed,	See response to comment 227

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				for example, on the lack of performance at school as a result of tiredness. Has this been researched? **Please add more references.**	
Acurable Limited	147	102	5.4.2	The authors state "We have included home pulse oximetry as a comparator, should home respiratory polygraphy be limited". Again, for the reasons explained before **(health inequalities associated with oximetry, recently discovered as a result of COVID)** this should not be stressed here or be given as an option.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	148	104	5.5.2	The inputs to the decision tree listed have been wrongly calculated and hence render the whole results invalid/wrong. More specifically: 1- Sensitivity and specificity are statistically obsolete metrics to evaluate diagnostic methods. Yes, they provide some value but they can be misleading because of the way they are defined (further scientific evidence supporting this statement can be provided. On the basis of this the model should have used either predictive values or likelihood ratios.) 2- In the case of AcuPebble SA100, **the EAG "invented" the results**, **as opposed to using the evidence**. More specifically they have reduced the performance metrics in some cases up to ~30%, which clearly is going to have a massive effect on the results of this, and ultimately any output associated to AcuPebble SA100 will be wrong. **The evidence was provided**. There is no justification not to use the PSG validation data, neither scientific nor ethical. **Please modify this.** 3- **Failure rates have been calculated wrong** (maybe for all the devices, but certainly for AcuPebble SA100). Hence the outputs and conclusions of this model are also flawed because of this. The rationale as to why this calculation is wrong (on the basis that failures from the protocol, incomplete tests due to physiological factors (i.e. patient sleeping less than 4 hours), and failure of the novel test) cannot be lumped together, since the effect this is going to have in the output is detrimental to the novel devices when in reality the novel devices provide advantages in many cases in relation to tests that failed for other causes. This needs to be recalculated for every device, considering the actual reasons why there is no output and whether those are a "failure" considering actual real world use case and the features a number of devices have that RP doesn't, and hence would allow, in some cases, for tests to be repeated the following night without any cost to the healthcare system or meaningful delays. **Please modify accordingly throughout the entire report.** 4- Data privacy is an important issue to consider. Some of the systems force patients to provide personal sensitive information and sign a disclaimer so that they are willing to share this. Note that this data is needed to be able to "feed algorithms" since these are based on ML. This needs to be discussed somehow because this goes against data privacy rules that are now pervasive in the NHS. Since every Trust follows a different process some estimation of percentage of Trusts that would be ok with this could have been incorporated into the model. **Please, find this information and discuss this in the report.**	With respect to each point 1, please see EAG response to Comment 94 2, please see EAG response to Comment 57 3, please see EAG response to Comment 128 4, please see EAG response to Comment 177

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	149	104	5.5.3	This model is obsolete considering **it is not taking into account as an input health inequalities**, which: 1- In the context of OSA diagnosed with oximetry is very relevant. Whilst this might have still been ok when the model was developed for CPAP, is not justifiable now when the world has become aware of the limitations of oximeters and there are worldwide investigations, enquiries and modifications of regulatory processes to determine what to do about the widespread use of oximetry; 2- It is also relevant in the context of novel technologies since they allow diagnostic access to those who are more vulnerable, live in rural areas, are less educated, etc. **Oximetry should not be considered as a comparator.**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	150	105	5.5.3	**Top paragraph:** This is again very simplistic. The quantifiable consequences of OSA go well beyond cardiovascular events. This might not have been the case a few years ago, but the evidence now is overwhelming. Focusing this study only on cardiovascular events and road car accidents is not only scientifically lacking but also misleading for the public. Furthermore, it will just lead to biased outputs of the model. **If this is going to be restricted to this please at least discuss the others properly, with references.**	Please note that consequences such as loss of productivity at work or school are beyond the remit of the NICE perspective. With respect to consequences such as fatigue and depression please see response to comment 167.
Acurable Limited	151	106	5.6.1	The authors refer to the fact the failure rate they consider will negatively impact a model (with other words). **This would be fine, except that the calculation of the "failure rate" is wrong, at least for AcuPebble SA100 (see other comments in this document). Hence, the output of the model is wrong. This needs to be corrected.** The authors also refer to the sensitivity and specificity being considered. Again, **the data they are using for sensitivity and specificity is wrong. In some cases data is made up** (with the false argument that they didn't have access to comparable information), reducing the systems' performance. I**n others, performance is inflated by ignoring the fact that data that was extracted with a post-hoc adjustment of the diagnostic thresholds** (for example, the authors quoted the sensitivity and specificity after having produced the ROC curve and determining for which diagnostic threshold, different to 5/15/30, the best sensitivity and specificity would be obtained. In some cases they adjusted not just the integer threshold but also up to two decimal places). **In other cases they took values of systems that are not approved by regulators to be in the market in the UK. All of this needs to be corrected.** And it is well established sensitivity and specificity are not the right ones to validate medical devices for diagnosis of OSA (scientific evidence for this can be provided)	With respect to the point on failure rates from the Devani study, please see our response to comment 128 (Acurable). With regard to the adjustment to compare AcuPebble to PSG please see our response to comment 138 (Acurable). With respect to studies using post-hoc optimisation, the limitations of the results from these studies (where post-hoc optimisation has been used), are highlighted in discussion of the accuracy evidence, and in the Discussion/Limitations (Section 6.3.2). With respect to the point on using sensitivity and specificity in the decision model, please see our response to Comment 42.
Acurable Limited	152	113		The authors admit that the model can account for population specific risks (and they do it- albeit wrongly, as will be explained later on in the document- to differentiate men and women). **However, they make no allocation to**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				differentiate risks of ethnic minorities or those with a dark skin, which in the context of OSA is very relevant. If it is not possible to do this because they have no data, t**he report should have a name that takes this into account and make a full disclaimer that this only refers to a white population. Anything else is potentially discriminatory.**	
Acurable Limited	153	115	5.7.3	The authors say that AcuPebble accuracy would be overestimated if there was a PSG study and this is the reason why they artificially make up performance results. First of all, there is absolutely no scientific grounds in what they are saying because: 1- There are many reasons behind potential differences between PSG and in-home RP which do not directly translate to accuracy, and the EAG has not investigated what those reasons are. We would be happy to create a full scientific document showing why deviations in accuracy happen for every single source of evidence they present us with. But based on what is written in this document there is nothing beyond a vague comment. Just as illustration of different reasons: 1-studies that show differences between PSG and RP are often done in different days, and it is well known that there is night-to-night physiological variability. 2-They are often done with different systems; and it is known that there is variability between systems (even if they are different variants from the same manufacturer). 3- The indexes are calculated based on real/observed sleep time in hospital and then based on recording time for the home one. And in some cases this does not extrapolate to new technologies because the novel technology (as is the case of AcuPebble) does internally calculate the sleep time and hence it is comparable to the in-hospital system. 2- The authors have got no idea of how AcuPebble works or the safety features the algorithms have built inside. Hence they cannot extrapolate results obtained in questionably extrapolatable studies based on comparing PSG to RP to what would happen in AcuPebble. 3- NICE was provided within the deadlines with data from a registered clinical trial (That had already finished the PSG cohort of patients) demonstrating the real accuracy values when comparing with PSG. However the EAG seems to have chosen to ignore this. 4- The EAG has completely misunderstood the maths of the paper they quote for the extrapolation, and used the incorrect equation from it, and hence their results are wrong (although this is irrelevant when considering they had the actual data and they didn't need to use a paper with zero citations published in a Q2 journal to make up performance numbers for AcuPebble SA100) **On the basis of this, the output of this exercise for AcuPebble SA100 are completely wrong, misleading and unethically presented. This needs to be corrected.**	With regard to the adjustment to compare AcuPebble to PSG please see our response to comment 138 (Acurable). Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024))

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	154	115	5.7.3	**In the third paragraph:** It is scientifically incorrect to mix data from different devices from the same manufacturer. If those devices had to undergo a regulatory process it means that they were not minor variations from one another. This needs to be investigated and proper scientific and regulatory grounds for taking this evidence needs to be provided.	Please see comment 24 from the manufacturers of WatchPAT 300 and WatchPAT ONE, where they confirm that these features of these devices are identical.
Acurable Limited	155	116	5.7.3	**In the first paragraph of page 116:** The EAG seems to have a double standard. For Sunrise it is assumed that the study done at home would have similar accuracy than the one in hospital. Again, no properly investigated scientific grounds for this with respect to RP. Major reasons for the differences in diagnostic output between RP at home and PSG in hospital are: 1- Night to night variability; 2- Different denominator conventionally used in the formulas (leading to it being quantified as recording time for studies done at home with RP), whilst in hospital neurological channels are supported by constant video monitoring; 3- Different sleeping positions caused by different systems; 4- More artefacts resulting from patients not being able to have a natural sleep in hospital; 5- Sensors are re-attached by technicians in the middle of the night in hospital whereas they are not at home. However, many of these reasons also apply to PSG at home. Still, the EAG decided to massively bring down artificially the numbers of performance for AcuPebble SA100, with flawed scientific arguments, but when it comes to Sunrise those numbers have been kept. Furthermore, Sunrise adjusts the thresholds post-hoc and this is not taken into account. Note also that AcuPebble SA100, and Sunrise Kelly's trials are the only ones that were carried out in the intended use environment, but this is not accounted for as inputs of the model.	The study by Kelly reports home PSG as the reference standard (rather than home RP). Our assumption is that home PSG is equivalent to laboratory PSG. Please note that we do not use the data from Kelly in the base case, but only in a scenario analysis. With regard to the adjustment of Devani, please see EAG response to comment 138. The company is correct that the model does not account for the use of accuracy data where post-hoc optimisation has been used. We do highlight the limitations of this in the Section on accuracy data (Section 5.7.3), and in the Discussion (Sections 6.2 and 6.3). With respect to the final point that there is no accounting for studies conducted in the home in the model: in the report we keep results separate for those where accuracy data used were obtained from a study in the home vs those in the clinic. We note that in the Erratum, the accuracy data used in the base case for all devices is taken from studies conducted in the clinic (not the intended use environment).
Acurable Limited	156	117	5.7.3	The authors seem to have ignored the landmark independent study of WatchPAT, which has been reported in the Journal of Clinical Sleep Medicine (*"Performance of peripheral arterial tonometry–based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort," Octavian C. Ioachimescu, MD, PhD, J. Shirine Allam, MD, Arash Samarghandi, MD, Neesha Anand, MD, Barry G. Fields, MD, MSEd, Swapan A. Dholakia, MD, Saiprakash B. Venkateshiah, MD, Rina Eisenstein, MD, Mary-Margaret Ciavatta, PAC, Nancy A. Collop, MD*) , which comprised 500 patients, and on the basis of which the American Association of Sleep Medicine issued a recommendation not to use PAT for diagnosis. **This study is important in the context of this work, because amongst other things, 72% of the population is African/American and hence has dark skin, which reiterates the issue with using any current technology based on PPG and with no flow indicator to diagnose OSA (mostly in dark skinned subjects). And in the context of this economic evaluation this differentiation should be made (i.e. two different**	With regard to IOACHIMESCU 2020 please see our response to comment 26 and comment 253 Please also see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				populations should at least be considered), mostly considering the review takes oximetry as a comparator, and two of the novel technologies are PAT based. Either the full report is changed to say that these findings only apply to male individuals with fair skin, or any result of any device based on signal extracted with PPG must have a note cautioning of this.	
Acurable Limited	157	118	5.7.3	**Top paragraph on this page:** the authors state that "to consider the uncertainties and limitations in the diagnostic accuracy estimates for the novel devices as discussed above and earlier section 4, we conduct a 'worst case' scenario analysis using the lower-bound of the 95% CIs for the accuracy estimates in the model." Whilst this approach has some merit, the way it is written is misleading. In order to put it in context three individual outputs should have been generated, one for the quoted performance metric, one for the lower bound and one for the highest bound. It is unclear whether the ranges in the tables correspond to this. **Please could what has been done be clarified in this paragraph.**	As suggested in the comment, we have undertaken an additional scenario analysis, assuming the upper 95%CI limit for sensitivity and specificity for all novel devices. Please see Section 5.10.2 and Appendix 9a.
Acurable Limited	158	118	5.7.4	The EAG justifies the fact that they used a piece of work carried out in oximetry in 2010 to justify the oximetry numbers, when there is an overwhelming amount of new evidence showing that any previously collected evidence in any application of oximetry needs to be questioned and potentially discarded if the performance values were not split between fair and dark skinned individuals. Taking this into account, this section is not just scientifically questionable but can lead to **very negative health inequality impacts. This needs to be corrected.**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	159	118	5.7.4	**Last paragraph:** The authors refer to studies evaluating accuracy of RP. However, accuracy of what with respect to what? Manual marking? Automatic scoring followed by interpretation? Semi-automatic scoring? It does make a difference. Also, the US AHRQ carried out a systematic review and meta-analysis in which they carried out a thorough evaluation of the quality of the evidence. The EAG could have relied on that since, although it is a bit old (2015) it is far more comprehensive that what they present in this work. **This needs to be scientifically unequivocally presented with references and no vague statements.**	In response to this comment, we have provided more detail in the report on the studies conducted by Xu and Pereira. Please see Section 5.7.4. As the NG202 and Khor 2023 SRs were conducted more recently than 2015, we have referred to these sources (NG202 and Khor 2023).
Acurable Limited	160	119	5.7.4	The basis of comparison between PSG and RP is potentially scientifically flawed because it is ignoring how the comparisons were made. For example, they take one study carried out in China (i.e. with a population that is different in nature to the UK population) to establish the "truth" for this report in terms of the performance of RP (and in fact the same study is used to artificially bring down the performance results of the regulatory clinical trial of AcuPebble SA100 carried out in the UK). However: 1- It is not stated exactly how the performance was measured/compared (was automatic scoring used? Did they use total sleep time for one and recorded time for the other?, etc…);	In response to this comment, we have provided more detail in the report on the studies conducted by Xu and Pereira. Please see Section 5.7.4.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				2- They don't take into account variability on different nights; 3- They don't consider the fact that in the home environment there was no one to attach the sensors and the artefact/signal loss would be different in different systems for different nights, in different environments, and taking into account patients attaching many sensors themselves, which are notoriously known for coming out of place. **The limitations need to be properly scientifically discussed.**	
Acurable Limited	161	120	Table 20	**Sensitivity and specificity estimates used in the model oximetry and RP:** This table further reiterates the fact in the previous point. It can be seen how in Xu et al, the specificity under 5 is significantly lower, which can be easily explained by different nights' variability since for low AHIs this is going to have a higher effect. Hence using Xu's report to extrapolate performance for other methods is a scientifically poor decision. The final section of the table is meaningless for the reader. Where do those numbers come from? (it is not clear from the text either). What is behind the drop in specificity? Why are there no confidence intervals? Considering how little evidence there is, how simplistic the explanation is, how questionable the other rows of this table already are, and how different these numbers are, **they must be explained for this to be scientifically rigorous, or removed completely.**	We describe in section 5.7.4 why our base case uses data from Xu et al, even though it is not ideal, Xu "was chosen as it evaluates RP in the home-setting, is compared to in-laboratory PSG, uses a named device (Nox-T3, that one of our experts noted as being representative of what is currently used in England), and is one of the most recent studies that we considered". We have edited Table 19 in an attempt to improve understanding for readers.
Acurable Limited	162	120	5.7.5	The numbers reported for test failure rate in RP based on personal communications seem to be well below what has been reported by many other sources. Test failure rates, in many cases, depend on what different sleep services account as "failure" and also how the sleep services are run. Hence the numbers cannot be decoupled from this explanation. A sleep service in which patients have to come to hospital, are trained, and do a "rehearsal", is not the same as a sleep service in which technicians go to a patient's house, or those which are overloaded and rely on temporary staff so that patients may not receive complete support. It is also worth bearing in mind, there are some incentives for providers not to have high failure rates, which results in undercounting. **Could numbers be given to support these statements? (considering how different they are from others reported around the world). How many patients in which period were sent home with RP? What training did they receive (ie. what was the pathway?). What was considered a failure rate? How many failed? How many centres were considered? What were the demographics of these centres?** Furthermore, it is interesting to note that the EAG decided to follow an undocumented conversation to estimate the failure rate, when they had at least one journal paper they could have extracted this from, and they didn't. In Devani et al, it was reported that 16/182 RP tests failed in the regulatory trial, and the reasons to consider failure are also given. That is 8.8%, as opposed to the 5%	We have provided more information and rationale on the data used in the base case for failures of RP. Please see Section 5.7.5 in the Erratum. Please also see Table 62 which gives further detail on failure rates. In addition to the one-way sensitivity analyses, we have also added a scenario analysis where we take the highest estimate of failure for RP that we found in our included studies: [redacted] from the Alsaif manuscript (see Section 5.7.5). This scenario increases the INMB estimates for all novel devices relative to RP but does not change the ranking of INMB between the novel devices (see section 5.10.2.2 and Table 46).

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				they quote. These were patients however that had been trained for about half an hour individually on the use of RP, something that many sleep services don't have the resources to do in the conventional diagnosis pathway. **This needs to be further investigated and supported with actual evidence, not what currently comes across as an anecdotal conversation.**	
Acurable Limited	163	121	Table 21	**Estimates of failure rates used in the model:** There is data here that is factually wrong. The numbers have not been double-checked by us for all the systems, but the data for AcuPebble is completely taken out of context and is mixing things that should not be mixed considering the aim of this health economics study. More specifically, the EAG has considered 6.74% represents the test failure rate of AcuPebble, when this is not correct. Out of the 12 studies that the EAG considered as failures to compute the 6.74%: 1- 7 patients forgot to do the study. This should not be an input for the model because this was a limitation of the validation protocol, not the system. Patients had to return the system to the hospital the following day in the clinical trial. Clearly if they forgot they also would not repeat RP considering how cumbersome this was (and also that the RP system was set up for one night). But in real world use of AcuPebble SA100 there are no constraints associated with when patients have to do the test. If they forget one night they can do it the following one. 2- 4 patients were given the wrong kit for the study. This is not a failure of the system but a limitation of the protocol since if a patient was given the wrong kit there was no option to course correct. In the real world, in the very unlikely scenario the patient was given the wrong kit this can be corrected by post with negligible delay. Note that in a clinical trial where patients are also following the conventional diagnostic pathway the level of attention to "giving the patient" the right kit for the trial is significantly less than it is when that kit is the one they are going to be diagnosed with. 3- 4 patients "played" with the app before going to sleep logging themselves out, and they had not been given the password to log back in). Again, this was a failure of the protocol not the system since it is not a problem that can happen with the commercial system because the password this referred to was specific for the clinical trial. There is no password in the app for patients. Based on this, there was only one failure, caused by a patient forgetting the phone in another room before going to sleep. And even that, would cause almost no consequences in real world use since this would be detected by the phone/app and the patient would be instructed the following morning to repeat the test a subsequent night (and this incurs no cost). Hence, 11 out of the 12 "failures" considered by the EAG were not failures caused by the system but by the clinical trial protocol. The remaining one, one might argue could be considered because patients could forget their phones in other rooms. But even if this is considered the failure rate in Table 21, for the purpose of this economic study it should have been 1/182= 0.5% or zero. **This**	The failure rates used in the model are now reported in Table 20. Please see our response to Comment 128 regarding changes to our calculation of test failures reported by Devani et al. for the AcuPebble device. We note that the INMB estimates for the novel devices are not sensitive to changes in failure rates between lower and upper 95% confidence limits (see Tornado diagrams in Figures 12 to 17, noting that failure rates are not shown in most diagrams as they are not one of the top 20 most impactful parameters).

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				needs to be changed and the health economics model re-run with the correct numbers.	
Acurable Limited	164	121	5.7.6	An unsubstantiated hypothesis regarding the time it will take to receive a diagnosis is made, on the basis that the evidence could not be found. More specifically, the authors have decided to assume it will take three months. This is not just arbitrary, but once again, it completely biases the outputs of their model towards something that doesn't need to be true, because it is considering a similar time as they assume with RP, when in reality in many cases clinicians have the results of the tests from these novel technologies immediately after the patients do them. Therefore, the model should have considered the "real case scenario" enabled by these tests to show the potential benefits if pathways were changed. Hence, their arguments and input to the models in terms of this are flawed too. Considering all systems the same is also inherently wrong since there are some systems that do not enable immediate access to results after the night finishes. **This needs to change, since it is completely biasing the model on the basis of a wrong assumption and inputs.**	Please see EAG response to comment 52 on the effect of novel devices on times to diagnosis or treatment. Regarding the point on considering all systems the same, please see our caution against making comparisons between novel devices on the basis of the model results (Section 5.10.1), and in our response to comment 260.
Acurable Limited	165	123	5.7.7	**Top paragraph:** the authors are using relative distributions of CV events from a publication dating back to 2007, with sources from the 90s. They then go on to say that the incidence rates of CV events have changed over time. How does the knowledge of this justify the fact they appear not to have updated this review? There is at least a recent publication, albeit in the Italian population, that maps the increased of different type of cardiovascular events as well as others with OSA. Has this publication been reviewed? Could they provide evidence in the text of a more up to date review?	In the report we state that there is evidence that incidence rates have changed, but it is unclear if the distribution of incident CV events has changed over time. We argue that on the basis of recent BHF data, we believe the distributions of events are still likely to be correct given that CHD is approximately twice as common as stroke (which is what is assumed in the model). Although we sought more recent evidence on the distribution of CV events (as stated in the report), we did not have the capacity to undertake a more thorough review of the evidence. This is a limitation of the model noted in the report. It is not clear which publication the company is referring to in this comment. As noted in our response to comment 167 below, we identified a paper by Carratù et al 2021 on relationships between OSA and cardiovascular risk in an Italian population. If this is the paper referred to in the current comment, we note that it does not report on the incidence or distribution of CV events.
Acurable Limited	166	123	5.7.8	**Road traffic accident (RTA) risk:** This section is too vague to be of any use for the reader. It should be extended to understand what the increased risks associated with OSA are.	We have edited the text in Section 5.7.8 to clarify the assumptions made. Note that the RTA risks for people with OSAHS who are not treated, or are inappropriately treated, are reported in Section 5.7.11.
Acurable Limited	167	124	5.7.9	**Mortality: Second paragraph on page 124**: the EAG justifies the fact of not updating the mortality risks on the fact there was limited evidence from non UK population. Whilst for other sections of their analysis it was decided to use data corresponding to non UK populations and healthcare pathways (despite this	*Estimates of mortality* Our justification for not using the more recent values in the base case analysis was not just based on studies including non-UK participants, but also after consideration of the age of the population, and potential lack of face validity in the estimates used.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				being unjustifiable in those contexts) for this one it was not, even when in this case there are other populations that could have been considered. For example, there is a recent study carried out in Italy that has quantified risk factors associated with OSA, including mortality. The Italian population is not so dissimilar to the UK one. This study should have been cited/taken into account, considering the sources they have used for this are very old. This paragraph seems to refer to mortality exclusively after cardiovascular events, without OSA, but could be expanded. Other: Why haven't the EAG taken into account the consequences of fatigue and the effect this has on quality of life? Studies have been done around this. Notably, Harvard Medical school and McKinsey's report is one that comes to mind "The price of fatigue" https://sleep.hms.harvard.edu/sites/default/files/assets/Images/The_Price_of_Fatigue.pdf?utm_medium=PANTHEON_STRIPPED. Trying to model the effects on quality of life on the indirect measures of life of those that suffered from cardiovascular events, whilst having some merit, is not taking into account even more obvious effects on the disease, for which there are references that the EAG has not taken into account. **Could the EAG try to include other effects? (maybe using the Harvard work as a starting point for "ideas" on what to look at?). And if those are not included could they at least descriptively (but not with a short paragraph) talk about all other benefits these technologies could bring that go beyond what they have analysed?**	We use values from these alternative sources in sensitivity analyses (please see Section 5.7.9). We believe the company is referring to this paper by Carratù et al 2021 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9285080/), where the risk characteristics of people in Italy with OSA are modelled to predict 10-year risk of CV. This is essentially the approach the EAG model uses - characteristics of the OSA population are used to estimate the 10-year risk of CV events based on a prediction model. As the estimates reported in this paper are also derived from a prediction model rather than being observed, they are not helpful for this part of the model. The company is indeed correct that we are seeking mortality estimates after the CV events. Moreover, having such estimates in a OSA-only population would be preferable, however, we could not find such data. Indeed, as noted above and in section 5.7.9, we had difficulties identifying such data in the general population. *Consequences of OSAHS* With respect to consequences of OSA on quality of life - fatigue, depression and other consequences of OSAHS (excluding CV risk and RTAs) are assumed to be captured in the model via the utility estimates. In particular, it is assumed in the model that people with OSAHS have lower utilities than people who do not have OSAHS. It is further assumed that people with OSA who are treated have higher utilities than those with OSAHS who are not treated (please see Section 5.7.10).
Acurable Limited	168	129	5.7.11	**Treatments offered for OSASH:** Once again, the authors state "based on expert opinion". This is not scientific and no piece of scientific evidence leading to a recommendation should rely on this. More information should be given, at least about which experts were asked, where their numbers came from, how they calculated them, how they were asked, etc. Otherwise any quantification is potentially wrong.	The experts who advised the EAG are listed in the Acknowledgments (EAR page 2). It is common practice, where data are lacking, to consult experts on their views regarding current practice. We did not undertake any formal elicitation of expert opinion. We have undertaken scenario analyses where there was variation in the response from experts (please see Section 5.10.2 for the scenario analysis where it is assumed that 75% of people diagnosed with mild OSAHS receive CPAP).
Acurable Limited	169	134	-	**Note that from here on, all the pages are numbered 256.**	This has been amended in the Erratum report.
Acurable Limited	170	256	5.7.12	First paragraph, last sentence**: The assumption that a failure of the equipment will incur additional cost per test is not correct for AcuPebble SA100**. Any test caused by the fault of the system is provided for free. So, **the corresponding input for this needs to be changed in the health economics model.**	Please see EAG response to comments 172 and 174.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	171	256	5.7.12	Second paragraph: **It is inaccurate to say that the price of the test for AcuPebble depends on the volume per week**. It depends on the volume per year, which is very different because there could be weeks of very small volume and weeks of very large, so the company facilitates this by having a long-term averaging for pricing. **Please check how this has been considered in the model, and also clarify in the text.**	We have updated the report to clarify that AcuPebble device costs depend on the volume per year (see section 5.7.12 paragraph 2, and the last two paragraphs of section 5.7.12). As shown in Table 29, we assume that 60 tests per week is equivalent to 3000 tests per year.
Acurable Limited	172	256	5.7.12	**Third paragraph:** The assumption that a failure of a test means that additional cost of NHS administration is incurred is simplistic to the point of being wrong for some of the systems. The advantage of some of the systems (such as AcuPebble SA100) is that a problem with the system or the test, as well as the reasons for this problem are instantaneously (post-night) **detected in the backend or in the mobile phone itself.** Hence, the patient does not need to return the system or come to the hospital to get a new one. Equally, unless it is the sensor hardware itself that fails (Negligible since it is very rare. Manufacturers can provide data) any other issue can be dealt with remotely. Hence there is neither a noticeable delay (patient can repeat a test the following night) nor an additional cost for the NHS (no administrative, no postage). And the manufacturer does not charge for tests that need to be repeated because of a problem with the system. **On the basis of this the input for the health economics model representing this assumption is wrong for AcuPebble and needs to be modified and outputs of the model recalculated.**	For AcuPebble and NightOwl, the costs of a repeat test due to a failure have been amended to reflect the point made in this comment. As in our response to comment 174, we assume in the base case that the AcuPebble device is enabled by the clinician such that the patient is automatically aware if a sleep study has been unsuccessful. In this situation, the only cost assumed for a repeat test with AcuPebble is £1 for additional adhesive. In scenario analyses, we assume this function with AcuPebble is disabled and so a repeat sleep study (due to a failed sleep study) incurs NHS costs for staff to identify the failure and request a repeat from the patient.
Acurable Limited	173	256 “140”	5.7.12, Table 30	**Table 30, Costs of a sleep study with AcuPebble by volume of sleep studies assumed:** This table and the paragraph **above** the section called “Cost of AcuPebble SA100” are misleading in implying that the cost of the second night study is due to a failure to obtain a diagnosis. This is not the case for AcuPebble SA100. The numbers provided for a second night have nothing to do with test failure. Test failures are provided for free. These numbers are because due to the night-to-night physiological variability existing in patients with OSA , and the simplicity of doing an AcuPebble test, some clinicians, of their own initiative, have implemented a pathway in which every patient has two-night studies, and this is reflected in the contract. Acurable offers the second night much cheaper. None of this is costs due to failure of the system. **This needs to be corrected, both in the text and in the table. And also the outputs of the health economics model need to be recalculated. And the diagnostic advantage of having multiple nights of testing (something that can’t effectively be offered by RP) also needs to be discussed (they can find plenty of references to justify this).**	We did not intend for this assumption to be made. We have amended the text in the paragraph above the section ‘Cost of AcuPebble SA100’ in 5.7.12 to: “… component and total costs for a successful one-night sleep study are shown in Table 30. In Table 31, the total costs for a second repeat sleep study due to failure to obtain a diagnosis or misdiagnosis of individuals with moderate-severe OSA as not having OSA, are presented.”
Acurable Limited	174	256	5.7.12, Table 30	**Paragraph following Table 30 in section with title “Costs of AcuPebble SA100”:** The EAG is assuming that a functionality of AcuPebble SA100’s mobile app, indicating to patients that they have to repeat a study, can be switched on and off, and if it is switched off that would require time from the clinician. **This is not just untrue but also concerning since this is “inventing” not just how the device works, but also how it is designed and engineered**. This	We make this assumption based on the company response to Question 3 in "DAP70_EAG request for information from Acurable additional questions [noACIC10102023" that there is functionality in the mobile app to detect any issues and alert the patient to repeat the sleep study. In their response, the company included

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				functionality cannot be switched off by the clinician. It is always there. Also, it is worth noting the EAG could have contacted the manufacturers should they have had any doubt but did not. **The inputs related to this must be changed in the health economics model.**	the following "Note this is an option that can be enabled/disabled depending on the clinician preference. " Our previous analysis assumed that this functionality was disabled in the base case. We have changed the base case and now assume that this functionality would be enabled by the clinician. In scenario analyses, we assume this function with AcuPebble is disabled and so a repeat sleep study (due to a failed sleep study) incurs NHS costs for staff to identify the failure and request a repeat from the patient.
Acurable Limited	175	256	5.7.12	**Last paragraph of section "Cost of AcuPebble SA100" and subsequent section "Cost of Brizzy":** The EAG state that it takes 20m to go through the report (as a worst case scenario). The company provided a range of 5-10m, **so the numbers used by the EAG are factually wrong.** It does not take longer on average than 10m per patient (and in fact it generally takes less). However, it appears as though, for AcuPebble information, for example time and costs (see previous comments), has been made up to negatively bias the outputs of the model, whereas for others optimistic times have not been questioned comparatively. For example, it is assumed that Brizzy's time for interpretation is the same as for AcuPebble, when this is not true. Brizzy has not received regulatory clearance (CE-mark) for fully automated diagnosis, whereas AcuPebble has. Brizzy's outputs require interpretation of signals that are surrogate of respiratory channels and hence can lead to longer interpretation time, and the list goes on (a more lengthy discussion can be provided). Also, they assume a similar case scenario for Brizzy as for AcuPebble and cost it the same, when in reality in the case of Brizzy the device has to be returned to the clinic for data upload (i.e. it is not instantaneous), whereas in the case of AcuPebble SA100 clinicians have access to all the information instantaneously. And in the case of AcuPebble an inconclusive/failed test can be repeated at zero cost (both test free and no human resources), **This needs to be corrected. The time and hence corresponding costs have been overestimated for AcuPebble SA100, which invariably results in a negatively biased output of the model. Once the inputs to the model are corrected, and the new outputs extracted, all the relevant figures, tables and descriptive parts of this report need to be redone.**	Although companies provided estimates on the time taken to review data and prepare a report, we have used in our base case analysis the assumption of 20 minutes. This is based on experts reporting that although novel devices can produce reports in shorter times, it takes NHS staff about 20 minutes to go through the report. We have added a scenario analysis to reflect the time that companies have stated it would take to prepare the report for their device (please see Section 5.10.2, Table 38 in the Erratum). This scenario led to a small improvement in cost-effectiveness for all novel devices compared with RP and oximetry: for example, the INMB for Acupebble versus RP rose from £141 in the base case to £150 in the scenario with the company stated time to review (at the £20,000 per QALY threshold). However, increases were similar for other novel devices, based on the stated times to review by the respective companies, and the ranking of devices by INMB did not change. We also conducted a sensitivity analysis, varying the cost for time to review data and prepare a report for all the novel devices (from 5 minutes of a band 5 to 20 minutes for a consultant) (see Table 78). Assuming 5 minutes for a band 5 member of staff improved the IMNB for each device compared to home RP by approximately £15, leading to NightOwl and WatchPAT 300 being considered cost-effective (INMB of £8 and £3 at the £20,000 per QALY WTP, respectively), and Brizzy having an INMB of £0. When compared to oximetry, assuming 5 minutes for a band 5 member of staff did not change the overall conclusions from those in the base case. Assuming 20 minutes for a consultant reduced the INMB for all devices in comparison to RP and oximetry, but did not change the overall conclusions. The tornado plots in Figures 12-17 (Appendix 9b) show the impact of assumptions on time to review data.
Acurable Limited	176	256 "142"	5.7.12	**Section called "Cost of NightOwl":** Again, the EAG makes an assumption in a particular case where the truth is provided by the manufacturers, based on real world use data. More specifically, in this section they say "For consistency with	Please see above response to comment 175.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				our assumptions across the novel devices, in the base case, we assume 20 minutes of band 6 sleep physiologist for report preparation". **In the case of AcuPebble SA100 this is not true**, the report is automatically prepared, and the manufacturers provided data for time estimations. However, the data has been changed to make the numbers worse, when in reality AcuPebble SA100 has got functionality that other systems do not have, precisely to facilitate clinician's work. Hence it is not right to change the numbers to make all the systems the same (considering they are different). Additionally, and concerningly, in this section the EAG state that the 15 minutes of nurse time to prescribe NightOwl are not taken into account, however in the section "Cost of AcuPebble SA100", what seems to be the equivalent is costed (creating the study on the web app). **Please clarify what is meant by prescribing.**	ResMed state that there is no requirement for preparation of the device by NHS staff. Thus, we do not include this cost. As described in response to comment 175, in the base case analysis we assume it takes 20 minutes to review data for all novel devices.
Acurable Limited	177	256	5.7.12	**Section called "Cost of Sunrise":** As the system is disposable, hospitals need to store a non-significant volume of capital equipment. The state costs associated with this should be accounted for in the health economics model. This is especially relevant for London hospitals. Additionally, has the EAG checked that it is actually possible for any of these companies to post the devices to the patients? In order for this to happen Information Governance in the NHS has to have given the clearance to share personal patient data. In the case of some technologies the system is by design forcing patients to share very sensitive unnecessary personal information, before they are allowed to do the test (this could have just been checked by downloading their apps). The reason for that seems to be because they are training ML algorithms in the background. As part of the permissions they seem to also be requiring that patients allow this to be shared with collaborators. This is borderline unethical and it is very likely because of this Data Protection Departments within the NHS would not allow postage to patients. Could this scenario be confirmed, for all the systems? And if there are systems for which this has not been implemented the economic modelling needs to re-do any hypothesis for those systems for which this is currently not realistic. And the data sharing requirements should be brought up in the report.	All manufacturers indicated that the novel devices could be posted, and this is assumed in the model (hence including the costs of postage). We do not include additional costs for storage of any of the novel devices. It would be very difficult to estimate storage costs, as they would vary between centres, depending on patient throughput and purchasing policies. Assessing compliance with data protection regulations is outside the scope of the EAGs work. Note that recent NICE guidance has stated the need for technologies to follow NHSE's Digital Technology Assessment Criteria (DTAC), which includes data protection.
Acurable Limited	178	256	5.7.12	**Section called "Cost of WatchPAT 300":** Unlike in AcuPebble (and maybe in others), could it be clarified whether in WatchPAT the cost of a second sleep study caused by a failure of the device is charged or has a cost? (the consumable is not a small adhesive for which patients receive more than they need. It is a full electronic probe, and there is no way of reusing the one they provide because the pneumatic pressure mechanism cannot be re-established after one use). And again the EAG makes no differentiation between the implications (neither financial or in terms of diagnosis delay) between devices in which the results are instantaneously available, and devices such as this one and Brizzy in which the only way of knowing whether something has failed is for patients to return the device to the clinic, as is the case, as stated, for WatchPAT	According to the manufacturer of WatchPAT, all consumables are included in the cost of the sleep study. This is shown in Table 30, and reported in the description of costs for WatchPAT 300 and WatchPAT ONE in section 5.7.12. It is also stated in this section that the cost of a sleep study due to failure would not be incurred for the WatchPAT devices. Although there is indeed variation in how the data from the different devices is obtained and downloaded, for all devices and comparators, the model does account for NHS staff time to check if the sleep study has failed and request a repeat sleep study. This

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				300. **This needs to be corrected throughout. The devices are all different and those differences need to be accounted for. Otherwise the outputs of their model will be misleading and misinformed.**	is reported for each device in section 5.7.12, Table 30 and Table 74 (Appendix 8).
Acurable Limited	179	256	5.7.12	**Section "Cost of WatchPAT ONE":** Once again the EAG states "Based on the experience of experts". Who are the experts? What were the exact questions asked? What credentials do the experts have? What assumptions or data are they basing these numbers on? This kind of generic statement should have no place in a piece of scientific work.	Please see response to comment 168.
Acurable Limited	180	256	5.7.12	**Section "Volume of sleep studies per week":** This section should be clarified, because it is considering number of sleep studies per week and it is assuming that there is no effect on Sunrise because it is assume 100 kits would be ordered in a batch, whereas there is an effect on AcuPebble and WatchPAT300 depending on whether 25 tests per week or 100 per week are ordered, when the reality is that in the case of AcuPebble SA100 at least, the pricing does not depend on tests per week but rather on tests average per year. Both are not necessarily equivalent. It depends on what are the contractual conditions behind the 100 tests/week of Sunrise. **Please confirm the assumptions made for the model are the correct ones.**	Please see response to comment 171 and clarifying text to this section of 5.7.12 in the Erratum report (renamed 'Volume of sleep studies).
Acurable Limited	181	256	5.7.12	**Section "Volume of sleep studies per week": The second paragraph once again makes the wrong assumption**: The cost of a repeat test for AcuPebble is zero provided that it is due to a fault of the system.	As described in Section 5.6,1, there are two reasons why a repeat study would be required. Where a repeat is required due to misdiagnosis there would be costs associated with that repeat. Therefore, we have kept the text as it is.
Acurable Limited	182	256	5.7.12, Table 31	**Table 31: This table is incorrect.** AcuPebble SA100 has amongst the fewest preparation steps than any of the other devices, the data is analysed instantaneously, there is no calibration, the output is totally automated (and the regulatory authorities have granted authorization for it, unlike for others), the patients attach the adhesives themselves, even if its reused and cleaned up, its size and shape makes this process faster or at least as fast as for others; and still the EAG has come up with a number for preparation of the device cost that is up to 3 times higher than for others. **This does not make sense**. **This needs to be corrected and if there is any doubt of how the device works the manufacturers can be contacted.**	As reported in the EAR, the company provided an estimation of 10-15 minutes of NHS staff time needed for preparation of the device, including cleaning and putting details on the web or mobile application for AcuPebble. We assumed 15 minutes for a band 4 member of NHS staff for such preparation. We have updated the model to assume this is 10 minutes. As an illustration of the impact of preparation time on the cost-effectiveness results, if we assume no preparation is required the INMB for AcuPebble compared to home RP increases by £6 to £147 at a WTP threshold of £20,000 per QALY gained.
Acurable Limited	183		5.7.13	"Costs of home oximetry and respiratory polygraphy": **Having oximetry as a comparator is implicitly promoting health inequalities**. See the abundance of comments in relation to this in other parts of this review.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	184		Table 34	**Cardiovascular event and road traffic accident costs:** Is this per year or over the lifetime of the patient? Please clarify in the heading. Also, why is the cost for traffic accidents including the police costs but not including the overall costs to the economy, which are also given by the Department of transport and are significantly higher than that? How many of these accidents lead to disabilities, and are the impacts of those disabilities incorporated in the model in any way? Other costs should be properly scientifically discussed at least.	CV events are costs per year (in that state); RTAs are cost per event. This has been updated in the report for Table 34 (now Table 33). The cost perspective for NICE is the NHS and Personal Social Services. Therefore, in base case analyses we only consider the NHS costs associated with RTAs. For one scenario analysis, we include police costs with NHS costs, but this is outside the scope and perspective for this analysis.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					The health impacts of an RTA are accounted for in the model via the utilities as detailed in Table 24 (now Table 23) in section 5.7.10.
Acurable Limited	185	153	5.8	**"Model assumptions":** The three model assumptions **have been wrongly quantified for AcuPebble SA100**, not by a small amount, and in some cases completely invented via incorrect 'adjustments', with absolutely no reason since the EAG had been provided with evidence which hasn't been referred to. **Hence, the outputs need to be recalculated after having corrected those assumptions.** Furthermore, the authors say that due to the lack of evidence, the differential time to diagnosis for each novel device and comparator is assumed to be the same. This is wrong scientifically because, regardless of whether there is evidence or not, the functionality of the novel technologies is different to the ones of the comparators, and the former, in many cases give **instantaneous access to the diagnostic results, which is something that the comparators cannot give**. Hence, potentially the time to diagnosis could be massively reduced thanks to the novel technologies, whereas it can't with the comparators. And even if they didn't this would be a problem associated with the current diagnostic pathways that were created around the characteristics of the technologies that were available at the time when those were defined, or to other factors such as the current waiting lists being too large to handle. The model should differentiate between what can be achieved with one and what can be achieved with another since otherwise clearly, if one assumes that their functionality and what they enable is not any different, the output of the model will result on something that shows that there are not advantages, when in reality this output is just the result of the assumption that, whilst being technologically different to the comparator, they don't improve in terms of the output. **This is a flawed assumption. This needs to be corrected taking into account the individual characteristics of the devices.**	We have re-run the analyses using accuracy provided by Acurable in the base case analysis. With respect to the point on time to diagnosis, please see our response to comment 52.
Acurable Limited	186		Table 35	**Population:** this assumption **promotes health inequalities** since it is implicitly assuming all other data applies equally to: 1- Those who have dark skin (for example some ethnic minorities) as to those with white skin, when that is not true (See comments about systems that rely almost uniquely on PPG signal). 2- Those from ethnic minorities that cannot shave a beard (See for example "Ramandeep Sahota (2021) COVID-19, beards and BAME: how ethnic minorities with religious beards are being let down, Journal of Occupational and Environmental Hygiene, 18:10-11, 477-480") 3- Women. The assumption of 70% being male is simplistic/obsolete and the conclusions of this report are promoting health inequalities). See for example Martins FO, Conde SV. Gender Differences in the Context of Obstructive Sleep Apnea and Metabolic Diseases. Front Physiol. 2021 Dec 14;12:792633. doi:	Note that Table 35 is now Table 34. 1- Please see comment 52 (Acurable) for the EAG overall response on the potential impact on oximetry and PPG on health inequalities. 2- Thank you for the Sahota 2021 reference. We note that some devices are not suitable for use for people with beards. We agree that this is a potential inequalities issue, both directly and indirectly due to the prevalence of beards for some ethnic minorities. 3- Thank you also for the Martins 2021 reference. This suggests OSA has been under-diagnosed in women: so the proportion of people with OSA who are women is likely to be higher in the community than in the clinics. However, for the purpose of estimating the cost-effectiveness of OSA tests from an NHS

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				10.3389/fphys.2021.792633. PMID: 34970158; PMCID: PMC8712658 or Jehan S, Auguste E, Zizi F, Pandi-Perumal SR, Gupta R, Attarian H, Jean-Louis G, McFarlane SI. Obstructive Sleep Apnea: Women's Perspective. J Sleep Med Disord. 2016;3(6):1064. Epub 2016 Aug 25. PMID: 28239685; PMCID: PMC5323064. **The outputs of the model and the whole report needs to be rewritten to make clear these findings refer only to those with white skin and in the cases of those devices which require beard shaving they might not be relevant to those ethnic minorities for whom beards are a religious requirement. It also needs to say that the outcomes might be biased against women, since it has been assumed that 70% of the population are male. This especially applies to women who are pregnant and women in peri-menopausal, menopausal ages.**	perspective, the relevant population is people referred for tests for suspected OSA, thus the clinic population is more relevant. There is a question of whether this average cost-effectiveness is appropriate for people with OSA at particularly high or low risk of related adverse outcomes, which will differ between men and women. However, our scenario analysis showed that cost-effectiveness results were similar for people at low vs. high cardiovascular risk. See Table 17 and section 5.7.7 for an explanation of this analysis, and Table 77 for the results for the lower and higher risk cohorts.
Acurable Limited	187		Table 35	**Diagnostic Accuracy:** The table stresses something that was already covered in previous parts of the report and is that the **EAG "made up" numbers for AcuPebble.** The reasons why this is completely unjustified, and the adjustment scientifically wrong are many and have been covered above. But most importantly, these numbers are wrong because Acurable provided NICE with PSG validation data within the deadline imposed to provide information and the EAG decided not to take that data but rather state in the report that they didn't have it or couldn't use it, using that for justification to entirely arbitrarily reduce the performance of AcuPebble SA100. **This is factually incorrect, and on the basis of this, the outputs of the model need to be recalculated for AcuPebble SA100 with this assumption corrected, taking the actual data into account instead.** The authors are oversimplifying the concept of "Diagnostic accuracy" of RP (This is explained in other parts of this document). **It needs to be specified how the accuracy was determined.** Regarding the section about WatchPAT, taking accuracy data from other products of a company assuming that it would be the same for the product under evaluation is not properly justified, and it might be completely wrong (depending on whether the right justification is found) . There are regulatory processes of new products for a reason. **This needs to be fully justified.**	We have re-run the analyses using accuracy provided by Acurable in the base case analysis. We have added further detail on the studies by Xu and Pereira which we use to inform the accuracy of home RP – please see Section 5.7.4. Please see comment 24 from the manufacturers of WatchPAT 300 and WatchPAT ONE, where they confirm that these features of these devices are identical.
Acurable Limited	188		Table 35	**Diagnostic pathway:** The EAG has assumed arbitrarily that it will take the same amount of time (i.e. three months) for a patient to be diagnosed with the novel technologies as it takes with the comparators. This assumption is unfounded. Many of the novel technologies allow for diagnosis much faster because the data is immediately accessible to clinicians. In the case of AcuPebble SA100 the technology has been granted regulatory approval for full automated diagnosis,	With respect to the point on time to diagnosis, please see response to comment 52. With respect to the comment on a repeat sleep study taking an additional month, this is incorrect. We assume that a repeat sleep

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				so in the optimum case scenario this would be instantaneous. **It makes no sense to review novel technologies if pathways taking advantage of new functionalities of the technologies are not considered.** And in this case the fastest diagnosis is enabled. **This needs to be corrected.** Furthermore, the EAG assumes that in case a test needs to be repeated due to failure of the first it would take an additional month. Again, **this is completely wrong** for many of the novel devices. Some of them are disposable and this would cause a delay. But some of them are reusable, an invalid test would be automatically noticed, and patients would be asked to repeat the study the following night. So the delay would be one day. **For these reasons these assumptions are wrong and the inputs of the model have to be changed, and consequently as the outputs change also the descriptive parts of this report.** Also, as a note of this: making the same assumptions for all the technologies is wrong since they are not only functionally different but they have also been granted different regulatory approvals in terms of intended use and hence it is potentially unsafe to consider them all under "the same bucket".	study due to misdiagnosis would occur a month later. This is intended to reflect the situation where a patient with moderate-to-severe OSA is misdiagnosed as having no OSA, yet their symptoms continue, and a month later they are asked to repeat the sleep study. This assumption is independent of the device used for the sleep study. Please see Section 5.6.1 (first paragraph below Figure 3), where we explain that there are two reasons for a second sleep study (due to a failure to provide a diagnosis, and due to misdiagnosis), and section 5.7.6 (second paragraph), where we state that we assume a delay of one month in diagnosis for people with moderate-severe OSA initially misdiagnosed as having no OSA. We have added clarification in section 5.6.1
Acurable Limited	189		Table 35	**Natural history:** Some of the assumptions here might be over simplistic. For example the financial implications of road accidents are underestimated. Refer to the corresponding section above.	Please see response to comment 184 on cost perspective for NICE.
Acurable Limited	190		Table 35	**Utilities:** This seems to be over simplistic. Other health economic studies by reputable institutions over the world could have been checked and taken as a starting point. For example, has the Harvard School of Medicine and McKensey work been considered?	Regarding consideration of the impacts of OSAHS and sub-optimally treated OSAHS, please see our response to comment 167. Please also see section 5.2 where we describe our systematic review of health-related quality of life studies, and section 5.7.10 where we explain the rationale for utility values used in the model.
Acurable Limited	191		Table 35	**Costs:** Oximetry in isolation should not be considered as an alternative for diagnosis after the evidence resulting from COVID, and considering the fact regulatory bodies around the world are all trying to see how to modify their regulatory processes to prevent what has/is happening, after it was discovered they do not work in those with dark skin. This is an insurmountable problem in those systems currently in the market due to the limitations of the sensing and electronic modalities. Additionally, it should be noted from a regulatory perspective that the ISO standard for validation still only requires 67% of data to be within 4% in 10 pooled subjects (out of which only two have to have dark skin and since the data is pooled the limitations of sensing and electronics allow the oximeter to pass regulations). The EAG should be up to date with the scientific evidence surrounding these systems and presenting them as a comparator is unacceptable from the point of view of health inequalities. **Please either remove oximetry as a comparator or make a statement in relation to dangers for those with dark skin in every paragraph this is mentioned, or clearly state this report applies only to those with fair skin.**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				The investigators (EAG) assume NHS staff time when a sleep study fails; to identify the study as a failure and speak to the patient to request a second study is done. **This assumption does not apply to AcuPebble SA100 since the patient is informed automatically.** Also the second study for AcuPebble SA100 in case of failure is free, whereas it seems this has been cost in the model. **On the basis of this the inputs corresponding to AcuPebble SA100 for the model need to change, the outputs recalculated and the corresponding sections in the text changed.**	With respect to the point on failure costs for AcuPebble, and the functionality of AcuPebble to automatically alert the patient, please see response to comment 172.
Acurable Limited	192		5.10	**Economic Analysis results:** The investigators are again comparing with oximetry when this should not be an option. This section is missing references. **Where are those formulas extracted from? Have they been defined by the EAG or are they commonly used and accepted by the academic community? Could they please give the formulas for every single metric they are using, rather than just describing it with words in some cases in the text? Also, could they give formulas in the main body of the text for the different fields that they appear in the subsequent tables?** Whilst this might be in an appendix (not checked) or in a piece of code submitted, it's poor scientific reporting practice not to give at least the basic information to be able to directly follow tables and/or figures. Without the definition of the formulas the tables that come after are not easy to audit. Also, when referring to advantages and limitations, are those widely known, or is it their interpretation? **This needs to be clarified and if they are widely known references given.**	The equations for incremental cost-effectiveness ratio and net monetary benefit are commonly used and accepted in economic evaluation. References for the equations have been added to section 5.10.
Acurable Limited	193		5.10.1	**Deterministic results:** This section says "The diagnostic pathway using oximetry is estimated to be the cheapest", without even discussing what the safety concerns associated with oximetry are. This is a concern because clearly for anyone who is looking to either increase profits or decrease costs this could lead to a direct decision to use oximetry, **potentially discriminating patients and promoting health inequalities.** The EAG also comes to the conclusion that WatchPAT 300 and WatchPAT ONE are associated with the highest estimated QALYs, driven by the estimated sensitivity of the WatchPAT devices. This is not just misleading but potentially untrue, as it results from the investigators having ignored the landmark piece of evidence on WatchPAT (applicable to all PAT devices in which the sensitivity depends on the PAT channel), which, interestingly, led to the American Association of Sleep Medicine issuing a notification saying that PAT should not be used for diagnosis. This was a clinical trial involving 500 patients (*Performance of peripheral arterial tonometry–based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort," Octavian C. Ioachimescu,*	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities. With regard to Ioachimescu 2020 please see our response to comment 26 and comment 253. With respect to WatchPAT 300 and WatchPAT ONE having the highest estimated QALYs, please see our response to comment 42.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				MD, PhD, *J. Shirine Allam, MD*, *Arash Samarghandi, MD*, *Neesha Anand, MD*, *Barry G. Fields, MD, MSEd*, *Swapan A. Dholakia, MD*, *Saiprakash B. Venkateshiah, MD*, *Rina Eisenstein, MD*, *Mary-Margaret Ciavatta, PAC*, *Nancy A. Collop, MD*). This trial is also relevant because 72% of the subjects were African American (which is the likely reason for it massive drop in performance) which reiterates the fact that any conclusion of any device based on PAT that has been obtained with a white population cannot be extrapolated to a dark skin population because the light transmission and absorption PPG is based on is very different. **It is worth noting that the EAG does say that results should be interpreted with caution, but this is neither justifiable nor sufficient given the problems highlighted in previous comments on bias evidence, made up data, lack of justification for using data, and potential promotion of health inequalities. We believe it is possible to do a much more thorough scientific work that removes a lot of the uncertainty and it has not been done.** In the penultimate paragraph it seems the EAG is assuming that if a patient receives an invalid diagnosis with oximetry and continues with symptoms the patient will have a RP test done within a year but the same assumption doesn't seem to be made with WatchPAT (is it assumed the test will be with WatchPAT again?), when the underlying reasons why a diagnosis is wrong in patients are generally the same: limitations of the signal from where SPO2 and/or PAT are extracted, ie the PPG signal. Those limitations are physical and relate to the sensing approach. **This needs to be taken into account.**	Regarding uncertainty in the model results, please see response to comment 69. Since the decision question is whether home RP can be replaced by the novel devices, in the WatchPAT strategy there would be no home RP, and so it is assumed that a repeat test for a misdiagnosis would be done with WatchPAT.
Acurable Limited	194		Table 36	Diagnosis row: **the numbers calculated in this row for AcuPebble SA100 are wrong because the assumptions were wrong**. This has been explained before. This needs to be recalculated. Intermediate outcomes: **All the numbers for AcuPebble SA100 in this subsection are wrong (factually wrong, not just interpretatively wrong. The EAG has used "made up" numbers, due to incorrect methodology and equations, as discussed in previous comments). These numbers need to be changed.** Long term outcomes: This is impossible to follow because the EAG does not give a proper definition of formulas in the body of the text. However they must be **wrong for AcuPebble SA100 considering all the inputs required by their model were wrong. This needs to be corrected.**	All base case analyses have been updated using accuracy data for AcuPebble that was supplied by the company.
Acurable Limited	195		Figure 5	**This figure is wrong for AcuPebble SA100 because it has been obtained with numbers that are wrong. This needs to be corrected.**	Please see response to comment 194.
Acurable Limited	196		Table 37	**All the calculations here for AcuPebble SA100 are wrong because they have been obtained with inputs to the model that are wrong. Hence they need to be recalculated.**	Please see response to comment 194.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	197		Figure 6	**This figure is wrong for AcuPebble SA100 because it has been obtained with inputs to the model that are wrong. This needs to be corrected.**	Please see response to comment 194.
Acurable Limited	198	163	5.10.1	**For reference, as the page numbers are all 256, this is regarding the paragraph after Figure 6**: the conclusions for Sunrise are potentially misleading. Sunrise performance data (as acknowledged in passing by the EAG but not taken into account in the economic model) was obtained with post-hoc diagnostic thresholds after seeing the ROC curve, and not corresponding to conventional clinical diagnostic ones. Any system that had done that to obtain performance would have come up with better numbers. However, this is not representative of reality. The model does not take this into account. **This needs to be clarified here too.**	The company is correct that the model does not account for this. In this section we report the results, and the limitations of the results - including limitations in the accuracy evidence - are stated in Section 6.2.2.
Acurable Limited	199	165	Table 38	**These numbers are wrong because the assumptions are wrong.** See previous comments. Also oximetry should not even be presented or considered an option (see previous comments).	Please see response to comment 195 regarding model assumptions, and comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	200	167	Figure 7	**This needs to be redone for AcuPebble SA100 because the inputs to the model are wrong.**	Please see response to comment 194.
Acurable Limited	201	168	Figure 8	**This needs to be redone for AcuPebble SA100 because the inputs to the model are wrong.**	Please see response to comment 194.
Acurable Limited	202		Table 39	This table is underdefined and hence can be meaningless for the reader. What is meant by "**correlation between first and second sleep study results**"? Is this a consideration of night-to-night variability? Or is it a reference to whether the initial and second result are likely to have a correlation in terms of misclassification of disease?	Note that Table 39 is now Table 38. It is the second definition: that the initial and second result are likely to have a correlation in terms of misclassification of disease. This has been clarified in Table 38 in the report.
Acurable Limited	203		Table 39	**In the sensitivity and specificity row:** This is unclear. Where and how has the lower limit of the confidence interval been used? This was mentioned this before in the report as well, but then in the tables there are three numbers (Two of them in parenthesis)	As Table 38 indicates, we use the lower and upper 95%CI for the sensitivity and specificity estimates of the novel devices in scenario analyses. Please also see comment 157.
Acurable Limited	204		Table 39	**Alternative decision tree parameterisation:** Could they explain in the table itself what they meant by "using raw accuracy data from NightOwl"?	Table 38 has been edited to clarify this.
Acurable Limited	205		Table 39	**Accuracy data for AcuPebble: The adjustment for PSG is unjustifiable, not just because the methodology is scientifically wrong, but also because they were provided with PSG data.** Also, what is meant by "unadjusted accuracy"? Is this the data from PSG comparison or from Devani et al?	Unadjusted accuracy relates to the data as reported in Devani (with no adjustment).
Acurable Limited	206		Table 39	**Third sleep study for false negatives:** PSG is being assumed for those with a false negative. That is not realistic. Many hospitals don't even have the ability to offer PSG, so in the best-case scenario patients would be offered RP in most places. **This needs to be reconsidered.**	In our discussion with experts, we understood that capacity was limited, but agreed that they could probably accommodate this, if needed. We therefore keep it as a scenario analysis in the model.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	207		Table 39	**Transport of novel devices to/from patients' home:** Has the EAG checked that this is a viable option for all devices? (ie. they all have clearance proven from Information Governance Departments within the NHS? Note that the amount of personal data collected varies and in some cases what is collected would prevent Information governance departments to allow for this). **This needs to be checked before making this assumption.**	Please see EAG response to comment 177.
Acurable Limited	208		Table 39	**Volume of sleep studies conducted per week:** Depending on how this has been inputted into the model it might be wrong for AcuPebble since it is not the number of studies per week that determines the price, but the average per year. Please confirm.	Please see response to comment 171.
Acurable Limited	209		Table 39	**Time for data review of novel devices:** This scenario is incorrect because it has assumed all devices have the same regulatory clearance and hence can offer the same performance, which is not correct. Review with AcuPebble can be done much faster than with others due to its regulatory approval for automated diagnosis. This needs to be corrected. The numbers cannot be the same for all devices because they are very different and they have different features, capabilities and intended uses. **And the numbers for AcuPebble cannot be "invented" just for the sake of making it worse so that they appear similar to others (the report repetitively says " for fairness". This must be corrected.**	Please see response to comment 175
Acurable Limited	210			**Cost of RTAs:** This is very much underestimated. See comments related to this in other sections.	Please see response to comment 184
Acurable Limited	211	171	5.10.2	**Scenario analysis for diagnostic accuracy estimates section:** **The conclusion of the novel devices having fewer QALY's have been obtained with the wrong input data for AcuPebble SA100 (unknown about the others). The analysis has to be repeated with the right data and the text changed accordingly throughout.** The conclusion that the "unadjusted analysis for AcuPebble is unrealistic because it assumes respiratory polygraphy has the same accuracy as PSG" is **simplistic and in this context completely wrong**. AcuPebble works in a completely different way to RP systems. The reasons for the differences reported between PSG and RP do not extrapolate to AcuPebble. Furthermore, the EAG has misunderstood and wrongly applied the formula used for adjustment (and on this note that formula is in a paper that hasn't been cited once, published in a "second class" (Q2) journal, hence it is not a widely used scientific methodology). Hence **the adjustment is also wrong**. Regardless this is irrelevant because the EAG, through NICE, had been provided with real world evidence (from a registered clinical trial that also demonstrates that their adjustment is wrong and arbitrary, and the performance of AcuPebble is maintained when compared to PSG). This has neither been acknowledged nor taken into account for their analysis. **All calculations that depend on this need to be redone because this is factually incorrect.**	Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024)) All base case analyses have been updated using accuracy data for AcuPebble from this trial. With regard to the adjustment to compare AcuPebble to PSG please see our response to comment 138 (Acurable).

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	212	172	Table 40	**For all scenario analysis tables:** The calculation of QALYs and all the subsequent numbers that depend on it are wrong for AcuPebble SA100 because the input data was wrong. Refer to previous comments that detail every single wrong assumption/input that has been made for AcuPebble SA100. **This needs to be changed.** Also, the data on this table does not consider: 1- The numbers of NightOwl have been obtained taking into account a device that cannot be sold in the NHS/UK. **Hence they need to be recalculated eliminating any assumption of input for this device.** 2- The EAG has missed the landmark publication of the results of WatchPAT, a clinical trial with 500 people showing much lower levels of sensitivity/specificity/accuracy, which would significantly change the results on the table (refer to previous comments). **In the context of this health economics study this trial is relevant and applicable since the changes in other versions of WatchPat do not affect the outputs that are relevant to this economic study, and in fact the study was done with a version that if nothing else would only improve diagnosis since it relied on two independent PPG extracted channels (so the information could be redundant if one changed)** 3- The tables assume a delay in diagnosis that is not representative of what all devices offer. Hence whilst some devices do not produce an instantaneous automated diagnosis others do. Still the EAG assumes the delay in diagnosis is three months. **This needs to be corrected.** 4- The fact that the sensitivity and specificity of Sunrise has been obtained with post-hoc thresholds which are different from the clinically accepted/established ones is not taken into account or mentioned. **This needs to be corrected.** 5- The numbers of the table have underestimated/simplified the prevalence of OSA in women since it has a different age distribution than in men. **This needs to be corrected.** 6- The tables fail to identify that in the case of devices that rely solely on PPG and no flow these numbers only represent the white population. **This needs to be explained.** 7- The comparison with oximetry should not be made (see all comments on health inequalities) and if it is made it needs to be clarified on the table that oximetry might not work for those with dark skin. The same for PAT since this technology is based on the same original signal which is the fundamental issue with people with dark skin.	All base case analyses have been updated using accuracy data for AcuPebble that was supplied by the company. In response to the other point made in this comment: 1.With regard to CE marking and regulatory approval please see our response to comment 100 (Acurable). 2, With regard to Ioachimescu 2020 please see our response to comment 26 and comment 253. 3, please see our response to comment 52. 4 – please see our response to comment 198 5-7 Please see comment 52 regarding the potential impact on oximetry and PPG on health inequalities.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Refer to previous comments for further explanation of what must be changed.	
Acurable Limited	213		Table 41	Same general comments as for table 40.	Please see response to comment 212
Acurable Limited	214		Table 42	Same general comments as for table 40.	Please see response to comment 212
Acurable Limited	215		Table 43	Same general comments as for table 40.	Please see response to comment 212
Acurable Limited	216		Table 44	See comments above in the document. **The entire column has to be re-calculated for AcuPebble SA100.** Even when in this table the authors claim the performance hasn't been adjusted there **are also other assumptions that are wrong/untrue** (failure rate, the fact that a functionality of the app that advises the patient to repeat the test without involving of the healthcare professional can be switched off, the fact that it takes the same time from the clinician than for other systems that have not been granted the CE-mark for automated diagnosis and hence signals not only have to be reviewed but also scored, etc.)	Please see response to comment 128, comment 174 and comment 175.
Acurable Limited	217	177	5.10.2	Section "Other scenario analyses that affect results": **This will need to be rewritten depending on the results when the tables are re-calculated.**	All analyses have been re-run.
Acurable Limited	218		Table 45	**This table needs to be modified and some numbers (the AcuPebble column but maybe also others) need to be recalculated with the right assumptions.**	All analyses have been re-run.
Acurable Limited	219		Table 46	Same general comments as for other tables.	All analyses have been re-run.
Acurable Limited	220		5.10.2	Additional comment: There should be separate tables considering the special cases of vulnerable rural populations since none of these tables considers this.	We note that the NICE scope highlighted potential health equality issues for people whose access to some technologies may be restricted (including rural or deprived areas with limited wi-fi access). However, we have not identified clinical or diagnostic data specific to such populations, which means that we cannot produce separate cost-effectiveness results. See comment 52 for general comments on how we expect NICE and the committee to account for potential health inequalities as part of the deliberative process.
Acurable Limited	221	180	5.10.3	**If a comparison with oximetry is going to be made it needs to be made clear this only applies to those with white skin.** Refers to previous comments in the document.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	222	181	5.11.1	**Challenges in modelling for children: Second paragraph**: This section sets up the context by saying that for children with no-comorbidities questionnaires are recommended by the BTS, under a guideline that was published in 2023. However, it should be considered that maybe this guideline had been created taking into account the constraints associated with diagnosis with existing technologies, and the fact that even for the emerging ones there hadn't been evidence validating their performance in children. In the same way, the recommendation of pulse oximetry is likely the result of not having anything better. Note that due to the invasiveness of the validation test required by	We recognise that as in adults, there is a potential equality issue regarding the use of oximetry for children from black, Asian and minority ethnic backgrounds. However, this section of our report is not making a case for (or against), the use of oximetry. It provides a factual description of the pathway for diagnosing OSA in children proposed in the recent BTS guideline, with the objective of defining a baseline of current practice, against which the introduction of home testing with novel devices could be evaluated in a new economic model. We conclude that "The simplest pathway would

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				regulators to approve oximeters, the ISO 806061 standard allows for oximeters to be used (pass regulations) in children based on the results for adults. Hence, there is even more uncertainty on the outputs of oximeters in children than in adults. Furthermore, again this is potentially discriminatory because of the limitations of the required regulatory testing, together with that of the sensing modality used to obtain oxygen saturation non-invasively, since it has been proven that oximeters do not work as they should in those with dark skin, this will happen also in children (note that the same will happen to any sensing modality that relies on PPG). It is also worth mentioning that the standard for regulation of oximeters only requires 67% of data points in validation to be within 4% of the gold standard. Hence, using oximetry only, for children, considering: no oximeter has been validated in children; they are proven not to work on those with dark skin, there is potential variability in excess of 4% in the desaturations, and children move more throughout the night than adults (and most oximeters in the market have not done regulatory testing with artefacts, since the standard allows for the subject to be in still conditions, should only be the last resort. It is therefore very possible that this is recommended by BTS due to a lack of better alternatives. **Since novel technologies are the subject of this NICE report, this section should discuss all of this.**	be in identification of OSAHS in children without comorbidities, where the use of novel devices as an alternative to pulse oximetry at home could be explored."
Acurable Limited	223		5.11.1	**Challenges in modelling for children:** Penultimate paragraph in this section: The EAG refer to "paediatric experts we have talked to". This is not scientific. Please could it be clarified: Who the experts are? How many were asked? How many centres in the UK they covered? What are the characteristics of these centres? What were the questions asked? The authors state "The simplest pathway would be in identification of OSAHS in children without comorbidities, where the use of novel devices as an alternative to pulse oximetry at home could be explored. However, such children appear to be a minority". This statement is both unscientific and wrong. It does not appear that a literature review has been conducted. Are the authors aware, for example, that it is estimated that 25% of children that are diagnosed with ADHD don't have ADHD? (there are publications on this. They should have been mentioned/discussed). Have the adverse effects and the financial costs associated with the incorrect use of anaesthetics in children with OSA been investigated? There is a lack of references or proper discussions on this. Have the reports that claim that approximately 20% of children in the bottom 10% of school have OSA, and the implications this will have over their lifetime, been investigated? **Please add references and discussions for all of this.**	The experts who advised the EAG are listed in the Acknowledgments (page 2). They include two Consultants in Paediatric Respiratory Medicine. We also spoke with a Consultant in Paediatric Respiratory and Sleep Medicine who is a Specialist Committee Member of the Diagnostic Advisory committee for this assessment. The statement that children without a co-morbidity are a minority of children being tested for OSA was based on estimates from two of the above experts. However, as the experts cited a wide range of estimates, and we have not been able to find a good published source, we have deleted this point from the text. Section 5.11.1 provides an overview of the challenges that we see in developing a health economic model to evaluate the use of novel home-based diagnostic devices in children. A pre-requisite for development of such a model is an understanding of how diagnostic tests are currently used and how they impact on the care pathway, and hence health outcomes and costs, for a defined population of children. We focussed our discussion in the opening section on 'population and clinical pathway' around the BTS guidelines, as these could potentially provide a framework around which a model of current practice could be developed, and against which novel home-based devices could be evaluated. This section is not intended as an exhaustive review of the epidemiology, clinical and socio-economic impacts of OSAHS in children.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	224		5.11.1	**Section "Lack of good quality accuracy data on novel devices in children":** Could the EAG clarify whether: 1- The Martinot 2022 study was done in a device approved by regulatory authorities to be used in the UK?; 2- Whether the thresholds for sensitivity and specificity were determined post-hoc and whether they matched the clinically conventionally used ones?	With regard to CE marking and regulatory approval please see our response to comment 100 (Acurable). We provided a critique of the Martinot 2022 study in section 4.4 of our report, noting that this study (along with others) used post hoc diagnostic thresholds, which potentially over-estimates diagnostic accuracy in practice.
Acurable Limited	225		5.11.1	**Section "Effective treatment for children", penultimate paragraph:** The authors state "it is noteworthy that polysomnographic abnormalities resolved without intervention in almost half of children in the watchful waiting arm". This is misleading and should be further discussed since in children time to resolution is very important due to the learning implications it has (there are studies going back almost a century, in the middle of a long list that most readers would miss, at a time when OSA in children wasn't even acknowledged, that observed how these children are not paying attention in class and are falling behind. A child that has behavioural problems and falls behind for years might eventually have the disease resolved but their entire life will be conditioned by it). **Please add more clarification for this.**	This statement relates to results from the CHAT trial (Marcus et al. 2013), discussed in the previous paragraph including the primary outcome, which was assessed at 7 months. For clarity, we have added a note in our report that the follow up period for assessment of PSG normalisation in this study was also 7 months.
Acurable Limited	226		5.11.1	**Section "Effective treatment for children", last paragraph:** Once again the EAG refers to "one of our experts" to come up with a very strong conclusion. Firstly, one person, no matter how knowledgeable that person is, is not enough to come to a conclusion unless that conclusion is based on quantifiable data (and then the nature of that data has to be reported, as well as the methods for quantification). Secondly, who is the expert? It is important to put names into this type of statement so that they are verifiable. Thirdly, what data was used to calculate this? Either more experts are consulted, and the details about how the data was calculated is described, or this section should be removed, especially the sentence "The likelihood of resolving this uncertainty over treatment effectiveness for future modelling might be low", since this sentence is speculative and not evidence based.	This paragraph is clearly labelled as being based on expert experience. We note that this experience is consistent with the pathway for children with disordered breathing in the 2024 ENTUK / RCS Commissioning guide on tonsillectomy (reference 4 in the EAR). For children with mild or moderate signs and symptoms, the ENTUK/RCS pathway recommends to "consider active monitoring prior to a decision on surgery." See response to comment 223 for information on the paediatric experts consulted by the EAG.
Acurable Limited	227		5.11.1	**Section "Longer-term impacts"**: The report includes speculative non-evidence supported strong statements. There is literature that could have been addressed in this section, but it is not used, instead relying on the vague statement "Experts point to…". Please either rewrite this Section using scientifically accepted sources that can be properly referenced or remove completely.	This section was based on comments and references on the epidemiology of OSA in children from the BTS guideline on sleep-disordered breathing in children. The guideline cited sources for a number of long-term adverse effects of OSA in children, but it also noted controversy over the extent and degree of reversibility of these effects (see section 5.11.2 of our report). We have added a reference to the BTS guideline in section 5.11.1 and revised the wording to be clear that the key uncertainties that impact on the potential for modelling are the extent and reversibility of long-term impact of OSA in children.
Acurable Limited	228		5.11.2	**Section Decision Problem, third paragraph**: The oversimplification and under-researched implications of underdiagnosed, undertreated OSA in children (as explained in comments above) could potentially lead to the wrong	See response to comment 223 for information about the paediatric experts consulted by the EAG, and comment 227 regarding our

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				recommendation. The EAG has, for example, not included the long-term consequences (in terms of effects associated with school fatigue and impaired learning outcomes) of OSA in children. Or the fact that a significant proportion of children who are being treated for ADHD don't have ADHD. If these had been properly researched this recommendation might be very different. **Please reconsider this, or remove the paragraph.**	conclusion on the current feasibility of modelling long term impacts of OSA in children.
Acurable Limited	229		5.11.2	**Decision problem, last paragraph:** This paragraph can be misleading since it leaves for interpretation whether what unnamed experts say or what the BTS say is the appropriate comparator for novel technologies, reality oximetry could be assertively discarded as a comparator for future models, since the implications of not doing this can potentially be hugely negative and discriminatory. As the report stands it is clear that there are gaps in knowledge on medical devices, regulatory frameworks and the recent global findings on the performance of PPG based systems, especially oximeters, which we ask be addressed and corrected. Refer to comments about the specific issues of using oximetry to inform the diagnosis in children.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	230		5.11.2	**Decision problem, last three paragraphs:** The key outcomes are questionable considering the EAG has not considered implications of OSA that are really important. There also might be interventions in the form of family counselling for some children not eligible for surgery that might also work. The comments in the last paragraph vaguely address some of the issues brought up before, but these are not properly justified or quantified, and only two references are given when in reality there are many others that contradict this.	See response to comment 227 above regarding outcomes. This section of the report presents EAG suggestions for a scope and potential specification for a future model based on our current understanding of the uncertainties and available evidence. We appreciate that there are differing opinions and evidence sources that we have not identified.
Acurable Limited	231		5.11.2	**Illustrative model structure section:** This is overly simplified since the arguments given by the EAG in previous sections are not properly justified, especially considering existing guidelines were conditioned/restricted by what was possible with what existed.	The illustrative model structure is simple because we do not believe that there is currently sufficient evidence to support a more complicated approach.
Acurable Limited	232		5.11.2	**Parameter requirements, Diagnostic performance:** Please specify the limitations of oximetry: oximeters already lack sensitivity in children, oximeters have not been tested in children and even in adults the standard for regulation's accuracy limit is 4% but only for 67% of the data being compared, and oximeters do not work well in those with dark skin.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	233		5.11.2	**Parameter requirements, Treatment use and effects:** this refers to adverse events during surgery, but how about the opposite? Does the anaesthetic not change in those children that have been identified with OSA to prevent risks of complications? (This is the case in adults, we are unsure about children. But we suggest that it should be looked into.). The last paragraph is missing many possible recommendations because the authors do not seem to have an in-depth, up to date understanding of OSA in children. It is recommended that either they do a proper literature review or eliminate any recommendation they give about any future work, or any hint on	The BTS guideline includes some recommendations on preoperative monitoring of oxygen saturation to predict risk of complications and improve surgical outcomes. We reference the relevant document that contains details of their review and recommendations.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				how to create an economic model. Something over simplistic for a serious problem is worse than nothing at all.	
Acurable Limited	234	191	6.1.1	**Discussion**; Clinical effectiveness evidence; Discussion of principal findings: Third paragraph: The authors say "limited data are available for outcomes relevant to the decision making problem such as time to making diagnosis". Although limited data might be available, this data should not be what an evaluation on an emerging technology is based on, because what they are referring to is that they are missing Trust data, when in reality any data from Trusts, even if it was available might be affected by a number of variables, some of which have nothing to do with the technologies. Illustrations of this: 1- NHS Trusts work differently and have got different amounts of resources, which might lead to a diagnosis being available but the patient not having been told (for example, Trusts in London might have a massive employee churn and new employees need time to be trained); 2- Currently waiting lists are very long due to Trusts not being able to fully recover from COVID. So what would make more sense is the comparison in terms of times between what novel technology would enable, if all other barriers were eliminated. The data for how long on average it takes for Trusts to get the RP systems back, as well as the time in the best-case scenario to have the data marked in the system is easy to obtain because there are many years of experience, and this data is representative of a patient's behaviour already. Any system that has to be returned to download data could be modelled in the same way. Any system that produces an instantaneous diagnosis should have that time discounted. **Please consider taking this into account.** The EAG is saying that some of the "suggested benefits" of novel devices are not founded on empirical evidence, when it comes to the time to produce a diagnosis**. This generalisation is simplistic and not strictly true.** If a device produces an automated diagnosis and this is immediately available to the clinicians, this does not need to be backed up by empirical evidence because the system is like that by design and this design has been audited and approved by regulatory bodies. This is not a manufacturer claim, as regulation is not based on "the word of manufacturers". The EAG should be able to differentiate regulatory from marketing claims. It is wrong to try to avoid differentiating these by making vague statements that put all the devices under the same umbrella. **Please correct this.**	See response to comment 52 for EAG response relating to modelling assumptions on how time to diagnosis is likely to differ between current technologies and the novel devices. We note that estimating this difference is not as simple as extrapolating from technical capabilities of the devices and regulatory evidence, because there will be variation in how new devices are used when adopted in NHS practice. For example, we understand that there is variation in how much reliance clinical services are prepared to place on automated diagnosis. The clinicians who we have spoken with who are already using novel devices for home sleep studies have told us that they access the raw data and manually review sleep studies for some or all patients. The extent of manual validation or coding varies between centres.
Acurable Limited	235		6.1.1	Last paragraph: The way the EAG calculated test failures is incorrect and unscientific. See corresponding section. **On the basis of this, this section needs to change.**	Please see response to comment 128.
Acurable Limited	236		6.2	See previous comments about oximetry. If oximetry is not eliminated as a comparator, **the report needs to be reminding the reader in every section that this should not apply to people with dark skin.**	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	237		6.2	Second paragraph: this is misleading and could potentially compromise the safety of people with dark skin, whilst also being discriminatory, by leading some healthcare centres to adopt oximetry as the diagnostic modality of choice on the basis of costs.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	238		6.2	**Third paragraph needs to be rewritten once the right numbers are input into the models, since the assumptions of the investigators were wrong (the majority) and unjustifiable (others).**	The results and discussion have been updated in light of the updated analyses.
Acurable Limited	239		6.3.1	**Strengths:** It is not a strength to inform the model via consultation with unnamed experts without giving their quantifiable assumptions for those estimations. Furthermore, please could it be clarified why using a previous model is a strength? There is no explanation as to why many of the statements given are a strength and many of the statements are very vague (for example statements along the lines of more recent data being used when found. **Please can this be rewritten and clarified.**	All decision models should undergo some process of discussion with experts in the field to clarify aspects of model structure and assumptions (please see https://pubmed.ncbi.nlm.nih.gov/22990082/). We believe that the attention that versions of this model have received, by experts in the field during development and in consultation processes is a strength. This is clarified in the report (please see section 6.3.1 (now section 6.2.1).
Acurable Limited	240		6.3.2	**Limitations:** This section should be explicit about what assumptions were made that limit the validity of the model, and for which devices. Every novel technology is again put "in the same bucket" despite the fact they are very different. For example, which devices had post-hoc thresholds. Also, the fact that they have limited the analysis of utilities to cardiovascular events, when there are others. **And the validity of any statement that is made about the functionality of characteristics of specific technologies should be verified with the manufacturers, to double check veracity.**	The text in section 6.3.2 (now section 6.2.2) has been edited to reflect this.
Acurable Limited	241		6.3.2	**Limitations:** Second paragraph: does the Martinot study for children also use post-hoc thresholds to quantify performance? If so, this paragraph should state that.	The text in section 6.3.2 (now section 6.2.2) has been edited to reflect this.
Acurable Limited	242		6.3.2	**Limitations:** Last paragraph: This should probably be removed together with the model. It is very simplistic and needs further literature review and research, so no suggestion is better than a potentially bad one.	We have kept this paragraph to summarise our approach given the limitations.
Acurable Limited	243	194	6.4	**Uncertainties:** Last paragraph on this page: mixing the results from devices that are different is not acceptable, not just scientifically but from the point of view of safety for the patients (for this reason, there needed to be a full regulatory submission and clearance for the different products). Just adding this as a limitation is not enough.	We have deleted this paragraph because the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024)) and updated analysis makes the indirect comparison unnecessary. Please see our response to comment 57.
Acurable Limited	244		6.4	Uncertainties: Third paragraph from the bottom: The analysis of uncertainty of what has been done for AcuPebble is not enough. See previous comments, but in a nutshell: 1- PSG comparison data was available. It was provided by the manufacturers (on time). Hence there is no uncertainty on this and this needs to be corrected. 2- Even if it was not, the method used for indirect comparison is incorrect, not just because: - one cannot infer from different nights the difference in accuracy between PSG and PG, - or infer from the results of RP to PSG that the comparison of a technology that works completely differently would be transferable, - or infer that the results from a Chinese population of that comparison would translate to the UK,	This paragraph has been deleted in light of the updated analyses.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				- or use a trial with very low statistical significance; but also because the authors have not understood or even used the correct formula from the paper cited to carry out the “indirect comparison”.	
Acurable Limited	245		7.1	**Conclusions, Implications for service provision:** Second paragraph (first bullet point): This paragraph needs to be removed, unless it clearly states the dangers of using oximeters for those with dark skin (with full explanation).	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	246		7.1	**Conclusions, Implications for service provision:** The third paragraph (second bullet point) has to be more assertive. It is “estimated” underplays the problem.	Please see comment 52 (Acurable) for the EAG response on the potential impact on oximetry and PPG on health inequalities.
Acurable Limited	247		7.1	**Conclusions, Implications for service provision:** Fourth paragraph (third bullet point): Limitations of PAT as a result of extracting it from PPG should be discussed, framing it in the context of the landmark Ioachimescu 2020 trial with 500 people (68). Also, there should be a comment about sustainability here. Disposing of a WatchPAT has a significant environmental cost. This needs to be commented upon.	Most of this paragraph has been deleted, because our updated analysis has changed these results. With regard to Ioachimescu 2020 please see our response to comment 26 and comment 253.
Acurable Limited	248		7.1	**Conclusions, Implications for service provision:** Fifth paragraph (fourth bullet point): A comment (with explanation of what it means) of Sunrise doing post-hoc adjustment of thresholds, so that performance values are given at an optimum point that does not correspond to the diagnostic thresholds and that is what leads to better results, must be made.	We have this paragraph because our updated analysis changed these results.
Acurable Limited	249		7.1	**Conclusions, Implications for service provision:** Sixth paragraph (fifth bullet point): Again here all devices are being put “into the same bucket” and that is neither scientific nor acceptable: 1- The uncertainty is already statistically determined in whatever metrics have been used to report their results. Hence it is not for the EAG to make a comment on this, mostly considering that it was the parameters chosen which were statistically wrong in terms of accounting for performance. Sensitivity and specificity have been long deemed not to be the most representative statistics to determine the performance of diagnostic devices. So this decision is directly introducing “uncertainty” in the device's evidence by developing a model based on poorly informed and in most cases poorly justified scientific assumptions; 2- They are also implicitly conveying the message that all the validation trials done by the devices have the same value in terms of evidence, when in some of the trials there hadn’t even been a power calculation to start with (and if there was this was not presented). **This paragraph has to be eliminated,** and if not the EAG should state that their conclusions should be taken cautiously because the model is simplistic, based in some cases in obsolete references, and with assumptions that in some cases are hard to justify.	Please see our response to comment 42 and comment 69.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited	250		7.2	**Suggested research priorities: This section has several critical issues, described below.** The EAG may not have sufficient knowledge about regulations, medical devices, ethics, and OSA to have the expertise to create a suggested research priorities section. However, this being potentially put forward in a report endorsed by NICE can be taken by others as "dogma", hence having it here can have a negative impact on the research community overall. **On the basis of this, it should be removed.** A brief analysis of why is given point by point below (this can be elaborated upon on request): 1- **Additional research to compare diagnostic accuracy of RP against PSG.** This is questionable considering that there are many factors that will affect the "accuracy" that have nothing to do with the technology of RP with respect to PSG. Amongst others night to night variability. There are situations in which RP will give a better representation of the state of the disease in a particular patient than PSG would, even leading to a different diagnosis, since RP at home might allow a more natural sleep and consequently better represent what truly happens to the patient. Hence, what would be the value of posing this scientific question, beyond what has been done already? 2- **Research on diagnostic accuracy of newer versions of devices.** This might be completely unnecessary. It depends on what changes a certain device has with respect to others. The evidence is already there because the devices have undergone a regulatory process which requires this evidence. This question should have been asked by the EAG to the manufacturers, not just because of this point but because if there are relevant differences already affecting accuracy the new numbers should have been used for the whole evaluation. 3- **Diagnostic accuracy of novel devices when used in the home.** The answer to the broader question they pose in this point as a research priority is already known: the answer is not generalizable since it depends on the technological characteristics of the devices, so it makes no sense to try to infer how the results would convert on the basis of what happened with devices that are technologically different. 4- **Home-based studies on children.** The investigators propose something that has already been widely considered. There is a fundamental reason why comparison studies in children are done in the clinic, mostly in those with comorbidities: someone needs to be making sure the children keep the RP-PG sensors on! Why would an ethics committee approve this? 5- **Indirect comparisons between novel and conventional devices and reference standards.** This makes no scientific sense because one cannot extrapolate to infer performance a comparison that is evaluating a technology that is completely different.	We emphasise that the EAG report is not 'endorsed' by NICE and entirely agree that it should not be taken as 'dogma'. The EAG is commissioned as an independent research organisation to conduct an assessment, which the NICE Diagnostic Advisory Committee will consider alongside advice and other sources of information. As with the rest of the report, the research priorities reflect the EAG's best understanding of the evidence, and we have sought to take account of feedback from all stakeholders. **Point 1**: Please see our response to comment 255 (BTS) below. We acknowledge these difficulties, but as highlighted in section 5.7.4, current evidence to compare novel devices against RP with a common reference standard is very weak. **Point 2**: We agree that the answer to the question of transferability of diagnostic accuracy evidence from older to newer versions of devices depends on what changes have been made to the device. We did ask this question of manufacturers where evidence was not available for newer versions of devices. However, we raise this as a research priority going forward. **Point 3**: We consider the transferability of evidence from sleep studies conducted in different settings to be an important, given that most evidence for devices intended for use in a home setting was collected in a clinic. **Point 4**: The transferability of evidence from clinic to home settings is an important issue for children, as for adults. However, we do understand that there may be practical and ethical challenges in designing studies to answer this question. **Point 5**: The use of indirect comparisons is standard in many NICE technology appraisals and diagnostic evaluations, where head-to-head comparative data is not available. Care is needed to identify, and where possible to adjust for, heterogeneity between studies, and there are some particular issues for meta-analysis of diagnostic accuracy data (see section 3.4.23 in the NICE health technology evaluation manual). See our response to comment 255 below for further discussion.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Regarding the conclusions in children, the EAG has not based these comments on proper literature review. They base them on conversations with an expert, but the expert is unnamed and the questions asked are not documented.	The additional research priorities for children are based on our own observations on gaps in data that would be required to develop a full health economic model: informed by our systematic review of health-related quality of life (section 5.2) and commentary and references on long-term impacts of OSAHS in the BTS guideline (Evans et al 2023).
Acurable Limited	251		8	**References:** It was stated throughout the report that there was no data comparing AcuPebble with PSG. That data was provided to NICE (within deadlines), so it was assumed in previous comments that maybe this data had not been passed on to the EAG. However, NICE certainly did pass this data on, because the name of the document provided appears as references 26,30.	Please refer to our response to comment 57 on the inclusion of the Phase 1 Virgen Macarena trial (Sanchez Gomez et al (2024))
Acurable Limited	252		8	**References:** It was implied throughout the report that there were no reports or trials of devices involving a wide range of co-morbidities and technologies had mostly been tested on patients with no-comorbidities. This again was surprising because AcuPebble SA100 has been validated in both adults and children with a very wide range of co-morbidities and these comorbidities were provided to NICE. The proof that the EAG had had access to documents with this information, at least for one trial is reference 24.	Please see our response to comment 114
Acurable Limited	253	248	Appendix 2, Table 50 and 51	Table 50 Inclusion/Exclusion screening worksheet: The eligible interventions include WatchPAT 200. However, Table 51 states that the reason for the exclusion of Ioachimescu 2020 is the intervention, which is listed as WatchPAT 200. Please could the EAG confirm why this study was excluded and why there are excluded studies which, based on the flow described in Table 50, should be included? Note we have not checked every single study in this list, so this may be the case for other studies as well.	Regarding the inclusion/exclusion screening worksheet, under the 'Intervention' heading it says "*WatchPAT 100, 200, 200U (Zoll/Itamar - these devices are not listed in the scope but we will include them at this stage)".* This was a note for the systematic review team to "include" studies of those versions so that we could set them aside to be considered for *possible inclusion* in the review in the likely event of there being limited/no data for WatchPAT 300/ONE (see next paragraph). It was not intended to suggest that all of these versions would be included (hence *"…at this stage").* We recognise, nonetheless, that this may appear confusing and have therefore removed this note from the inclusion/exclusion screening worksheet. At the completion of study screening, it became apparent that there was insufficient available evidence for WatchPAT 300/ONE, specifically on diagnostic performance. Hence it became necessary to consider the studies of older versions we had set aside. At this time, we made a reasoned decision to include WatchPAT 200U studies as supporting evidence, but not to include the earlier versions (e.g. WatchPAT 200), which we understood used a different algorithm. We acknowledge that this wasn't clear in our report and we have made this more explicit in the Erratum: see section 4.1 for an explanation of this decision.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

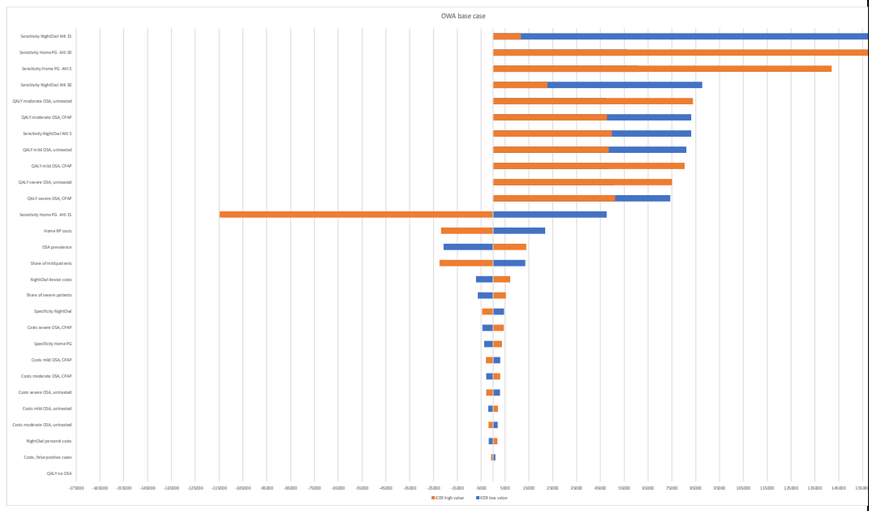
Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					Ioachimescu 2020 was therefore not included in the systematic review as it uses WatchPAT 200, as confirmed by the study authors (see response to comment 26 above). However, Ioachimescu 2020 was used as a source of evidence for test failure rates in the economic evaluation. This was because evidence on failure rates in the study by Mueller (2022) (WatchPAT 300) had been overlooked, and no data in the study by Storey et al (WatchPAT ONE) or supporting evidence studies of WatchPAT 200U (Pillar (2020) and Tauman (2020). We selected Ioachimescu 2020 because it had a significantly large sample size compared to other available studies. This was a pragmatic decision taking into account the overall balance of strengths and limitations across the evidence base. Please note that data from Mueller 2022 are used in the updated results for both of the WatchPAT devices (See Section 5.7.5 for more details).
British Thoracic Society	254	12, bullet 2	Service implications	This may be beyond the EAR's scope but there is a broader discussion to be had about implications of missing mild OSA and downgrading mod-severe OSA to mild. Firstly, the latter is of questionable relevance since NICE (NG 202) recommended CPAP as first line treatment in mild disease as well, depending on symptom severity. Furthermore, at all severity levels including mild OSA, this point regarding service implications could be misinterpreted as inferring that CPAP is the essential and only useful treatment for OSA symptoms and, importantly, downstream health consequences. This is oversimplistic and inconsistent with published evidence. It is well recognised that the approach to patients with OSA should, if done properly, be multifaceted, addressing lifestyle factors, sleep hygiene, weight loss, and cardiovascular (CVS) risk modification. This applies particularly in milder cases, where CPAP is more likely to be inappropriate, and in all cases where CPAP is clinically inappropriate, declined or not tolerated. The evidence regarding the benefits of CPAP in milder OSA (including when defined as such based on symptom level rather than sleep study findings) remains doubtful, particularly regarding CVS risk. In any case, the evidence points increasingly to hypoxic burden as being, alongside consideration of all CVS risk factors (particularly BP), the key determinant of whether CPAP might have potential to positively impact CVS outcomes in an individual (Randerath ERJ 2021, Xu et al Thorax 2023, Trezepizir AJRCCM 2021, Azarbarzin EHJ 2019). Therefore it could be argued that oximetry is sufficient when employed alongside expert evaluation of symptom level and with reference to CVS risk. This was fed back to the NG202 guideline committee at the draft consultation stage and remains relevant.	Thank you for these comments. We appreciate the importance of a taking a multifaceted approach to treatment of OSA and do not assume that CPAP is the only or essential treatment. In our analysis, we assume that all people diagnosed with OSA would be offered 'conservative management' (which may include advice on lifestyle, sleep hygiene and risk modification) and that some people diagnosed with mild OSA, and all people diagnosed with moderate or severe OSA, would also be offered treatment with CPAP or MAD (see section 5.7.11 in our report). This does mean that both missing mild OSA and downgrading moderate-severe OSA would result in some patients not being offered treatments from which they would benefit.
British Thoracic Society	255	13 and elsewhere in report	Research recommendations	Bullets 1&3: The EAR makes it clear that heterogeneity of published data is a major factor affecting the DAP70 analyses. Unfortunately it is not clear that research can adequately address this issue, which is due in large part to inescapable real life confounders. These include night to night variability of OSA, differing respiratory event scoring criteria (partly influenced by tech used) and home vs. clinical testing scenarios. While the widening range of sleep diagnostics fuels expectations of ever more precise measurement, this may be missing the point. OSA severity criteria are arbitrary and correlate poorly with symptoms, important health consequences and treatment outcomes. Ref	The suggestions in this section are based on key evidence gaps that we encountered in addressing the decision problem set by NICE. They are suggestions for priorities to reduce uncertainty over the relative clinical and cost-effectiveness of the included tests, rather than specific research recommendations. It is beyond our scope, and we do not have the expertise, to provide more specific guidance on study design. We have changed the heading of this subsection in the Scientific Summary to "Suggested research priorities" to match that in section 7.1.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				comment 1, the assessment and management of OSA is (or should be) more complicated than the existence of CPAP and polysomnography might suggest. However, if the view remains that this recommendation is important and achievable, then in its current form it is very general. More precise guidance on what is required would help guide potential investigators in designing trials that might provide data useful to future NICE evaluations. Bullet 2: Suggest seek data from manufacturer if not already done. Could also get more product info from them - if little difference in key tech. features then may be irrelevant. General: Oximetry is still used by a significant proportion of NHS sleep services. It is cheaper and more accessible than other tests, which is of particular importance given the current waiting list pressures on sleep diagnostics. The lack of comparative evidence for oximetry has been highlighted in this report. Given these points and the high levels of uncertainty recognised to be affecting many areas of this report, should the EAR include oximetry in its research recommendations? Evaluation of oximetry vs. novel wearables vs. polygraphy is suggested to provide data on the clinical and cost effectiveness, as well as ease of application and extraction of data for other purposes such as assessing hypoxic burden and quantifying CV risk for example.	We agree that the 'real life confounders' are difficult to escape. Nevertheless, the existing evidence base presents a challenge in estimating the relative effectiveness and cost-effectiveness of the novel devices compared with RP and oximetry: given that most of the evidence on the diagnostic accuracy of the novel devices comes from comparisons with a PSG reference standard, and that the evidence base for the diagnostic accuracy of home RP and oximetry compared with PSG is weak (see section 5.7.4), We recognise that there are significant challenges in resolving the heterogeneity of current evidence on diagnostic accuracy, and that making indirect comparisons between tests might not be feasible. However, there would also be challenges in designing a trial, given heterogeneity of patient populations, variations in practice, and differing opinions on the appropriateness of oximetry. We have added further comment on these points in the suggested research priorities sections in the Scientific Summary and in section 7.2.
British Thoracic Society	256	90-91	5.1.3, points 4 and 8 re repeating test if negative oximetry	To assume that all negative oximetry tests would lead to a repeat test (oximetry or higher level) is flawed. It does not represent clinical practice in NHS sleep services currently using oximetry as entry level sleep diagnostic. Ref above points in support of this. Clinical context (nature/extent of symptoms; other cardiovascular risk factors (particularly refractory BP which may respond to CPAP)) is essential part of assessment alongside sleep study results, and hypoxic burden (detected by oximetry) is increasingly supported by evidence to be key determinant of CVS risk. Similar feedback was given by stakeholders at the draft consultation stage of NG202.	Thank you for his comment, highlighting the different factors involved. To clarify, in our analysis we assume that there would be repeat tests for people who truly have moderate-severe OSAHS but who have been misdiagnosed as having no OSAHS. If oximetry is negative for people without OSAHS or for people with mild OSAHS, we do not assume any repeat tests.Our inclusion of a repeat test for those with moderate-severe OSAHS who are diagnosed with no OSAHS, is based on an assumption that these individuals would continue to have symptoms which would lead to a repeat sleep study.
British Thoracic Society	257	149	5.7.13	We accessed the NHS supply chain and found the cost of a Nonin 3150 oximeter device to be [redacted] to [redacted]	Thank you for supplying this information. Using a cost of [redacted] for the oximeter device leads to a slight increase in the estimated cost per sleep study with oximetry from £17.86 (assuming the patient collects the device in person, as in our base case analysis) to [redacted]. This is due to assumptions on the number of times each device would be used, and the life years for the device. Given that all novel devices are estimated to be cost-effective compared to oximetry, this very small difference to the costs has minimal impact on the results.
British Thoracic Society	258	256	5.7.15	Cost of semi-bespoke mandibular advancement splints. The device used in TOMADO (confirmed by TQ, CI of TOMADO, stakeholder organisation member) was the same device although the molding kit was an earlier iteration, superseded by currently available Sleeppro Custom/S2. Advise use up to date price for this.	Thank you for this information. We do not model semi-bespoke devices in the analysis, only the customised devices. However, this was unclear in the report and we have updated our report to reflect this. Please see section 5.7.15.

ResMed	259	72	4.5.1 Diagnostic accuracy (people over 16 years of age)	***Dijnostic accuracy of the devices in scope of this assessment was measured with different markers (AHI for some devices and ODI for other devices). It seems v*** to base a comparison of innovative devices on the same diagnostic marker instead of mixing diagnostic markers and treating them as being interchangeable. Moreover, scoring rules for sleep studies have been changed over the years in regard to the definition of hypopneas. The original rules were known as the Chicago Criteria from 1999, then there were the American Academy of Sleep Medicine (AASM) 2007 rule and the current guideline is the AASM 2012 rule. The differences in the scoring rules are summarised (Benjafield et al. Lancet Repir Med 2019): 1) **1999 guideline:** ≥50% decrease in flow OR a clear reduction in flow that does not reach ≥50% AND is associated with either an oxygen desaturation of ≥3% or an arousal. 2) **2007 guideline:** ≥30% decrease in flow from baseline with an associated oxygen desaturation of ≥4%. 3) **2012 guideline:** ≥30% decrease in flow from baseline with an associated oxygen desaturation of ≥3% OR an associated arousal. As published by Duce et al. in JCSM 2015 the different scoring rules effect the AHI result and hence prevalence and severity classification. They showed the median AHI from AASM 2012 guideline was approximately 90% greater than the AASM 2007 guideline and approximately 15% lower than the Chicago 1999 guideline. This makes it challenging to compare sensitivity and specificity results from different studies with different technologies if different scoring rules were used. When considering ODI as the metric, as above with AHI scoring there are different scoring rules that have been utilised over time. ODI based on ≥4% desaturation and ODI based on ≥3% desaturation. As published by Ling et al. in SLEEP 2012 there are interrelationships between different oxygen desaturation thresholds, body mass and AHI. They showed that there are differences in the accuracy of ODI thresholds for detecting moderate to severe OSA based on AHI. ODI 3% achieved a significantly higher area under the curve (AUC) of receiver operating characteristics than ODI 4% for moderate to severe OSA. ODI 3% also achieved significantly higher AUC than ODI 4% in non-obese subjects. This highlights that there are differences between the ODI 3% and ODI 4% and further confirms the importance of identifying the scoring rules used for comparisons. When it comes to the decision of which diagnostic marker is best, consistency across diagnostic and therapy may be appropriate. As AHI is frequently used across the globe, and into the NICE definition of OSAHS severity, the cost-effectiveness analysis should be based on AHI (NICE NG 202 Guideline, 2021). Finally, there are confounders when it comes to diagnostic accuracy of ODI. Notably, BMI has an influence on diagnostic accuracy of ODI (Ling et al. 2012). This might introduce additional uncertainty and bias in the model as the diagnostic validation studies featured patient populations with different BMI levels.	Thank you for this useful information. We are mindful of the differences between the diagnostic indices and we agree that ideally assessment of the performance of the novel devices should use the same marker. We are limited, however, by the variety of measures the respective study investigators have chosen to report. We have given greater emphasis to this issue in the presentation of our findings and discuss it this as a limitation of our review findings. For instance, this detail is now added to Tables 17 and 18 in the discussion of accuracy data used in the model (please see Section 5.7.3).

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Conclusion: AHI and ODI are not interchangeable. Therefore, a comparison of innovative diagnostics should be based on the same diagnostic marker to avoid introduction of additional uncertainty and bias. **References:** 1. Massie F, Van Pee B, Bergmann J. Correlations between home sleep apnea tests and polysomnography outcomes do not fully reflect the diagnostic accuracy of these tests. J Clin Sleep Med. 2022;18(3):871–876 2. Rashid NH, Zaghi S, Scapuccin M, Camacho M, Certal V, Capasso R. The Value of Oxygen Desaturation Index for Diagnosing Obstructive Sleep Apnea: A Systematic Review. Laryngoscope. 2021 Feb;131(2):440-447.). 3. Letter from Professor Chris Carlin 4. Benjafield et al. estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis Lancet Respir Med 2019 Aug;7(8):687-698 doi: 10.1016/S2213-2600(19)30198-5 5. Duce et al. The 2012 AASM Respiratory Event Criteria Increase the Incidence of Hypopneas in an Adult Sleep Center Population J Clin Sleep Med 2015;11(12):1425 –1431 6. *NICE NG 202 Guideline, 2021, https://www.nice.org.uk/guidance/ng202/chapter/terms-used-in-this-guideline#sleep-study* Ling IT, James AL, Hillman DR. Interrelationships between body mass, oxygen desaturation, and apnea-hypopnea indices in a sleep clinic population. Sleep. 2012 Jan 1;35(1):89-96.	
Resmed	260	89-190	5 ECONOMIC ANALYSIS	***There is high uncertainty in relation to the diagnostic accuracy data of the devices in scope of this assessment. Furthermore, scenario analyses revealed that the outcome of the health economic model is very sensitive even to small changes in diagnostic accuracy input parameters. Therefore, a more robust approach to estimating diagnostic accuracy of the innovative devices in scope and the comparator should be considered.*** It has been stated in the EAR that the diagnostic accuracy (sensitivity and specificity) data for the devices in scope of this assessment are associated to a high degree of uncertainty (see e.g. pages 5, 12, 76, 77, 88 of the EAR). This uncertainty is of high relevance for the assessment as the model input parameter with the highest impact on the model outcomes is sensitivity of the test strategies compared in the model. Even small changes in this input parameter have substantial impact on the model outcome demonstrating that the model is not robust in its current form. As it was not feasible to code a full sensitivity analysis in the model provided for review within the short time for providing feedback, a simplified model version was developed. This simplified version featured a similar decision tree for assessing the diagnostic outcome of test strategies. The decision tree was used to calculate the distribution of patients over the OSA severity grades (from no OSA to severe OSA) according to test result and true OSA severity grade similar to the original model. The distribution of patients was linked to the cost and	Thank you for this examination of the results of the economic model, and for highlighting Massie et al 2018. Regarding the choice of inputs for the sensitivity and specificity of NightOwl, we chose to use data from Lyne et al. 2023 in the base case analysis. The version of NightOwl used in Lyne et al 2023 is the disposable version to be launched in the UK. Massie et al 2018 and van Pee et al. 2022 use the reusable version of NightOwl. Since the company confirmed to NICE that the only difference between the reusable and disposable NightOwl devices is whether the battery can be re-charged, and that the sensors and software are identical, we have conducted scenario analyses using data from Massie et al 2018 and van Pee et al 2022. The results can be found in Tables 44 and 45 in Section 5.10.2. As we state in the Report: “When the data from Massie et al 2018 are used to inform the accuracy of NightOwl, it is estimated to dominate respiratory polygraphy: it is associated with greater QALYs at lower cost. This is due to the higher sensitivity estimates at both AHI diagnostic cut-offs for NightOwl from Massie et al, than from Lyne 2023, see Table 44. When van Pee 2022 data inform the performance of NightOwl, NightOwl is estimated to be less costly (by £108) and

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				QALY outcomes of the different treatment strategies according to true underlying OSA severity grade similar to the approach taken in model provided for review. The main simplification in the test model was that only CPAP vs. no treatment was considered for any OSA severity and that only two alternative diagnostic strategies were compared (Home RP and NightOwl). The outcome of the sensitivity analysis varying input parameters by +/-20% is outlined in the figure below.  Further scenario analysis conducted manually in the model provided by NICE substantiated the findings from the simplified model version: **Scenario 1:** For this scenario, we used the contingency table from Massie et al. 2018 to update the sensitivity and specificity in the model. Sensitivity low threshold: 0.978 Specificity low threshold: 0.800 Sensitivity high threshold: 0.967 Specificity high threshold: 0.829 Further, test costs for NightOwl were adapted to reflect no additional costs for re-testing as discussed in part 3 of this response. **Result:** NightOwl becomes a dominant strategy vs. Home RP and ranks 1 in terms of INMB at an ICER of 20k GBP. |	Costs				QALY	vs RP		vs oximetry		Time to treatment (months)	ICER				INMB at £20,000
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Device/test	Diagnosis	Treatment	CVE	HSA	Total	Total	Incr costs	Incr QALYs	Incr costs	Incr QALYs		v oximetry	v HomeRP	v oximetry	v HomeRP	v oximetry
Oximetry	£ 64,31	£ 1.901	£ 5.245,71	£ 358,08	£ 7.571	14,055					6,17		£ 8.488			£ -
Home RP	£ 224,21	£ 2.611	£ 5.240,68	£ 284,60	£ 8.361	14,144					6,00	£ 8.488		£ 8.488		£ 1.075
Acupebble SA100	£ 84,38	£ 2.394	£ 5.242,34	£ 307,48	£ 8.029	14,114	-£ 332	-0,029	£ 457	0,064	6,01	£ 7.194	£ 11.283	£ 7.194	not cost-effective	£ 814
Bricsy	£ 76,29	£ 2.425	£ 5.241,04	£ 302,90	£ 8.045	14,124	-£ 315	-0,019	£ 474	0,074	6,00	£ 6.449	£ 16.182	£ 6.449	not cost-effective	£ 996
NightOwl	£ 108,93	£ 2.684	£ 5.240,07	£ 275,89	£ 8.309	14,154	-£ 51	0,010	£ 738	0,103	6,00	£ 7.150	£ 5.017	£ 7.150	Dominant	£ 1.326
Sunrise	£ 97,46	£ 2.579	£ 5.240,78	£ 287,00	£ 8.195	14,139	-£ 166	-0,004	£ 624	0,089	6,00	£ 7.034	£ 38.402	£ 7.034	not cost-effective	£ 1.192
WatchPAT 300	£ 90,64	£ 2.713	£ 5.242,23	£ 277,38	£ 8.323	14,146	-£ 38	0,002	£ 752	0,095	6,00	£ 7.893	£ 16.888	£ 7.893	Dominant	£ 1.159
WatchPAT ONE	£ 113,39	£ 2.713	£ 5.242,23	£ 277,38	£ 8.346	14,146	-£ 15	0,002	£ 774	0,095	6,00	£ 8.132	£ 6.688	£ 8.132	Dominant	£ 1.130	 **Scenario 2:**	less effective (by 0.001 QALYs) than respiratory polygraphy, but the reduction in QALYs is estimated to be cost-effective compared to the reduction in costs, see Table 45. This is driven by a slight increase in the sensitivity of NightOwl at the AHI ≥ 15 diagnostic cut-off when van Pee data are used (0.91) compared to when data from Lyne are used (base case analysis, 0.89), and a decrease in the specificity of NightOwl at the AHI ≥ 15 diagnostic cut-off (0.76 from van Pee and 0.82 from Lyne in the base case). This seemingly unintuitive finding is driven by the fact that with more people having mild OSAHS being misdiagnosed as having moderate-severe OSAHS they have a greater chance of receiving CPAP (rather than conservative management), which is associated with utility gains. This pattern can be seen with the scenario analysis where more people with mild OSAHS are assumed to receive CPAP than in the base case analysis, please see section below." We have highlighted this in the discussion of the Scenario analysis results, and in our overall discussion. We agree with ResMed that there is indeed high uncertainty in the model results depending on the estimates of accuracy used for NightOwl (and also for home respiratory polygraphy). This is mainly driven by the very small differences in QALYs between the novel devices and home respiratory polygraphy.. We have stated our reasons for choosing Xu over Pereira 2013 or the values used in the NG202 economic model (please see Section 5.7.4). As with the use of alternative sources for the accuracy of NightOwl, we highlight in our scenario analysis results, and our discussion, the impact of these different data sources on the model results. With respect to the point on the non-linearity of the results, please see our response to comment 42. With regard the point about ranking, we have not reported a fully incremental analysis due to the many differences between the data source used in the model for the devices, and do not provide ranks for the devices in the report. Although, a fully incremental analysis can be seen in the excel model, that is a legacy of the NG202 economic model and we do not refer to it.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				For this scenario, we used the sensitivity and specificity data for Home RP that was calculated based on the meta-analysis conducted for the previous NICE OSA guideline update: Sensitivity low threshold: 0.945 Specificity low threshold: 0.580 Sensitivity high threshold: 0.842 Specificity high threshold: 0.890 Further, test costs for NightOwl were adapted to reflect no additional costs for re-testing as discussed in part 3 of this response. **Result:** NightOwl becomes a dominant strategy vs. Home RP and ranks 4 in terms of INMB at an ICER of 20k GBP.	

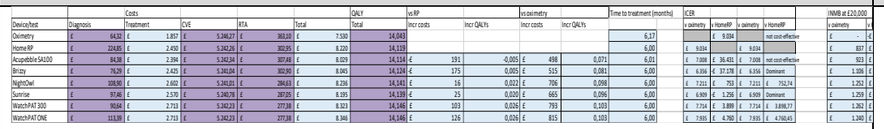
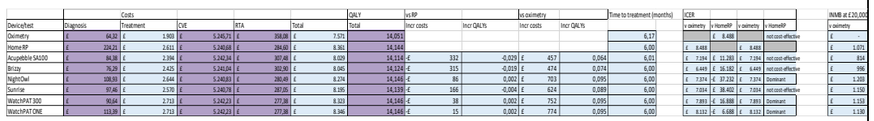
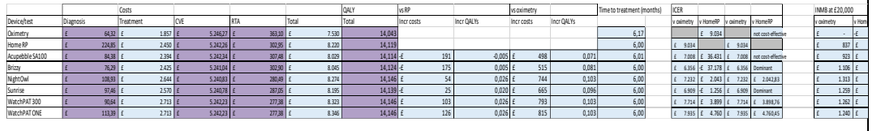
	Costs					QALY	vs RP		vs oximetry		Time to treatment (months)	ICER				INMB at £20,000	
Device/test	Diagnosis	Treatment	CVE	RTA	Total	Total	Incr costs	Incr QALYs	Incr costs	Incr QALYs		v oximetry	v HomeRP	v oximetry	v HomeRP	v oximetry	v
Oximetry	£ 64,32	£ 1.857	£ 5.246,27	£ 363,10	£ 7.520	14,043					6,17		£ 9.034		not cost-effective	£ -	£
Home RP	£ 224,85	£ 2.450	£ 5.242,26	£ 302,95	£ 8.220	14,119					6,00	£ 9.034		£ 9.034		£ 837	£
Acupebble SA100	£ 84,38	£ 2.394	£ 5.242,34	£ 307,40	£ 8.029	14,116	£ 191	-0,005	£ 498	0,071	6,01	£ 7.038	£ 36.431	£ 7.038	not cost-effective	£ 923	£
Bricsy	£ 76,29	£ 2.425	£ 5.241,04	£ 302,90	£ 8.045	14,124	£ 175	0,005	£ 515	0,081	6,00	£ 6.356	£ 37.378	£ 6.356	Dominant	£ 1.106	£
NightOwl	£ 108,89	£ 2.557	£ 5.241,48	£ 289,72	£ 8.197	14,134	£ 23	0,015	£ 667	0,091	6,00	£ 7.322	£ 1.558	£ 7.322	Dominant	£ 1.154	£
Sunrise	£ 97,46	£ 2.570	£ 5.240,79	£ 287,05	£ 8.195	14,139	£ 25	0,020	£ 665	0,096	6,00	£ 6.909	£ 1.256	£ 6.909	Dominant	£ 1.259	£
Watch PAT 300	£ 90,64	£ 2.713	£ 5.242,23	£ 277,28	£ 8.323	14,146	£ 103	0,026	£ 793	0,103	6,00	£ 7.714	£ 3.899	£ 7.714	£ 3.898,77	£ 1.262	£
Watch PAT ONE	£ 113,39	£ 2.713	£ 5.242,23	£ 277,28	£ 8.346	14,146	£ 126	0,026	£ 815	0,103	6,00	£ 7.935	£ 4.760	£ 7.935	£ 4.760,45	£ 1.240	£

Scenario 3:
For this scenario, sensitivity and specificity data for NightOwl were calculated based on the contingency tables from Massie et al. 2018, van Pee et al. 2022 and Lyne et al. 2023**:**
Sensitivity low threshold: 0.935
Specificity low threshold: 0.803
Sensitivity high threshold: 0.915
Specificity high threshold: 0.809
Further, test costs for NightOwl were adapted to reflect no additional costs for re-testing as discussed in part 3 of this response.
Result: NightOwl was not cost-effective at a WTP of 20k GBP vs. Home RP and ranks 3 in terms of INMB at an ICER of 20k GBP.

	Costs					QALY	vs RP		vs oximetry		Time to treatment (months)	ICER				INMB at £20,000	
Device/test	Diagnosis	Treatment	CVE	RTA	Total	Total	Incr costs	Incr QALYs	Incr costs	Incr QALYs		v oximetry	v HomeRP	v oximetry	v HomeRP	v oximetry	v
Oximetry	£ 64,32	£ 1.903	£ 5.245,71	£ 358,09	£ 7.571	14,055					6,17		£ 8.448		not cost-effective	£ -	£
Home RP	£ 224,21	£ 2.611	£ 5.240,68	£ 284,60	£ 8.361	14,144					6,00	£ 8.448		£ 8.448		£ 1.071	£
Acupebble SA100	£ 84,38	£ 2.394	£ 5.242,34	£ 307,40	£ 8.029	14,114	£ 312	-0,029	£ 457	0,064	6,01	£ 7.194	£ 11.283	£ 7.194	not cost-effective	£ 814	£
Bricsy	£ 76,29	£ 2.425	£ 5.241,04	£ 302,90	£ 8.045	14,124	£ 315	-0,019	£ 474	0,074	6,00	£ 6.449	£ 16.182	£ 6.449	not cost-effective	£ 996	£
NightOwl	£ 108,90	£ 2.602	£ 5.241,01	£ 294,63	£ 8.236	14,143	£ 124	-0,003	£ 665	0,096	6,00	£ 7.358	£ 47.337	£ 7.358	not cost-effective	£ 1.142	£
Sunrise	£ 97,46	£ 2.570	£ 5.240,79	£ 287,05	£ 8.195	14,139	£ 166	-0,004	£ 624	0,089	6,00	£ 7.034	£ 39.403	£ 7.034	not cost-effective	£ 1.150	£
Watch PAT 300	£ 90,64	£ 2.713	£ 5.242,23	£ 277,28	£ 8.323	14,146	£ 38	0,002	£ 752	0,095	6,00	£ 7.893	£ 16.888	£ 7.893	Dominant	£ 1.153	£
Watch PAT ONE	£ 113,39	£ 2.713	£ 5.242,23	£ 277,28	£ 8.346	14,146	£ 15	0,002	£ 774	0,095	6,00	£ 8.132	£ 6.688	£ 8.132	Dominant	£ 1.130	£

Scenario 4:
For this scenario, we used the sensitivity and specificity data for Home RP that was calculated based on the meta-analysis conducted for the last NICE OSA guideline update (see scenario 2 for updated input). Sensitivity and specificity data for NightOwl were calculated based on the contingency tables from Massie et al. 2018, van Pee et al. 2022 and Lyne et al. 2023 (see scenario 3 for updated input). Further, test costs for NightOwl were adapted to reflect no additional costs for re-testing as discussed in part 3 of this response.
Result: NightOwl has an ICER of 753 GBP vs. Home RP and ranks 3 in terms of INMB at an ICER of 20k GBP.

Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				 **Scenario 5:** For this scenario, sensitivity and specificity data for NightOwl were calculated based on the contingency tables from Massie et al. 2018 and van Pee et al. 2022: Sensitivity low threshold: 0.949 Specificity low threshold: 0.750 Sensitivity high threshold: 0.929 Specificity high threshold: 0.790 Further, test costs for NightOwl were adapted to reflect no additional costs for re-testing as discussed in part 3 of this response. **Result:** NightOwl was a dominant strategy vs. Home RP and ranks 1 in terms of INMB at an ICER of 20k GBP.  **Scenario 6:** For this scenario, we used the sensitivity and specificity data for Home RP that was calculated based on the meta-analysis conducted for the last NICE OSA guideline update (see scenario 2 for updated input). Sensitivity and specificity data for NightOwl were calculated based on the contingency tables from Massie et al. 2018 and van Pee et al. 2022 (see scenario 5 for updated input). Further, test costs for NightOwl were adapted to reflect no additional costs for re-testing as discussed in part 3 of this response. **Result:** NightOwl has an ICER of 2,043 GBP vs. Home RP and ranks 1 in terms of INMB at an ICER of 20k GBP.  **In summary, small and reasonable changes in sensitivity and specificity of the diagnostic devices yielded substantial changes in model outcomes and rank in terms of INMB of the technologies in scope of the assessment.** All scenarios considered were based on published evidence. In addition, it seems that there is a U-shaped association of diagnostic accuracy parameters and the ICER calculated by the model. The non-linearity of this association makes the model outputs very hard to interpret (especially given the uncertainty of the model input).	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Furthermore, it seems that the rank according INMB is not the best approach to summarize the model results as the uncertainty regarding diagnostic accuracy input and the unclear relationship of most impactful model input and model output is not fully understood. This is demonstrated with scenario 2, where sensitivity and specificity of the comparator (Home RP) were changed to values from an evidence synthesis. This resulted in substantial changes in ICERs and rank according to INMB for the innovative diagnostic devices without any significant changes in input parameters for the innovative devices (variation of retest costs for NightOwl alone does not change the model output substantially). This finding is further substantiated with scenarios 5 and 6. The diagnostic accuracy for NightOwl was the same in both scenarios but the diagnostic accuracy for Home RP was higher in scenario 5. In contrast to what would be expected the cost-effectiveness of NightOwl improved to dominance when the diagnostic accuracy was increased for the comparator Home RP. This underlines the non-linear relationship of diagnostic accuracy and model outcome again. Instead of using individual studies to populate input variables of model, we would suggest to synthesise data or take the most robust study whenever more than one study is available. This would be applicable to the Home RP comparator and for NightOwl. We are not aware if this would apply to the other innovative devices, too. As an example, the diagnostic accuracy for Home RP in the previous NICE OSA guideline update was calculated based on a meta-analysis. This seems to be a more robust approach than selecting a single study as model input. For NightOwl, data synthesis of the studies by Massie et al. 2018, van Pee et al. 2022 and Lyne et al. 2023 may be considered since the NightOwl mini and reusable are essentially the same device relying on the same sensor, algorithms and design. Alternatively, the most robust study out of the three may serve as input for the model. We would support Massie as the most robust study out of the three as it included more patients than the two other studies and used double scoring to minimize interscorer variability. The Lyne et al. study seems the least robust as it enrolled the lowest number of patients out of the three and did not rely on double scoring to minimize interscorer variability. **Conclusion:** The model provided for review is very sensitive to small changes in sensitivity and specificity input data with substantial impact on model outcome. Furthermore, there is considerable uncertainty around the sensitivity and specificity input data for all diagnostics considered in this assessment including Home RP. **References:** 1. Massie F, Mendes de Almeida D, Dreesen P, Thijs I, Vranken J, Klerkx S. An evaluation of the NightOwl home sleep apnea testing system. J Clin Sleep Med. 2018;14(10):1791–1796.	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				2. NICE NG 202 Guideline, 2021, *https://www.nice.org.uk/guidance/ng202/chapter/terms-used-in-this-guideline#sleep-study* 3. Bart Van Pee, Frederik Massie, Steven Vits, Pauline Dreesen, Susie Klerkx, Jagdeep Bijwadia, Johan Verbraecken, Jeroen Bergmann, A multicentric validation study of a novel home sleep apnea test based on peripheral arterial tonometry, Sleep, Volume 45, Issue 5, May 2022, zsac028, https://doi.org/10.1093/sleep/zsac028 Lyne CJ, Hamilton GS, Turton ARE, et al. Validation of a single-use and reusable home sleep apnea test based on peripheral arterial tonometry compared to laboratory polysomnography for the diagnosis of obstructive sleep apnea. J Clin Sleep Med. 2023;19(8):1429–1435	
ResMed	261	78 99	4.5.6 Test failure rate (people over 16 years of age) 5.3 Overview of economic evidence in the company submissions	***For the NightOwl device there are no additional costs for re-testing.*** The NightOwl has a built-in battery that is capable of recording up to 10 nights of sleep data. The battery is usable for 3 years from manufacturing. Therefore, the device can be used intermittently and is not required to be used on consecutive days. This provides a large window of time for retesting in case of a failed sleep study or a negative test result that is suspected to be a false negative. The same device that was used for diagnosis can be used for the evaluation of treatment success. Therefore, no device costs apply to the treatment evaluation. In addition, there is evidence that diagnosing over 3 days, would provide for more accurate results and reduce failure rates (Zou et al. 2023). Such an approach could be considered with the NightOwl device. In the Lyne paper which shows a failure rate of 12,1%, the authors state in the discussion *"The trial involved single-night in-laboratory testing, and therefore real-world failure rates could not be assessed. However, we believe our results are generalizable to a health-literate population. Also, given that single-night simultaneous data acquisition was employed, we could not assess for night-to-night variability of OSA or whether the diagnostic classification accuracy of the NOM and NOR improves with multi-night testing".* As the clinical team will receive a flag in the online software in case of failure of a NightOwl sleep study (which could be less than 4 hours recording), the patient can be asked by e-mail or phone to make a second test until there is 1 or several valid test(s). This means that the real-world failure rate is likely lower than what is reported in the Lyne et al paper. **Conclusion:** For the NightOwl device there are no costs for repeat testing regardless of failure rates and rate of treatment evaluation. **References:** 1. Zou D, Vits S, Egea C, Ehrsam-Tosi D, Lavergne F, Azpiazu M and Fietze I (2023) A new approach to streamline obstructive sleep apnea therapy access using peripheral arterial tone-based home sleep test devices. Front. Sleep 2:1256078. doi: 10.3389/frsle.2023.125607	Thank you for this clarification. Following-up this point we asked the company, who confirmed that patients would be expected to retain the device in case of further need in the near future for re-testing to clarify a diagnosis, or for assessing the impact of treatment. Thus, the costs of repeat testing have been updated in Section 5.7.12, and in the calculation of treatment costs where a sleep study is assumed, to reflect that no additional device costs would be incurred. With respect to the issue of testing over 3 nights, we did not find any studies reporting on the accuracy or failure rates over 3 nights. Therefore, our analyses are restricted to the consideration of a single-night sleep study.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Lyne CJ, Hamilton GS, Turton ARE, et al. Validation of a single-use and reusable home sleep apnea test based on peripheral arterial tonometry compared to laboratory polysomnography for the diagnosis of obstructive sleep apnea. J Clin Sleep Med. 2023;19(8):1429–1435	
ResMed	262	104 106-110	5.5.2 The diagnostic pathway: the decision tree 5.6.1 Decision tree	***Diagnostic decision tree structure does not seem to discriminate accurately between moderate and severe OSA*** In the decision tree, sensitivity for mild, moderate and severe OSA is simplified according to the AHI cut-offs of AHI≥5 and AHI≥15 calculated in the parameter sheet. Accordingly, moderate and severe OSA are diagnosed with the same sensitivity input value. In contrast to this, the publication referenced for Home RP states that sensitivity is 0.93 for moderate OSA and 0.63 for severe OSA (Xu et al. 2017). This seems to lead to a substantial overestimation of true positive patients with severe OSA. In the routine care patient pathway this may seem to have a lower relevance as moderate and severe patients are offered the same treatment options. In contrast to this, the lack of accurate discrimination of test sensitivity for moderate and severe OSA respectively in the model has a substantial impact on model outcome as costs and QALYs accrued over the patient/cohort lifetime horizon differ substantially for treated and untreated moderate and severe OSA patients. It needs to be emphasized again that the model is very sensitive to small changes in test accuracy input parameters. Accordingly, this lack of correct discrimination of test sensitivity for moderate and severe OSA introduces bias and uncertainty of unknown direction and size. Further, the input parameters for Home RP used in the model are based on the 4% desaturation rule. There seem to be inconsistencies regarding the 3% vs 4% desaturation rules when looking at the studies used as input for the model. We strongly recommend validating the scoring rules in the diagnostic validation studies, making them transparent in the EAR and comparing/including only those studies with the same scoring rules as these will influence sensitivity and specificity parameters. This is of utmost importance as the model outcome is very sensitive to these input parameters. With the current input data that is based on different scoring rules, different diagnostic markers and that is not referencing accurately the test sensitivity for moderate and severe OSA respectively unnecessary uncertainty is introduced. This seems highly problematic as we do not fully understand the amount of uncertainty that is introduced and whether this may skew the model results in a certain direction. We refer to item 2 of this response for a discussion of impact of model input on outcomes and their relationship/association. **References:** Xu L, Han F, Keenan BT, Kneeland-Szanto E, Yan H, Dong X, Chang Y, Zhao L, Zhang X, Li J, Pack AI, Kuna ST. Validation of the Nox-T3 Portable Monitor for Diagnosis of Obstructive Sleep Apnea in Chinese Adults. J Clin Sleep Med. 2017 May 15;13(5):675-683.	With regard to the parameterisation of the decision tree, please see our response to comment 94, and our scenario analyses where data from 4 x4 contingency tables are used (Table 43 in Section 5.10.2). We agree that the approach used in the base case does have limitations due to the use of sensitivity and specificity to discriminate between OSAHS severity, and the implications this can have – in particular, the company's point here regarding discrimination of individuals with moderate and severe OSAHS, because they do indeed have different utilities. The scenario analysis using the 4x4 contingency tables allows for better discrimination of moderate and severe. As in our response to comment 260, we do make clear in the report that the model is particularly sensitive to the accuracy data used (for the novel devices and respiratory polygraphy). With respect to the second point, we have clarified the desaturation rules used to obtain the accuracy data in the model (please see sections 5.7.3 and 5.7.4).

Section B Economic model - Comments

Stakeholder	Comment	Description of problem	Description of proposed amendment	Result of amended model or expected impact on the result (if applicable)	EAG response
Association of Respiratory Nurses	1	No comments on the economic model. It looks to cover all the relevant aspects,			Thank you.

<table>
<tr>
<td>ZOLL Itamar</td>
<td>2</td>
<td>We believe that the specificity and failure rate values that were used in the economic evaluation of the WatchPAT300 and WatchPAT One, are inaccurate, and thus reduce the economical value of the WatchPAT devices.</td>
<td>We recommend replacing the specificity and failure rate values in the economic calculation with the values we suggested.</td>
<td>
<table>
<tr><td>Device\ICER Vs Oximetry</td><td>Before amendment</td><td>After amendment</td></tr>
<tr><td>WatchPAT 300</td><td>7,893</td><td>7,854</td></tr>
<tr><td>WatchPAT One</td><td>8,132</td><td>8,105</td></tr>
</table>
</td>
<td>Please see our response to comments 25 and 26 in Section A.</td>
</tr>
<tr>
<td>Sunrise</td>
<td>3</td>
<td>We recommend using the 5.88% failure rate (1/17) reported in Alsaif (2022) for the Sunrise device, as detailed in your literature review focusing on a home environment. This is in preference to the rate reported in Kelly (2022). Our suggestion stems from the fact that the data in Alsaif (2022) is more recent, collected in 2022 and 2023, compared to the 2020 data from Kelly (2022).

The rationale for this recommendation is twofold. Firstly, it takes into account the advancements and updates made to the Sunrise device and its user instructions since 2020, including improvements in Bluetooth connection stability.

Secondly, our recommendation aligns with the observations in the report: ”It should be acknowledged that some of the factors contributing to failed tests were not anticipated by the study investigators and, with the benefit of hindsight, were preventable. The expectation is that the learning from these instances will have prompted necessary changes to testing protocols, device features and user instructions to avoid similar failures occurring again. If this is the case then novel device failure rates in clinical practice would be lower than those reported in the studies, all other factors being equal.”

Given these developments, we believe that the failure rate from Alsaif (2022) more accurately reflects the current performance of the Sunrise device.</td>
<td>Sunrise failure rate = 5.88%</td>
<td>ICER vs oximetry = £6,990 (before £7,034)
ICER vs home RP = £39,306 (before £38,402)</td>
<td>Please see our response to comment 27 in Section A.</td>
</tr>
</table>

Sunrise	4	Could you kindly make the following updates for accuracy and consistency? ● Replace 'In base case analyses, we assume that should a sleep study fail, the full cost of a new device would be incurred to undertake a second sleep study.' with 'In base case analyses, we assume that there is no additional sleep study or device cost to the NHS for a failed sleep study with the Sunrise device.' to correct erroneous information. Indeed, in the event of a device failure, the manufacturer replaces the device at no additional cost to the NHS. Thank you for your attention to these details.	Cost of a failed sleep study with Sunrise device: ● Posted = £4.42 + £8.43 + £5.67 + £2.99 = £21.91 ● Collected = £4.42 + £8.43 + £5.67 = £18.92	ICER vs oximetry = £6,960 (before £7,034) ICER vs home RP = £39,915 (before £38,402)	Please see response to comment 37 in Section A.
Sunrise	5	Another potential approach for the base case analysis is to combine data from the Pepin 2020 and Kelly 2022 studies. By aggregating the 2x2 confusion matrices (with the number of patients) for both cut-offs, a more comprehensive assessment of performance can be achieved (even if this combination is primarily influenced by the Pepin 2020 study, due to its significantly larger patient cohort). This method presents the following diagnostic accuracy: ● Sensitivity at low cut-off: 0.907 (95% CI: 0.881 - 0.929) ● Specificity at low cut-off: 0.942 (95% CI: 0.895 - 0.979) ● Sensitivity at high cut-off: 0.925 (95% CI: 0.899 - 0.950) ● Specificity at high cut-off: 0.833 (95% CI: 0.797 - 0.869) This analysis encompasses a total of 407 patients, combining 376 patients from Pepin 2020 and 31 patients from Kelly 2022. The aggregation of data from these two studies not only enhances the statistical power through a larger sample size but also provides a broader	Sunrise diagnostic accuracy: Sensitivity at low cut-off: 0.907 Specificity at low cut-off: 0.942 Sensitivity at high cut-off: 0.925 Specificity at high cut-off: 0.833	ICER vs oximetry = £7,028 (before £7,034) ICER vs home RP = £59,784 (before £38,402)	Please see response to comment 35 in Section A.

		perspective on the diagnostic performance across different patient populations.			
Sunrise	6	Issues 1 + 2 + 3	Sunrise failure rate = 5.88% Cost of a failed sleep study with Sunrise device: ● Posted = £4.42 + £8.43 + £5.67 + £2.99 = £21.91 ● Collected = £4.42 + £8.43 + £5.67 = £18.92 Sunrise diagnostic accuracy: ● Sensitivity at low cut-off: 0.907 ● Specificity at low cut-off: 0.942 ● Sensitivity at high cut-off: 0.925 ● Specificity at high cut-off: 0.833	ICER vs oximetry = £6,944 (before £7,034) ICER vs home RP = £62,715 (before £38,402)	Please see responses to comments 27, 35 and 37 in in Section A.
Sunrise	7	The model assumes, on the basis of expert's opinion, that time to diagnosis and to treatment are identical from one device to another (3 months, decreased at 1.5 months in a one-way sensitivity analysis). Using the preliminary results from the Sunrise SOSAT study, we conducted a similar analysis by substituting 3 weeks for the previously mentioned 3 months time to diagnosis. However, this adjustment did not alter Sunrise QALYs. Yet, we assume that prompt treatment initiation should theoretically influence factors such as utility so this result is not so intuitive to us: is there an explanation about the lack of influence of time to treatment on the QALYs and ICER? Concerning the utilities, one-way sensitivity analyses are carried out on this parameter (Table 69 and Figures 12-17) but only slight information and explanations are reported in section 5.10.3 of the report and the chapter about uncertainties.	Should section 5.10.3 of the report (focusing on the results of the one-way sensitivity analyses) provide more information and discussion about the lack of influence of some parameters on cost-effectiveness results, including utilities? Should it also be discussed in the chapter about uncertainties?		Thank you for highlighting this point. Investigation led us to identify an error in the model, where time to diagnosis and treatment parameters for all novel devices was set to that for oximetry. This did not impact on the base case analysis (as all strategies were assumed to be the same), however, it meant that the one-way sensitivity analyses were not correct. This has now been corrected. After this correction, we find that changes in assumed times to diagnosis and treatment for novel devices have a small impact on cost-effectiveness results (see our response to comment 52 above). We also report an additional scenario analysis with a reduced time to treatment for novel devices (see EAR Section 5.10.2). This scenario did not change the cost-effectiveness conclusions.

Acurable Limited	8	General Comments Please note, these problems are also identified and discussed in Section A above. These have been included in the below format as well as they relate directly to the economic model. While we note that "Comments on the model must be given in sufficient detail to enable your changes to be replicated by the model owners." and that “where the EAG are unable to replicate the results claimed by implementing the changes described, comments will be rejected without further consideration”, we still believe that it is essential to take these into account, as both Committee and EAG may well have the knowledge to implement the changes and therefore review impact.			
Acurable Limited	9	Test failures are calculated incorrectly. This is explained in detail in comments in Section A.	**The EAG has considered that the test failure rate for AcuPebble is 6.74%, when this should be 0.5%.**	The model has not been re-run due to time constraints.	Please see our response to comment 128 in section A.
Acurable Limited	10	In the “parameters” tab of the economic model, the specificity and sensitivity of some devices are stated to be given with respect to ODI threshold, when in the EAR they correspond to AHI. This may be a typo but please correct this.	Use AHI thresholds for consistency for all devices, or correct if this is a typo.	This is assumed to be a typo, but should be corrected. If it is not, then the same indexes should be used for all devices otherwise the model is inconsistent.	Thank you, this has been updated in the model.
Acurable Limited	11	Sensitivity and specificity of AcuPebble is incorrect, as described in detail in Section A. The incorrect values as quoted from the model (which we also believe should say “AHI” not “ODI”, as per comment above) are: Sensitivity of Acupebble SA100 at low threshold (ODI ≥ 5): 0.86 Specificity of Acupebble SA100 at low threshold (ODI ≥ 5): 0.73 Sensitivity of Acupebble SA100 at high threshold (ODI ≥ 15): 0.82 Specificity of Acupebble SA100 at high threshold (ODI ≥ 15): 0.89	The values which should be used, from the performance evaluation of AcuPebble SA100 compared to PSG, using AHI are the following **at high threshold (AHI≥15) :** Sensitivity: 92.86% (CI: 76.50% to 99.12%) Specificity 97.14% (CI: 85.08% to 99.93%) [redacted]	The model has not been re-run due to time constraints.	Thank you for providing these data. Please see our response to comment 138 in Section A.

			Alternatively, since the performance does not significantly degrade from PSG to PG, the pooled performance data taken into account both studies could be used: **At high threshold (AHI≥15) :** Sensitivity 93.83% (CI: 86.18% to 97.97%) Specificity 96.21% (CI: 91.38% to 98.76%) [redacted]		
Acurable Limited	12	The assumption that a failure of the equipment will incur additional cost per test is not correct for AcuPebble SA100. Any test caused by the fault of the system is provided for free. So, the corresponding input for this needs to be changed in the health economics model. There is no charge for a test where a diagnosis cannot be made due to an issue with the tech. This is explained extensively in Section A.	The cost of an AcuPebble repeat (due to failure to get dx) should be £0. This could be amended in [Costs - Tests, C34] in the model, if two-night studies are never considered in the model. Note that if two-night studies to account for night-to-night variability are considered, then £14 is the cost provided by the manufacturer for the second night of a study.	Model has not been rerun.	Please see our response to comment 174 in Section A, and Table 31 in the Report detailing the costs associated with repeat sleep studies due to failures or misdiagnosis.
Acurable Limited	13	The time it will take to receive a diagnosis has been assumed to be three months for all technologies, on the basis that more evidence could not be found. This biases the outputs of the model by assuming a similar time as with RP or PSG, when in reality in many cases clinicians have the results of the tests from novel technologies immediately after the patients do the test. Note that this cannot be generalised to all technologies, since some of them require patients to return the device and upload the signals to see the results, others require semi-manual marking, and others have	The model should consider the “real case scenario” enabled by the novel technologies to show the potential benefits if the pathways were changed, considering a “time discount” with respect to RP, depending on their functionality and what they have received regulatory approvals for. In some cases, this might be reduced to 1 month, or 0.5 months. For AcuPebble SA100, where the diagnosis is automated, this is immediately available, therefore the time to diagnosis is 0 months.	Model has not been rerun, but this change is expected to portray a much more realistic view of outcomes when using novel technologies, as one of the main benefits for some of them is the reduced reporting time and therefore ability to clear waiting lists and accelerate time to diagnosis.	Please see our response to comment 52 in section A.

		regulatory approvals for fully automatic diagnosis. Hence, time has to be "discounted" in reference to RP accordingly (i.e. a technology for which data uploading and manual interpretation is required will be equivalent to whatever time is assumed for RP, whereas in the other extreme technologies for which automatic diagnosis has been approved by regulators enable instantaneous diagnosis (which for practical reasons could be assigned a minimum time).			
Acurable Limited	14	The failure rate of PSG is assumed to be 0%, because "we assume this sleep study is perfect". PSG can be assumed to be perfect in terms of its gold-standard diagnostic performance, but that does not include its failure rate. There is no evidence that PSG failure rate has been researched in this report. Adjusting PSG failure rates (if they do differ from 0% in actuality) would lead to more representative ICER values, and qualitatively better represent devices against the gold-standard.	Evidence of the PSG failure rate, preferably measured within healthcare systems representative of the UK, should be supplied. If the value is not 0, then a similar rule as RP and HSATs could be used (i.e., assume at most 2 tests) which would then be reflected in extra manual analysis time and other associated costs. Furthermore, clinic based tests require significant scheduling, which may affect the ability of a patient attending a sleep clinic to receive a prompt repeat test.	Model has not been rerun. Adjusting PSG failure rates (if they do differ from 0% in actuality) would lead to more representative ICER values, and qualitatively better represent devices against the gold-standard.	As with the simplifying assumption in the model that PSG is perfect, we have also assumed that the failure rate is perfect, though acknowledge that this may not be the case. We have further highlighted this simplification in the report (see section 5.7.5). The use of PSG is only assumed in a scenario analysis. Due to a very low proportion of the cohort with moderate-severe OSAHS being misdiagnosed as having no OSAHS after two sleep studies, there is no substantial impact on the model results when use of PSG as a third sleep study is assumed.
Acurable Limited	15	The failure rate of RP is stated as 5%, which seems to be well below what has been reported by many other sources.	In a journal publication (Devani et al), it was reported that 16/182 RP tests failed in the, and the reasons to consider failure are also given. That failure rate is 8.8%, as opposed to the 5% quoted. There are other published sources that can be used reporting failure rates of up to 20%. A failure rate of RP somewhere between 8.8% and 20% should be used in the models (note that in Devani et al, that reported 8.8% patients, although consecutive, had been extensively training on the use of RP because they were part of a trial in which RP was going to be the reference test and they would have no option of repeating it (as per protocol). Hence it is likely that in the real world situation the number can be as high as 20% as reported by others.	Model has not been rerun. This impacts how all the novel devices compare to RP.	Please see our response to comment 162 in Section A.
Acurable Limited	16	Sensitivity and specificity are statistically obsolete metrics to evaluate diagnostic methods. They provide some value but can be misleading because of the way they are defined (further scientific evidence supporting	In addition to outputs with sensitivity and specificity, the model output should also be generated using predictive values or likelihood ratios as input parameters for	The model has not been re-run due to time constraints in producing these comments. The expected impact is that the model would be more robust, with significantly less uncertainty over the performance of diagnostic devices.	Please see our response to comment 42 in section A.

		this statement can be provided). On the basis of this the model should have used either predictive values or likelihood ratios.	each technology. And recommendations should be based on likelihood ratios.		
Acurable Limited	17	The quantifiable consequences of OSA go well beyond cardiovascular events (see **Section A, point 105 for more detail**). This might not have been the case a few years ago, but the evidence now is overwhelming. Focusing this study only on cardiovascular events and road car accidents is not only scientifically lacking but also misleading for the public, and biasing the outputs of the model.	Other dis include increased risk of developing type 2 Diabetes, depression, or loss of productivity at work or school. Each has distinct effects on HRQoL and mortality.	Model has not been rerun. The cost of false negatives would be much greater, affecting devices' ability to meet WTP thresholds.	Please see our response to comment 167 in Section A.
Acurable Limited	18	No reference to benefit of multi-night testing, which is not effectively possible with RP but is with many of the novel devices.	Model should include a scenario analysis of how multi-night testing can be enabled and may well address many of the limitations mentioned.	Expected impact is both on failure rate of test, and also impact of having two nights rather than one in terms of potential misdiagnosis of a patient (i.e., a single test can be used if the second is invalid, and the second can be used to improve the diagnosis if both are valid).	We did not find any studies reporting on the accuracy or failure rates for multi-night sleep studies. Therefore, our analyses are restricted to the consideration of a single-night sleep study.
Acurable Limited	19	Whilst this is not a direct outcome of the analysis, the environmental cost (of production, delivery, analysis, upkeep) of the novel and established devices could have been analysed. Some manufacturers may choose to increase spending on more environmentally friendly materials, or provide a solution that is inherently less environmentally destructive.	This could be captured through carbon pricing methods, taking into account the materials, number of units used (and disposed of) per study, shipping emissions and computational costs, and then related to the failure rates of studies in much the same way as the existing costing model.	The effect of this would be to increase costs across the board, but differentiate between devices such as RP and reusable novel devices, and other disposable devices.	This is beyond the scope of work for the EAG report. Moreover, how best to incorporate these environmental costs are part of on-going work in the research community
Acurable Limited	20	The time to review an AcuPebble SA100 report has been incorrectly and arbitrarily assumed to be 20 minutes. Reporting time varies across novel devices depending on whether regulatory approval is for automated diagnosis or not. The same estimate cannot be applied to all equally. AcuPebble SA100 is the only one with regulatory approval for fully automatic diagnosis. Hence it cannot be assumed to have the same time for review of the diagnostic report as others that do require a much more detailed level of interpretation.	For AcuPebble SA100, the company provided a range of 5-10min which should be used instead.	Model has not been rerun. Expected outcome is for the model to be more representative of the benefits provided by automated diagnosis.	Please see our response to comment 175 in section A
Acurable Limited	21	The AcuPebble "repeat test required" notification functionality cannot be switched off. Evidence as a result of the contrary is factually incorrect. If a patient sleeps less than 4 hours, they will always automatically be asked to repeat the study by the AcuPebble SA100 system, via the app, without requiring any involvement of the clinician, and without needing to return to the clinic.	Any analysis that indicates it is possible to "switch this feature off" (such as the parameter "AcuPebble notifies pt of failure") should be removed as this is always the case.	Model has not been rerun. Expected outcome would be to remove inaccuracy in terms of what happens when there is a user-related test failure	Please see our response to comment 174 in section A.
Acurable Limited	22	The economic model has been run with models of devices that are not approved by regulatory authorities in the UK.	At least one of the models of NightOwl (as said by the EAG) has not been approved to be sold in the UK. Please remove all evidence and assumptions based on this	Model should be rerun with only assumptions corresponding to existing commercial systems.	With regard to CE marking and regulatory approval please see our response to comment 100 (Acurable).

			model from the economic study. Taking into account the current situation with waiting lists for Notified Bodies the process of approval might take 1.5-2 years, at which point this whole report might be obsolete.		
Acurable Limited	23	The sensitivity and specificity of Sunrise used in the economic model have been obtained with post-hoc thresholds, which are different from the clinically established ones.	Remove from the model any values of sensitivity and specificity that have been obtained with post-hoc adjustment of thresholds. If this is not possible because all the values have been obtained like that, there is at least one publication where, although post-hoc, the values for sensitivity and specificity could be extracted using the 5, 15, 30 thresholds. Use those ones instead.	Please re-run the model.	Please see response to comment 198 in section A.
Acurable Limited	24	It is uncertain whether all the systems comply by design with the Information Governance requirements of the NHS. Without that, manufacturers providing fulfilment services is not realistic	Check with manufacturers that they have been approved in terms of information governance by at least one Trust. If they have not, the assumption of those particular manufacturers posting systems to patients needs to be reconsidered. This will affect the values input into the model for some devices.	Re-run the model accordingly depending on what manufacturers say.	Assessing compliance with data protection regulations is outside the scope of the EAGs work. Note that recent NICE guidance has stated the need for technologies to follow NHSE's Digital Technology Assessment Criteria (DTAC), which includes data protection.
Acurable Limited	25	Costs of having to hold stock of devices.	While this cost might be negligible for some systems. For those that are not reusable (Sunrise?) and for which manufacturers have not been granted permission to send devices to patients, high volume healthcare centres will need to store a large number of stock. This stock will incur a state cost which will be non-negligible (once boxes, accessories, etc are taken into account). Should any of these assumptions be correct this needs to be taken into account in the model.	Re-run the model	Please see response to comment 177 in Section A.
Acurable Limited	26	The cost for traffic accidents includes the police costs but does not include the overall costs to the economy, which are also given by the Department of transport and are significantly higher than that. Also, how many of these accidents lead to disabilities and are the impacts of those disabilities incorporated in the model in any way? These have not been discussed, nor has a methodology been proposed for considering only a subset of state costs in the scenario analysis, instead of extended costs such as lost productivity and the burden on affected parties.	Neglecting these costs should be either properly justified or alternatively properly modelled for the model.		Please see response to comment 184 in Section A.
Acurable Limited	27	The determination of WatchPAT systems' sensitivity and specificity has excluded significant evidence, such as that provided by a landmark evaluation of the WatchPAT 200 in 500 people with varying skin tone	The conflicting evidence from a range of other studies on WatchPAT should be taken into account, not least the results included in Ioachimescu. Considering the demonstrable effects of skin colour on PPG	This would require a full re-evaluation of WatchPAT performance if included. As such we have not rerun the model. The total expected effect is not clear, for instance as Ioachimescu	With regard to Ioachimescu 2020 please see our response to comment 26 and comment 253

		(Ioachimescu et al., which was used only for failure rates). A further meta-analysis of 17 studies was not acknowledged (Iftikhar et al., 2022, J Clin Sleep Med). Whilst included evidence reports low specificities, Ioachimescu reports only 65% sensitivity for moderate-severe (with 91% specificity). On the other extreme the EAG has "mixed and matched" evidence results from different WatchPAT models. There needs to be a critical assessment as to why some results are representative and others are not. And if not either the results are "pooled" and sensitivity and specificity are calculated together with confidence intervals, or the most extreme values have to be computed.	and PAT based systems, a scenario analysis should be conducted that adjusts results based on skin tone.	observed a higher specificity at moderate-severe.	
Acurable Limited	28	In addition to adjusting the diagnosis time down from 3 months for all novel devices, the method of data transport and analysis should be taken into account. HSAT devices that store data locally and rely on its safe return and processing incur subsequent personnel time costs for interfacing with the device and uploading its data, in addition to the further delay waiting for analysis. This difference in time has not been acknowledged in the model. Leaving the analysis as it groups all novel HSATs in this regard, despite the clear implications of disposability, remote data upload, and local storage.	The model for time allocation should reflect the individual data uploading and handling processes for the devices. For instance, some devices may require no data handling time (effectively instantaneous reports after the patient has uploaded) and others may require 5-10 minutes (or referral to manufacturer estimated times) to be factored in.	The model has not been rerun.	These costs have been accounted for in the model and do differ depending on the novel device. Please see section 5.7.12 of the report, and Table 30.
Acurable Limited	29	We posit that the costs of a low specificity test are not sufficiently accounted for in the model. The primary cost of a false positive in the current model is the subsequent cost of care and treatment, which is capped at a year. This is relatively minimal in comparison to the cost of untreated sleep apnoea, due to RTAs, QoL reduction, and cardiovascular attacks (in fact we believe it should be even higher, after taking into account other likely diseases and costs, as mentioned in comment 9). As the available interventions for OSAHS are not potentially harmful or that damaging to QoL, it may be tempting to accept low specificity, and overdiagnose and overtreat, especially in light of the current underdiagnosis of OSAHS. And yet, the cost of a false positive is to degrade the usefulness of all positive results in the eyes of both patients and doctors. Patients may require confirmatory testing (e.g., in-lab PSG) should their symptoms continue. A misdiagnosis could lead to the misattribution of symptoms of other underlying health problems	The EAG should follow up on and model downstream effects of false positives, as well as consider the effects of excessive false positives on trust if they cannot be modelled economically.	Model has not been rerun. This would likely lead to a higher penalisation of false positives, and a more accurate representation of the social costs of systems that are more suited to screening rather than primary diagnosis tools.	We have undertaken an additional analysis assuming treatment costs continue for two years for false positives. Please see Section 5.10.2. We also highlight the error that was found in the model, which relates to this point, and was incorrectly benefitting devices with low specificity at the low diagnostic cut-off. Please see our response to comment 42.

		to sleep apnoea. In such a case, a patient would be denied the proper care for an extended period of time. Given these risks, it is not unreasonable to expect clinicians to adopt complementary tests to increase confidence. As such the costs could include secondary testing (much as with false negatives), as well as the costs due to lack of treatment for other diseases in the differential diagnosis.			
ResMed	30	Tab “Costs (tests)” cells B84, B85, B91, B95, B96, C95, C96, F95, F96, G95, G96, J95, J96 The model states re-test costs of 0-90 GBP for NightOwl. As stated in section A item 3 above, the NightOwl battery is usable for three years from manufacturing. Therefore, no additional costs for re-testing apply for the NightOwl device. Since there is a notification in case of a failed sleep study there is no need to review data in such cases. Therefore, reduced personnel costs apply for re-testing with NightOwl due to a device or data acquisition failure. Furthermore, no costs for postage apply to NightOwl as devices are sent out from ResMed to the patient directly at no additional charge.	We propose to amend the model to reflect no additional costs for re-testing with the NightOwl device.	The amendment results in slightly lower costs for NightOwl that do not change the model outcome substantially.	Please see our response to comment 261.
ResMed	31	Tab “Parameters” cells E19-22 and Q19-AF22 The study used as input for Home RP sensitivity and specificity enrolled only 80 patients, which is low for a diagnostic validation study. Furthermore, the study enrolled a Chinese population with unknown applicability to the English NHS setting.	We propose to use the data synthesis from the last NICE OSAS guideline update as input data for Home RP or choose another more robust study with a patient population that is comparable to the patient population in the NHS.	See Section A item 2	Please see our response to comment 260 in section A.
ResMed	32	Tab “Parameters” cells E31-34 and Q31-AF34 The study used as input for NightOwl sensitivity and specificity enrolled only 94 patients, which is low for a diagnostic validation study. Furthermore, the study did not rely on double scoring to minimize interscorer variability.	We propose to synthesis data from the published NightOwl studies or choose another more robust study with higher patient numbers and double scoring of PSG data (Massie et al 2018 or van Pee 2022).	See Section A item 2	Please see our response to comment 260 in Section A

ResMed	33	Tab “1st year costs and utilities” cell L9 Costs for evaluating treatment success are included by multiplying cells C9 and K9. Therefore, even if re-test costs are set to zero in the “Costs (tests)” tab the full costs are still considered for treatment evaluation.	The formula on cell L9 should reference the re-test costs in the “Costs (tests)” tab.	No major impact on model results is expected.	Thank you for pointing this out, it has been updated
ResMed	34	Tab “Decision Tree” cell I78 In the decision tree severe OSA is diagnosed with a sensitivity that is similar to the sensitivity of the AHI≥15 threshold. In contrast to this the referenced publication states a sensitivity of 0.63 for severe OSA. This seems to lead to substantial overestimation of true positive severe OSA cases for Home RP. The same oversimplification holds true for the other diagnostic devices. For NightOwl the sensitivity for moderate and severe OSA is similar. Therefore, the model simplification has no substantial impact on the NightOwl pathway. We are not aware whether there is a substsantial impact for the other innovative devices.	The sensitivity at different thresholds should be reflected accurately.	The proposed change is anticipated to have an impact on ICER for all devices.	Please see our response to comment 94, and our scenario analyses where data from 4 x4 contingency tables are used.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Section A: External Assessment Report - Comments :

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Association of Respiratory Nurses (ARNS)-Kerri Smith	1	-	-	I am emailing to confirm that ARNS do not have any further comments to submit regarding the Novel home-testing devices for sleep apnoea/hypopnoea syndrome: economic model.	Noted, thank you.
Sleep Apnoea Trust Association (SATA)-Ann Nevinson	2	-	Overall / general Based on comments document and track changes report	These comments build on and from those we supplied based on the earlier version of the report sent out for consultation and the EAG responses document plus the track change current version of the systematic review and evaluation report. We assume that to aid committee discussion members will see or have seen all versions for completeness . In particular we note that given these are described as " home" testing devices evidence of their being tested in home as opposed to clinic settings in a context likely to be representative of the population of the UK has not been found . We also note the comments which seem to indicate that patient experience has not been featured in the modelling work but even if it had been included it is felt this would not have impacted on the results.	As stated in the Response to comment 4 in the first round of comments: "Any potential for greater acceptability of the novel devices over the comparators is not directly captured in the model. As the sleep studies occur over just a few nights at most, any attempts to incorporate acceptability in terms of QALYs would have a negligible impact on the model results. Acceptability may be captured indirectly via the failure rates for the novel devices (should they be lower than those for the comparators). Given the difficulties and likely small impacts on the model results of trying to capture these potential benefits, such claims are best dealt with as part of the deliberative process and

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					depend on the evidence for greater acceptability over the comparators."
Sleep Apnoea Trust Association (SATA)-Ann Nevinson	3	6	Track changes (TC) Limitations and Conclusions	In the limitations section you do state that there is " high uncertainty " over the cost effectiveness results (from the model ?) and acknowledge the reliance on clinic data. In moving to the conclusions section you use the word " estimate " in describing that novel devices are a cost-effective alternative to oximetry . It is understood NICE often uses particular words to classify the strength of its conclusions but we are unclear if that is the case in the use of "estimate" which may merit discussion or description in a terms section to clarify . As one reads further into the report the uncertainties , variations and limitations are clearly described it seems important that those who may simply read what is an opening section of a report and take decisions cant leap to conclusions without understanding the findings in context	We have clarified in the report that we are referring to high uncertainty over the cost-effectiveness results from the model for the comparison of novel devices with respiratory polygraphy. With respect to their comparison with oximetry, there is little uncertainty in the model results. Please note that the use of the word "estimate" is solely based on the fact that our cost-effectiveness analysis results are estimates (based on the inputs and model structure).
Sleep Apnoea Trust Association (SATA)-Ann Nevinson	4	12	TC Conclusions Implications for service Provision	This section is clear in addressing the above point and the final bullet point which emphasises the high uncertainty over the relative diagnostic accuracy estimates for all devices advises caution in interpreting the results	Noted, thank you.
Sleep Apnoea Trust Association	5	13	TC research priorities	Agreed.	Noted, thank you.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
(SATA)-Ann Nevinson					
Sleep Apnoea Trust Association (SATA)-Ann Nevinson	6	237	TC discussion	There is an " expectation " of learning to prompt changes described here which is important not simply to improve failure rates but to improve the results and experience for patients	Noted, thank you.
Sleep Apnoea Trust Association (SATA)-Ann Nevinson	7	238 - 243	TC Section 6.2 & 6.3	The changes made in this version do set out the limitations and uncertainties linking with comments made to earlier consultation and above	Noted, thank you.
Sleep Apnoea Trust Association (SATA)-Ann Nevinson	8		General	To conclude we again would repeat as we did in our previous comments. We note the extensive work that has been undertaken in carrying out this assessment in particular the effort made in trying to consider available evidence and make adjustments to draw reasonable comparisons . We trust that there will be many points of learning in addition to the research recommendations that will to enable the work and evidence on the devices to be further developed to overcome the current variables and uncertainties identified. SATA very much wish to see the development of improvements in NHS sleep services to enable effective accurate diagnosis of OSAHS for patients .	Noted, thank you.
Zoll Itamar-Lior Solomon	9	122	5.7.4	Diagnostic accuracy of RP: The analysis quotes two validation studies of RP vs PSG, however, most patients in these studies did not perform simultaneous measurement of RP and PSG, which is standard practice in validation of measurements. The results of these studies showed high sensitivity and specificity of the RP compared to the PSG, but due to the lack of simultaneous	*No simultaneous measurement in RP diagnostic studies* Thank you for highlighting these limitations of the studies informing the diagnostic accuracy of RP. We have

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				measurement these results represent a lower quality comparison. The assessment doesn't take into account that RP devices usually do not measure sleep time (which also seems to be the case in the two validation studies that were used as reference). To emphasize this point, we would like to refer you to Escourrou et al 2015, a large-scale multi-centre study that assessed the results of 5,745 RP (AKA PG) tests and 5304 PSG tests. The study showed that the lack of sleep time measurement in RP led to a lower AHI score compared to the PSG tests. The study found that "Overall, patients investigated using PG are likely to have a 30% lower AHI on average, compared to patients investigated by PSG", and concluded that "PG interpreted using standard guidelines results in underdiagnosis and misclassification of OSA". The lack of sleep time measurement (which serves as the denominator in formula of the AHI calculation) represents a significant disadvantage of the RP and oximetry devices compared to most of the Novel devices that are assessed in this report. We also wish to address an issue that bears clinical significancy, yet wasn't properly addressed in the assessment – Sleep Apnoea phenotyping: In recent years, sleep medicine started looking beyond the all-night average AHI, referring to different physiological elements that could impact the disease severity, but may not be reflected in the all night average AHI. Addressing these elements in the diagnostic process could impact the diagnostic conclusion and the therapeutic pathway. These elements include: 1. Differentiation between obstructive or central sleep apnoea 2. Detection of REM related sleep apnoea 3. Detection of positional related sleep apnoea. Patients with Central Sleep Apnoea may not benefit from standard CPAP therapy, patients with REM related sleep apnoea may need therapy even if their all-night average AHI is mild, and patients with positional sleep apnoea might be better off with positional therapy, instead of CPAP. Being	amended the report to make clear that not having simultaneous measurement is a limitation of both studies. (Please see Section 5.7.4) *Lack of sleep time measurement in RP* Thank you for the summary of Escourrou et al 2015. We agree with the point made about total sleep time (PSG) vs total hours recording (RP). We have added a sentence to section 5.7.4 which highlights this limitation to the report. *Sleep phenotyping* Penotyping in OSA was not specifically mentioned in the NICE Scope for this diagnostic assessment, and it was rarely mentioned in the studies included in the systematic review. Whilst sleep phenotyping is an advancing area of interest in the study of OSA it and has potential to inform treatment pathways, we are unable to make an specific conclusions and recommendations based on this report.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				able to phenotype sleep apnoea has an impact on the therapeutic pathway, and thus we believe it should have been addressed in the assessment. While WatchPAT can detect all the elements mentioned above, the other novel devices are unable to detect some or even all these phenotyping parameters, and, as we mentioned before, most RP cannot detect sleep time and sleep stages, and some can't even detect body position. Pulse Ox devices are unable to detect any of these parameters. In addition, we wish to update that last year WatchPAT devices received FDA clearance for flagging Atrial Fibrillation, and we expect CE approval for this new feature to be granted during 2024. This new capability could assist in detection of asymptomatic arrhythmia patients during sleep studies, and potentially direct them to a cardiologist for full diagnosis.	
Zoll Itamar-Lior Solomon	10	269	Appendix 2 Table 56	Ioachimescu et al. 2020: We couldn't not notice the repeated attempts by Accupebble to draw the attention to the Ioachimescu et al study, probably with the intention to tarnish the value of WatchPAT (which is the subject of this validation study). Therefore, although the authors stack with their decision to exclude this study, we would like to address it in our comments. When the Ioachimescu study was published we immediately noticed that key data in the study is questionable. A quick glance at the table of the PSG results, indicate that the ODI values are so low that the integrity of the study can be easily challenged. It is visible when comparing the median AHI 3% (18.4) and the median ODI 3% (2.5). For the sake of argument, if we imagine that the median is a specific person, that person had on average 18.4 respiratory events but only in 2.5 of these they also desaturated at 3% or more. That scenario is highly unlikely in patients' population of older obese men, with high risk for sleep apnoea. Furthermore, the measurement of oximetry in WatchPAT and in the PSG is performed with the same technological method, and thus one should expect the ODI results of both devices to be rather similar. However, the median ODI 4% of the PSG is 0.5, while in the WatchPAT it is 10.9, more than twenty times higher. When we drew the attention of Prof Nancy Collop (a co-author of this	Noted, thank you.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				study and the chief editor of the JCSM, where the study was published) to the problematic data in the study, she immediately invited us (the company) to write a letter to the editor, and indeed, a letter to the editor, signed by Itamar's medical director, was published by the JCSM, titled – "Oximetry data affects quality of gold standard" Moreover, the questionable PSG oximetry results were brought to the attention of members of the NICE committee that was in the process of writing the NICE Sleep Diagnostic Guidelines, and as a result, the Ioachimescu et al study was excluded from the NICE guidelines' references. Finally – both Atlanta VA and Emory university are still using WatchPAT devices as their primary RP to this day.	
Zoll Itamar-Lior Solomon	11	32	1.3.1 Accupebble SA100	In the description of the accupebble device it is stated that there is a possibility to add to the device a compatible third-party oximetry. In other words, the device itself doesn't have oximetry measurement as a default. In the NHS, different sleep centres use different desaturation thresholds for AHI calculation – some use 3% while others prefer 4%. For example, in Royal Free NHS trust, where the Devani et al study was performed, sleep apnoea diagnosis is based on 4% desaturation threshold, while in the validation study, a 3% desaturation threshold was applied. We estimate that this dual approach could lead potential users of the Accupebble to request the usage of oximetry. We are assuming that the device is calibrated only to one of the two thresholds, since the validation studies provided by the company referred only to 3% threshold. Therefore, it is not clear if the change in threshold requires oximetry measurement. We assume that it does. We believe that the cost per test provided by the company doesn't include the cost of a compatible oximetry device and the time invested in analysing it. If this is true, we request that these costs will be included in the financial assessment, for a more balanced comparison.	The publication by Devani et al., 2021 clearly states that automatic output of the AcuPebble device is: • Diagnosis based on AHI defined by the current recommended AASM criteria (ie, with ≥3% as the threshold for oxygen desaturation). • Diagnosis based on AHI defined by the AASM criteria, but with the exception of having ≥4% as the threshold for desaturation. • Diagnosis based on ODI considering ≥3% desaturation as the threshold for events.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					• Diagnosis based on ODI considering ≥4% desaturation as the threshold for events This is also reiterated in information on AcuPebble provided by Acurable to NICE.
Zoll Itamar-Lior Solomon	12	76-77	Table 9 Accuracy of Novel devices	We noticed that in all of Sunrise's studies (Pepin et al, Kelly et al) Dr. Jean Benoit Martinot was amongst the authors. It appears that he is the father of the two founders, and current leaders of Sunrise. Furthermore, we noticed that in Pepin et al 2020, four of the authors were employees of Sunrise. We believe that this is a clear conflict of interest that could jeopardise the conclusions of this assessment, if it is based only on these studies. We suspect that this argument is in line with the already existing identification of these studies as high risk of bias. The same argument applies to Messie et al and Van Pee et al, as both were the founders and leaders of Ectosense (the company that developed the NightOwl device and was later acquired by ResMed), at the time of the publication of these studies. We request that the evaluation of the devices accuracy will be based exclusively an independent studies that are cleared of any conflict of interest.	Author conflict of interest statements were provided in cited material for all included studies in this assessment apart from two studies, encompassing all the novel devices. The EAG considers conflict of interest a potential risk that applies to the evidence for all devices included in this assessment. It is not uncommon for authorship of scientific publications to include representatives from the technology sponsor. The purpose of our through and independent critical appraisal of the studies is to identify any risks of bias in these studies.
Zoll Itamar-Lior Solomon	13	151	5.7.12	If a second test with WatchPAT One is needed, the assessment calculated "15 minutes for a band 4 to prepare the device and organise re-sending, or collection, of the device for the repeat sleep study".	The company are correct that we assume 5 minutes for "quick registration of the device" before it is sent to patients.

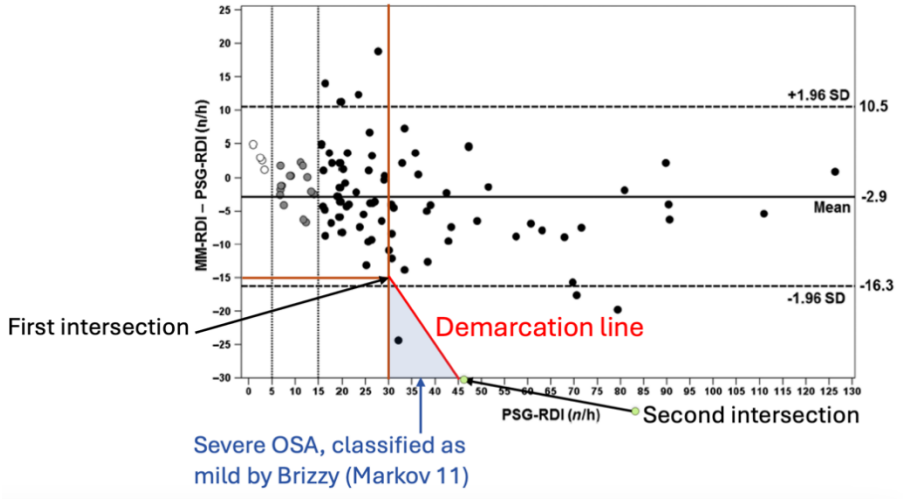
DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				However, since it is a disposable device, the only task left for the band 4 is to register a new device and send it, which you calculated earlier as a 5 minutes task.	The 15 minutes for a band 4 member of staff highlighted here includes 5 minutes for registering the device for a repeat sleep study, and time for organising postage of the new device. We therefore believe this is a reasonable assessment of the resources needed to prepare and send a second device when a repeat test is needed.
Nomics S.A-Nicolas Stefenatto	14	114-	5.6.2	**Concerning the modelling of cohort 11 – True severe Conservative management** **Issue 1.** The simplifying assumption that true severe cases of OSAHS (33% of the starting cohort) when misdiagnosed by Brizzy (11%) are considered mild (93%) or without OSAHS (7%) is not supported by diagnostic performance evidence. The EAG base case estimates 34 cases of this type in a starting population of 1,000 adults with suspected OSAHS I.e., qualifying for cohort 11. Due to limited evidence, the authors had to design the decision tree so that test results of moderate-severe OSA were grouped together rather than being separated into moderate and severe OSA. For Brizzy, at least, this means that in the model there is an overestimation of the number of true severe OSA cases that would receive a 'mild' OSA test results (cohort 11). Indeed, most cases of true severe OSA that are not correctly diagnosed would be diagnosed as 'moderate' OSA. These cases are underestimated in the model because of the combined grouping of moderate and severe OSA.	As the company states, in the base case analysis our decision tree model is a simplification of the OSAHS classification of individuals. We do not have access to the raw data from the Martinot 2017 study. We have tried to extract these data from Figures in the paper – Fig 2 and Supplementary Fig 4. However, due to overlapping data points, we have been unable to do this such that we can replicate the sensitivities and specificities reported in Martinot 2017. We believe that using any extracted

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				The article already used by the EAG for Brizzy (DAR ref 21, Martinot et al. 2017, Respirology) provides some evidence that the model overestimates the number of patients in cohort 11: 1. In the article, Figure 2 presented a Bland–Altman plot of the concordance between the Brizzy-calculated index and the polysomnography index. Using this plot, a line can be traced to represent the demarcation between patients with severe OSA who will be classified as at least moderate (above the line) and patients with severe OSA who will be classified as no or Mild OSA (i.e., the line were the Brizzy index is lower than 15n/h). This line starts at the intersection of the X-axis at 30 n/h and the Y-axis at -15n/h, and passes at the intersection of 45n/h on the X-axis, and -30n/h on the Y-axis (see attached screenshot for clarity). This demarcation line demonstrates that most severe OSA patients will be classified as having moderate if not severe OSA.	data would introduce yet more uncertainty. We therefore acknowledge that the simplification of the parameterisation of the decision tree is a limitation, and may well lead to the performance of Brizzy being underestimated. Please note that the simplification of the decision tree is discussed in Section 5.6.1 of the EAR, and that a scenario analysis for devices where we do have the 4x4 contingency table data improves the cost-effectiveness for some novel devices, [redacted]

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				 2. In the supplementary information, Table S2 "Differences between the two scoring methods (MM analysis and conventional PSG)" shows that the index calculated by the PSG has a very low delta compared to the index calculated by Brizzy, with the 95% CI boundaries being -4.20 and -1.49. However, this CI is based on the entire sample, not just true Severe OSA. 3. In the supplementary information, Table S1 "PSG-RDI and MM-RDI scores by categories of OSAHS severity", shows that the mean "MM-RDI" (i.e., the index) is 60.85 ± 25.28. 4. In the supplementary information, Figure S4, "MM-RDI results by 4 levels of OSAS severity defined by PSG (AHI)," shows the spread of the adjusted index calculated by Brizzy in each level of OSA. While the index spread is greater in severe OSA, most values remain nested about the 15n/h threshold.	

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Nomics S.A-Nicolas Stefenatto	15	135-144	5.7.11	Concerning the modelling of cohort 11 – True severe Conservative management Issue 2. Individuals sorted to cohort 11 are not modelled as benefitting from treatment. This may be an oversight in original conception since this approach contrasts the modelling of true mild OSAHS (cohorts 5 to 6d) where treatment is applied. The two cohorts ought to be modelled the same since the only difference (true disease status) is unknown to the treating clinician. The absence of treatment benefit for cohort 11 cases contrasts the stated disease management according to NICE clinical guideline NG202, which recommends the use of CPAP in people with mild OSA who have symptoms that affect their quality of life and usual daytime activities (Recommendation 1.5.2). Symptoms of OSAH are present in true mild cases, which is evident in the model based on the reduced baseline utility for true mild cases. There is no scenario analysis which truly represents the impact of these compounding issues 1 and 2, however, the scenario exploring a higher proportion receiving CPAP treatment for true mild cases (75% versus 25% in the base case) illustrates the expected direction of change in the INMB. The INMB of Brizzy versus RP in the EAG base case is -£15.17 at the £20,000 threshold in the deterministic analysis, and -£3.29 in the probabilistic analysis. With even a moderate level of treatment benefit given to cohort 11 (adjustment for Issue 2) we would expect a revised estimate of INMB to be positive. Further, the likelihood of cost-effectiveness at the £20,000 threshold would increase further if misdiagnosed cases of severe OSAHS were modelled as moderate cases rather than mild (adjustment for Issue 1). Considering a model adjustment for Issue 2 only, giving treatment benefit to misdiagnosed severe OSAHS, we redirected the 34 individuals of cohort 11 into the true mild cohort 5 (both strategies). Thereby effectively awarding treatment benefit to all individuals in the model considered to have mild OSAHS, correctly or incorrectly. This simple change increased the Brizzy INMB vs RP from -£15.17 to £257.79 at the £20,000 threshold.	In the base case analysis 20% of individuals who are diagnosed as having mild OSAHS would receive CPAP, regardless of their true severity, please see Section 5.7.11. This is reflected in the model on the spreadsheet “Diagnostic Results” where the second table (titled: Distribution of patients to the interventions) splits the cohorts to the relevant interventions. Specific to this comment, formula in cell G43 specifies that 20% of those from cohort 11 would receive CPAP. Furthermore, those individuals would benefit from CPAP since they are truly severe, as indicated in Table 28. We believe this misunderstanding may have originated from a cell within Cohort 11 erroneously stating that the intervention is Conservative management, and we apologise for this. The assumption that 20% of patients diagnosed as having mild OSAHS would receive CPAP, is based on expert

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				These issues particularly impact Brizzy since the key input parameter defining the proportion of the starting cohort entering cohort 11 is test sensitivity at the AH1= 15 (higher) cut-off. We are concerned that the EAG base case as it is presently estimates a INMB which may give support to a negative recommendation; therefore, we urge the EAG to adapt the approach taken to the modelling of cohort 11 by awarding some treatment benefit.	opinion (see Section 5.7.11). As the company point out there is a scenario analysis where this percentage is increased to 75% (again based on expert opinion), albeit with a lower compliance rate. In this scenario all novel devices are estimated to be cost-effective compared to respiratory polygraphy at £20,000 and £30,000 per QALY gained. Please see Section 5.10.2 and Table 47 for these results.
Acurable Limited- Orsina Desi & Emilio	16	6	Conclusions	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' The authors state *"It is difficult to assess the cost effectiveness of the novel devices compared with respiratory polygraphy, and the relative clinical and economic effects of the different novel devices are unclear"*. This statement is potentially misleading. This can be understood to mean that the relative clinical and economic effects are unclear in general, when in reality, if devices were to be analysed in isolation, taking into account the wider clinical and economic effects rather than just those that have been considered for this economic analysis, the picture would be very different. In other words, it is the assumptions made in the context of this work that make the results unclear. We understand the authors have already declared the limitations before. However, considering that this report is supposed to be accessible to the public, the conclusions should not leave room for interpretation with a statement as assertive as this.	In the previous response to comments, we justified the restriction of clinical and economic effects to those we have modelled. For instance, as NICE's scope is NHS and PSS we only consider RTA costs that fall on the NHS, furthermore we do not consider loss of productivity at work or school. As the previous paragraph – titled Limitations - highlights limitations of the modelling (as the company point out), we do not believe it needs to be repeated here.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Could it please be reiterated in the conclusions why the results are unclear? We suggest stating that *"the clinical and economic effects of the different novel devices are unclear due to the limitations of the study itself, which is restricting the analysis to a subset of benefits"*, or something along those lines.	
Acurable Limited- Orsina Desi & Emilio	17	13	Conclusions - Suggested research priorities	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' The authors suggest as a research priority that "The diagnostic accuracy of home based polygraphy is compared against a laboratory PSG standard". This point should be removed since it has the potential to lead people to do research that is potentially flawed from the outset and hence would not be ethical. The reason for this is that the only way of doing the suggested research would be by testing patients on two completely different nights (as a person cannot be simultaneously in the sleep clinic doing PSG and at home doing RP), and there is a very significant night to night variability in the diagnostic indexes in a very significant number of patients with OSA. Even if the uncertainty of that variability was somewhat statistically quantified based on what is known this would be so large that it would render this exercise pointless.	We acknowledge that there would be uncertainty, though we do not consider that this would be of the magnitude suggested. Our view is that the research recommendation is still important.
Acurable Limited- Orsina Desi & Emilio	18	29	1.1	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' Last paragraph of Background; Description of the health problem: Poor school performance should be added to the list of effects in children.	In section 1.1 (last paragraph of Background; Description of the health problem) poor school performance has been added to the list of effects in children (EM)
Acurable Limited- Orsina Desi & Emilio	19	30	1.2	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf'	The text has been amended as suggested

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Diagnostic tests for OSAHS; Respiratory polygraphy; We suggest this paragraph should say “signals directly resulting from airflow in the respiratory tract”, rather than “airflow”, because what the nasal cannulas measure is pressure and/or temperature. Also, the list is missing pulse or heart rate, and should say “peripheral oxygen saturation” rather than "oxygen levels".	
Acurable Limited- Orsina Desi & Emilio	20	31	1.2	‘The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf’ The first two sentences at the top of page 31: When describing the autogenerated sleep report, it should say “with details that may include….”, rather than “with details including”.	The text has been amended as suggested.
Acurable Limited- Orsina Desi & Emilio	21	31	1.3	‘The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf’ First paragraph, Description of diagnostic technologies under assessment: This should also include comment on the fact that these devices might allow a more natural sleep, hence the signals will be more representative of the true physiological state of the patient (and hence of the presence, absence, and severity of the disease).	The text has been amended as suggested.
Acurable Limited- Orsina Desi & Emilio	22	32	1.3	‘The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf’ First line of page 32, Description of diagnostic technologies under assessment:	Nasal airflow has been deleted.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Strictly speaking no device measures “nasal airflow”. What any device that measures “airflow” is measuring is a signal directly generated (i.e. not a surrogate) by the movement of air in the respiratory system. Temperature, sound, pressure are then all under the umbrella of “airflow” as ways of measuring the signal generated and hence should be treated in the same way. The text should be rephrased to better represent this.	
Acurable Limited- Orsina Desi & Emilio	23	40	1.4.1	‘The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf’ Care pathways in adults: In the paragraph starting with “Although both the AASM and NICE…”, which discusses some limitations of home RP: added to “potentially interrupting sleep”, it should also say “affecting the patient’s sleep position and consequently natural patterns”.	The following text has been incorporated “affecting the patient’s sleep position and therefore natural sleep patterns.”
Acurable Limited- Orsina Desi & Emilio	24	125	5.7.5	‘The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf’ Test failure rates: The sources to estimate failure rate of RP should be referenced here. Otherwise, this remains a conversation with someone of the Newcastle Regional Sleep Service which is not scientifically supported. It also appears slightly odd that over a period of six months, the count comes out exactly to 1000 studies. How was this data collected and evaluated? This is important as the failure rate number is significantly lower than what has been reported in published studies. We note that as a scenario analysis, failure rate has now been taken from the Alsaif manuscript. As this data is redacted we cannot comment on it, but we highlight that another potential source of quantifiable evidence the authors have, should it be higher, is	The data provided were by month, and we have no reason to believe that the total number for those 6 months adding to 1000 is suspicious. We state in our report that reasons for failures from the Newcastle data were due to adequate data. We have now expanded this description to: “For RP and oximetry, it is reported that failures recorded in the Newcastle Regional

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Devani et al, since patients there followed the conventional pathway for RP and there is a diagram showing how many RP tests failed. Please incorporate references and include an explanation of how the data from Newcastle was quantified.	Sleep Service data, were due to a lack of adequate data. signals missing, too short, unable to gain useful information from test to draw conclusion" (Please see Section 5.7.5 of Report) 
Acurable Limited- Orsina Desi & Emilio	25	127	Table 20	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' Estimates of failure rates used in the model: The first row should be amended to include, in addition to communication with one sleep service, the addition of published evidence, since this exists. An explanation of which methods were followed to extract the data in Newcastle and what the definition of a "failed test" was should also be included.	The failure rates shown in Table 20 are those used in the model, and so we have added the Alsaif data to this table, please see updated Table 20. As we do not use any other failure rate data for RP in the model we have not added anything further to Table 20. Information on the RP failure rates from Newcastle is provided in the text prior to this table and so we do not repeat this in Table 20.
Acurable Limited- Orsina Desi & Emilio	26	127	5.7.6	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf'	In the one-way sensitivity analyses, we change each parameter, in turn, to assess the impact it has on the cost-

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Time to diagnosis and treatment: It is unclear what has been assessed in the one-way sensitivity analysis described here. Please could the EAG clarify whether the time to diagnosis or treatment has been increased and decreased just for the novel devices, or for the comparator of home RP as well, in this analysis? In the response to comments on the previous EAR (comment 52), the EAG state that "the tornado plots in Appendix 9b show [..] when time to diagnosis or treatment is assumed to be 1.5 months for the novel devices and 3 months for home RP", however in Table 78 of the EAR; Parameter values used in the base case, probabilistic analysis and one-way sensitivity analysis, it appears that the times to diagnosis or treatment in the one-way sensitivity analysis have been changed for all devices: novel, oximetry and RP. From the report text, section 5.7.6, it is unclear whether there is a typo in the table. Please could the purpose of this one-way sensitivity analysis be clarified? For the avoidance of doubt, we believe it is important for an analysis to be run considering the reduced time to diagnosis. As appears to be reported in Alsaif (we are not sure because the information is redacted), and as we have seen in preliminary results shown to us from an independent (central London Hospital) research study of time to diagnosis when using AcuPebble rather than RP, there is growing evidence that using novel devices would reduce the time to diagnosis significantly, and this should be taken into account.	effectiveness results. The company is correct that the time to diagnosis and time to treatment have been changed for all devices: novel devices, oximetry and RP. However, these changes are not done simultaneously. For instance, taking time to diagnosis for AcuPebble, in the one-way sensitivity analysis, we would change this to 1.5 months, record the results, then change it to 6 months and record the results. This is what is presented in the tornado plots in Appendix 9b. The one-way sensitivity analysis serves to assess the impact of individual parameters on the model results, using reasonable upper and lower values. Thus, for many of the one-way sensitivity analyses conducted, we use the upper and lower 95% confidence limits. We have added clarification of the purpose of the one-way sensitivity analyses to section 5.10.3.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					In addition to these one-way sensitivity analyses, we have run a scenario analysis on time to diagnosis and treatment based on the Alsaif data. Please see Table 38 and results in Table 77.
Acurable Limited- Orsina Desi & Emilio	27	158	Table 33	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' Road Traffic Accidents: The Department of Health publishes costs that go beyond the NHS and the police. These should be included, considering they are costs for the taxpayer and ultimately any NICE evaluation has an effect on how tax payer money is spent and how much goes to the NHS. Most specifically "Lost output" should be amended. We note the EAGs response to comment 184 in the previous response, that the cost perspective for NICE is the NHS and Personal Social Services, however feel that the above could be mentioned in the report.	We have clarified in the report why we only include NHS costs associated with RTAs. Please see Section 5.7.17.
Acurable Limited- Orsina Desi & Emilio	28	159	5.8	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' Model assumptions: As mentioned in comment 011 above, we are aware of preliminary results from an, independent from Acurable, research study which shows that using novel devices (in the case of this particular work, AcuPebble) instead of RP can significantly reduce the time to diagnosis. We note also that while it is not possible to see the data from Alsaif due to redaction, the report suggests that this is also reported in this work.	As reported in Table 38, we undertake a scenario analysis adjusting the time to diagnosis and treatment for the novel devices.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
Acurable Limited- Orsina Desi & Emilio	29	214	7.2	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' The authors suggest as a research priority that "The diagnostic accuracy of home based polygraphy is compared against a laboratory PSG standard". This point should be removed since it has the potential to lead people to do research that is potentially flawed from the outset and that would be unethical. The reason is that the only way of doing this is by testing patients in two completely different nights (as it is obvious a person cannot be simultaneously in the sleep clinic doing PSG and at home doing RP), and there is a very significant night to night variability in the diagnostic indexes in a very significant number of patients with OSA. Even if the uncertainty of that variability was somewhat statistically quantified based on what is known this would be so large that would render this exercise pointless.	Please see our response to comment number 17 above.
Acurable Limited- Orsina Desi & Emilio	30	360	Table 78	'The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' In this table, the time to diagnosis has been reduced to 1.5 months for Respiratory polygraphy as well as for novel devices, however the response to comment 52 on the previous version of the EAR suggests that the purpose of this parameter change is to assess the impact of reducing time to diagnosis for novel devices, while keeping RP at 3 months. Please could the purpose of this one-way sensitivity analysis be clarified, and an analysis where the time to diagnosis with novel devices is shorter than RP be conducted, if it has not been?	Please see response to comment 26 above.
Sunrise - Fabien Crespo	31	35 (of no track	1.3.4	Thank you for the changes. We confirm that the latest versions of the Sunrise HCP user manual and of the Sunrise patient user manual do not state "Do not use the	Noted, thank you.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
		change s file)		device with a pacemaker or similar implanted device since it could impair its functioning" anymore.	
Sunrise - Fabien Crespo	32	35	1.3.4	Could you please make the following update to correct a typo error? Thank you. • Replace '…respiratory effort related arousal (RERA) index, respiratory effort), awakening and arousal index…' with '…respiratory effort related arousal (RERA) index, respiratory effort, awakening and arousal index…'	The text has been amended as suggested.
Sunrise - Fabien Crespo	33	58	4.2.1	Could you please make the following update to add more detailed and comprehensive information? Thank you. • Replace 'Six studies (Devani et al., 2021; Martinot et al., 2017; Van Pee et al., 2022; Lyne et 23; Sanchez Gomez et al., 2024 and Pepin et al., 2020) reported sample size calculations and all were subsequently adequately powered, except for Pepin et al., (2020) which recruited just under the minimum target number of patients in one of the study groups.' with 'Six studies (Devani et al., 2021; Martinot et al., 2017; Van Pee et al., 2022; Lyne et 23; Sanchez Gomez et al., 2024 and Pepin et al., 2020) reported sample size calculations and all were subsequently adequately powered, except for Pepin et al., (2020) which recruited just under the minimum target number of patients in one of the study groups (46 patients instead of 50 patients for the non-OSA group).'	This text has been amended as suggested..
Sunrise - Fabien Crespo	34	79	4.5.1	We believe it is important to mention that Sunrise underwent evaluation using both an RP reference standard and a PSG reference standard. Moreover, Sunrise is unique for being assessed with in-lab PSG and home PSG. It is also important to highlight the recognized limitations of RP in	We have revised the text in this section to note that Sunrise has been assessed against RP (Alsaif) and PSG (home PSG in

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				evaluating clinical effectiveness and to emphasize the preference for PSG as the reference standard for such evaluations. The Alsaif 2023 study reinforces this by demonstrating instances where Sunrise accurately identified patients overlooked by RP. Thank you.	Kelly et al; lab based PSG in Pepin et al. It is not appropriate to discuss the limitations of RP and preferences for PSG in this section, this is mentioned elsewhere in the report. In this section the focus is on the performance of the novel devices.
Sunrise - Fabien Crespo	35	87	4.8.1	Could you please make the following update to correct a typo error? Thank you. • Replace 'Martinont et al (2022)' with 'Martinot et al (2022)'	This has been updated as suggested.
Sunrise - Fabien Crespo	36	126	5.7.5	Failure rates for Sunrise are taken from Kelly 2022 in base case analysis and from Alsaif 2023 in a scenario analysis. To reflect the advancements and updates made to the Sunrise device and its user instructions, including improvements in Bluetooth connection stability, could you please mention in this section what is already mentioned in the discussion (see below)? Additionally, could you please run scenario analyses without the failure occurrences associated with Bluetooth connection and sensor association for Kelly 2022 (corresponding failure rate of 1/38) and Alsaif 2023 (corresponding failure rate of 1/40)? Thank you. *It should be acknowledged that some of the factors contributing to failed tests were not anticipated by the study investigators and, with the benefit of hindsight, were preventable. The expectation is that the learning from these instances will have prompted necessary changes to testing protocols,*	We have not conducted additional scenario analyses as requested here by the company. Instead, we refer to the results of the one-way sensitivity analyses for Sunrise failure rates (please see Table 78 and tornado plots in Figure 15, Appendix 9). These results are based on using the lower and upper 95% confidence limits around the estimated failure rate of 4/38 from Kelly 2022. The lower limit used in the sensitivity analysis is 0.77%, which is much lower than the 2.6% (1/38) and 2.5%

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				device features and user instructions to avoid similar failures occurring again. If this is the case then novel device failure rates in clinical practice would be lower than those reported in the studies, all other factors being equal.	(1.40) noted here by the company. The tornado plots show that even this very low estimate of 0.77% for Sunrise has little impact on the cost-effectiveness results compared to RP. Furthermore, their impact when compared to oximetry is even less significant, and is not visible in Figures 15a and 15b.
Sunrise - Fabien Crespo	37	127	5.6.7	"The systematic review of clinical effectiveness did not identify any studies reporting data on the time to diagnosis or time to treatment associated with use of any of the novel devices." We suggest that the work from Alsaif 2023 deserves mention here, thank you.	Thank you for highlighting this omission, the report has now been amended to include Alsaif 2023, please see Section 5.7.6.
Sunrise - Fabien Crespo	38	149 154 348	5.7.12 5.7.12 Appendix 8	We confirm that in the event of a device failure with Sunrise, we will replace the device at no additional cost to the NHS. It seems there was a misunderstanding with the information provided earlier. This policy aligns with the practices considered for other single-use devices in the external assessment report the current practice for Sunrise in the UK market. Could you please make the following update on page 149, review the table 31, the table 74 and the economic model accordingly? Thank you. • Replace 'In base case analyses, we assume that should a sleep study fail, the full cost of a new device would be incurred to undertake a second sleep study.' with 'In base case analyses, we assume that there is no additional sleep study or device cost to the NHS for a failed sleep study with the Sunrise device.'	As stated in the previous response to comments (comment 37): "In response to an EAG request for information, the company reported that [redacted]

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					[redacted] Due to a preference from NICE to avoid CIC data from the base case analysis, where possible, we assume that costs of devices for any repeat tests are included. However, we report on a scenario analysis [redacted] We have run the model assuming no additional device costs for any repeat studies needed due to failures of the Sunrise device. The INMB gained for Sunrise vs RP increases to £134 at £20,000 per QALY (from £127 in the base case), and £107 at £30,000 per QALY gained (from £100 in the base case).

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
					Compared to oximetry, the INMB for Sunrise increases to £1,158 at £20,000 per QALY gained (from £1,152 in the base case) and £2,045 at £30,000 per QALY gained (from £2,039 in the base case).
Sunrise - Fabien Crespo	39	295	Appendix 5	We believe additional clarifications are necessary regarding the application of post-hoc analysis: 1. Pepin 2020: This study utilised post-hoc analysis to establish the thresholds for ORDI to be used with the Sunrise device for adults. These thresholds are clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included. This approach is standard practice in the development of a novel device and its comparison with the reference standard. It establishes thresholds for the novel device based on study findings, which are subsequently applied to the commercially available product. Thus, the performance metrics derived from these studies accurately reflect real-world clinical usage. 2. Martinot 2022 (child): This study utilised post-hoc analysis to establish the thresholds for ORDI to be used with the Sunrise device for children. These thresholds are clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included. This approach is standard practice in the development of a novel device and its comparison with the reference standard. It establishes thresholds for the novel device based on study findings, which are subsequently applied to the commercially available product. Thus, the performance metrics derived from these studies accurately reflect real-world clinical usage. 3. Kelly 2022: This study utilised post-hoc analysis to confirm the thresholds for ORDI to be used with the Sunrise device for adults. The first threshold of 9.53 was very close to the first threshold established by Pepin	Thank you. However, the information provided does not add much beyond what is already available. Our position on the issue of post-hoc analysis remains the same. Therefore we have not amended our risk of bias judgements for these studies or revised the text in the report.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				2020 (7.63) and the second threshold was the same as observed previously by Pepin 2020. This allowed to confirm the good selection of the thresholds for ORDI to be used with the Sunrise device for adults. 4. Alsaif 2023: The primary objective of the study was to assess the time to treatment decision using the Sunrise device compared to home respiratory polygraphy. It did not intend to reassess the thresholds to be used with the Sunrise device, which would have been unfeasible due to the study's design anyway. The Sunrise sleep reports, which were automatically generated, were directly used for treatment decisions. The thresholds for ORDI for adults indicated in the Sunrise report are the ones from Pepin 2020 and these are clearly indicated, and to further aid in result interpretation, a colour-coded severity scale is also included. Considering these clarifications, could you please review the assessments of the studies below and the corresponding table 8 on page 72? Thank you. • Page 323 - Pepin et al., 2020 Regarding index test - risk of bias - judgment: Please see clarification 1 above. We recommend adjusting the associated risk to “LOW”. Additionally, more detailed and comprehensive information should be incorporated into the comments. Regarding index test - concerns regarding applicability - judgment: We recommend adjusting the associated concern to “UNCLEAR” given that the study was conducted in a sleep laboratory rather than a home setting. The post-hoc analysis to establish the thresholds for ORDI to be used with the Sunrise device for adults does not introduce concerns regarding applicability given that these thresholds are used and clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included (please see clarification 1 above).	

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				• Page 311 - Martinot et al., 2022 (child) Regarding index test - risk of bias - judgment: Please see clarification 2 above. We recommend adjusting the associated risk to “LOW”. Additionally, more detailed and comprehensive information should be incorporated into the comments. Regarding index test - concerns regarding applicability - judgment: We recommend adjusting the associated concern to “UNCLEAR” given that the study was conducted in a sleep laboratory rather than a home setting. The post-hoc analysis to establish the thresholds for ORDI to be used with the Sunrise device for children does not introduce concerns regarding applicability given that these thresholds are used and clearly indicated in the Sunrise report, and to further aid in result interpretation, a colour-coded severity scale is also included (please see clarification 2 above). • Page 301 - Kelly et al., 2022 Regarding index test - risk of bias - judgment: Please see clarification 3 above. We recommend adjusting the associated risk to “LOW”. Additionally, more detailed and comprehensive information should be incorporated into the comments. Regarding index test - concerns regarding applicability - judgment: We recommend adjusting the associated concern to “LOW” given that the study was conducted in a home setting. The post-hoc analysis to confirm the thresholds for ORDI to be used with the Sunrise device for adults does not introduce concerns regarding applicability (please see clarifications 1 and 3 above). • Page 295 - Alsaif et al., 2023 Regarding index test - risk of bias - question 2:	

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				Please see clarification 4 above. Moreover, it has been confirmed to us that Pepin 2020 is mentioned in the draft manuscript. We recommend adjusting the associated assessment to “Yes”. Additionally, associated comments should be reviewed. Regarding index test - risk of bias - judgement: Considering the remarks above, we recommend adjusting the associated risk to “LOW”. Regarding index test - concerns regarding applicability - judgment: Considering the remarks above, we recommend adjusting the associated concern to “LOW”. In addition, could you please review the following text passages (non-exhaustive)? There is significant repetition, which is misleading considering the clarifications provided. It is necessary to remove and/or add more detailed and comprehensive information. Thank you. • Page 9 - “Risk of bias assessments of the studies indicated a low risk of bias for most domains, however there were instances of high or unclear risk of bias for some domains, including bias in the analysis of the index test.” • Page 12 - “The cost-effectiveness analysis is limited by the availability and quality of data for many of the model components, including limited accuracy data in the home environment and the effects of post-hoc optimisation of thresholds for sensitivity and specificity.” • Page 64 - “Post-hoc analysis was performed to optimize the diagnostic cut-offs” • Page 69 - “Post-hoc analysis was performed to optimize the cut-offs”	

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				• Page 70 - "A high risk of bias was judged in the conduct or interpretation of the index test in three studies (Kelly et al., 2022, Martinot et al., 2022 and Pepin et al., 2020) all of which used post-hoc analyses to optimise diagnostic cut-off points, potentially over-estimating novel device diagnostic accuracy." • Pages 70-71 - "Regarding applicability to the decision problem, most studies were judged as low concern for the patient selection and the reference standard domains. However in many studies it was unclear whether the conduct, or interpretation of the index test was relevant to the decision problem. This judgement was made for all studies where the novel testing device was used in a sleep laboratory (concomitant to PSG testing), rather than its intended setting (i.e. the patient's home). Two studies were also rated unclear for this domain although they were conducted in a home setting. Alsaif et al (2023) did not report on the thresholds used in their study and Storey et al., 2022 did not report details of the conduct and interpretation of the index test. For four studies, the judgements were of high concern – in Kelly et al., 2022, Martinot et al., 2022 and Pepin et al., 2020 all used post-hoc analyses to optimise diagnostic cut-off points, while in Martinot et al., 2015 diagnostic accuracy results for against PSG or any other reference standard were not reported and the study was conducted in a sleep laboratory rather than the home setting." • Page 77 - "Post-hoc optimisation of the diagnostic cut-off points for the novel device against reference standard cut-offs" • Pages 79-80 - "It is also important to note that at least three of the studies performed post-hoc optimization of the diagnostic cut-off points for the novel device against reference standard cut-offs (Kelly et al 2022; Martinot et al, 2017; Pepin et al, 2020). In this approach, optimal cut-offs on the ROC curve (defined as maximum sensitivity and specificity values simultaneously) are assessed and diagnostic performance metrics for this cut-off are estimated. However, this approach can be open to selective reporting of results from the cutoffs that perform well, thus over-estimating diagnostic accuracy. Pepin et al (2020), for example, sought to optimise the clinical performance of the RDI derived from the Sunrise system analysis in ruling in a diagnosis	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				of OSA at the two reference thresholds of PSG of at least 5 events/h or at least 15 events/h (leading to the classification of participants as not having OSA, or having OSA with comorbidities or having OSA irrespective of comorbidities, respectively). They used ROC curves and defined the trade-off between true-positive rates and false positive rates at PSG-RDI of at least 5 events/h and at least 15 events/h. The optimal diagnostic cutoff was adjusted, and the diagnostic PSG-RDI cutoffs of at least 5 events/h and at least 15 events/h were extrapolated to Sr-RDI cutoffs of at least 7.63 events/h and at least 12.65 events/h. Whilst the results of the Pepin et al study can be at increased risk of bias subsequent studies using the cut-offs established by Pepin would not be at such risk because the thresholds will be pre-specified. (NB. The Sunrise manufacturer confirmed to NICE that the device uses pre-specified thresholds established in Pepin et al., 2020). Notably, Kelly et al 2022, a more recent evaluation of Sunrise, did not report using pre-specified thresholds from Pepin et al., 2020. Instead, a post hoc analysis was done to optimise the cut-off points of MM-ORDI for diagnostic decisions, compared with reference standard cut-off values of obstructive PSG-ORDI. It is unclear why Pepin's thresholds were not used, but it might be because the diagnostic indices are not the same (Pepin used RDI, Kelly used ORDI)." • Page 88 • Page 119 - "We note that both studies use post-hoc optimisation of thresholds, which is likely to overestimate accuracy of the devices, as highlighted in the clinical effectiveness risk of bias assessment (see Section 4.4)." • Page 119 "Using accuracy estimates based on MM-ORDI 9.53 cut-off; using accuracy estimates based on MM-ORDI 12.65 cut-off;" • Page 207 - "Moreover the data used in the base case analysis to inform the accuracy estimates for novel devices are all derived from a clinical setting, with three based on post-hoc optimisation of thresholds, which is likely to overestimate the accuracy of the devices." • Page 208 - "Post-hoc optimisation of diagnostic thresholds within accuracy studies (such as for Sunrise in Pepin et al (2020) and Kelly et al (2022)."	

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG Response
				• Page 209 - "Data on the diagnostic accuracy of the novel devices in children is sparse: with only one published study reporting sensitivity and specificity (Martinot 2022 for the Sunrise device, which used post-hoc optimisation);"	
Sunrise - Fabien Crespo	40	357-358	Appendix 9	Could you please make the following updates? • Replace values for 'Sunrise: low specificity' by 0.94 (0.91, 0.97) • Replace values for 'Sunrise: high sensitivity' by 0.92 (0.90, 0.94) This is to correct typo errors and align with Pepin 2020 and the table 18 on page 121. The values in the economic model seem to be correct. Thank you.	These typos have now been amended, please see Table 78.

Section B Economic model - Comments

Stakeholder	**Comment**	**Description of problem**	**Description of proposed amendment**	**Result of amended model or expected impact on the result (if applicable)**	**EAG response**
Zoll Itamar-Lior Solomon	1	WatchPAT One price: We noticed that the price of the WatchPAT One that we provided for the calculation is the list price, which is £80. However, the average selling price of the device based on 12 months records is £67	Please change the price of the WatchPAT One to £67.	ICER Vs OX - £13,424 (was £9,163) ICER Vs RP - £8,617 (was 8,777) *See comment #3 regarding these values	We can confirm that the updated ICERs given by the company here are correct. In addition to the ICERs, the INMBs for WatchPAT One assuming this lower cost compared to RP are -£22 at the £20,000 per QALY gained threshold and -£55 at the £30,000 per QALY gained threshold.

DAP70 Novel home-testing devices for diagnosing obstructive sleep apnoea/hypopnoea syndrome:

Zoll Itamar-Lior Solomon	2	The model includes posting cost of £5.98 for the WatchPAT300. However, we know that none of our clients is posting WatchPAT300 devices. When posting is needed they prefer to use the WatchPAT One.	Please remove the posting cost of £5.98 from the model	ICER Vs OX - £32,891 (was £16,172) ICER Vs RP - £7,888 (was 8,515) *See comment #3 regarding these values	The results of a scenario analysis assuming that novel devices are collected in person is provided in Table 77. These analyses show that at a threshold of £20,000 per QALY gained, WatchPAT 300 would have a INMB of -£6 (compared to -£13 in the base case where postage is assumed).
Zoll Itamar-Lior Solomon	3	Inconsistency in the excel in the Scenario result live table	We noticed that in line 5 “Incr cost (vs ox)” the values the formula are multiplied with cell E2, which is the cost of the RP, and the opposite occurs in line 13 “Incr cost (vs RP) where the formula are multiplied with cell F2 (Oxi)		Thank you for highlighting this inconsistency in the labelling of the table. We confirm that it is purely a labelling error and does not affect any results.
Zoll Itamar-Lior Solomon	4	Revised specificity in your reply to our comment (#25, Section 5.7.3) you advised us that due to the uncertainty of the estimation of specificity in the Tauman et al study, you will use a higher specificity, of 0.806. We could not identify this revised value in the Excel sheet.	Please make sure the new specificity value (0.806) replaces the 0.25 value in the excel and recalculate.		There has been a misunderstanding. Our response to the previous comment was not that we would use a higher specificity estimate of 0.806 in the base case. We stated that in the base case analysis we would use the specificity of 0.25 from Tauman et al, as it is the source of all of the diagnostic performance estimates for WatchPAT 300 and WatchPAT ONE. Instead, we pointed out that in one-way sensitivity analyses, we use an upper estimate of 0.806, which is higher than the

					company's original suggestion for this input value (0.60). Please refer to our response to comment 25 in the original response to comments.
The comments below relate to the page numbering in the clean version of the report; 1. REDACTED EAR ERRATUM [ACIC] No Track changes.pdf' Acurable Limited- Orsina Desi & Emilio	5	It appears that in the one-way sensitivity analysis assessing the impact of reducing the time to diagnosis for novel devices to 1.5 month, the time to diagnosis for RP has also been reduced to 1.5 months rather than being kept at 3 months.	An analysis should be conducted to assess whether the reduced time to diagnosis with novel devices has an impact on the model results. In Table 78, the time to diagnosis for RP in the one-way sensitivity analysis should remain 3 months. This is particularly important as there is growing evidence to support the statement that novel devices reduce time to diagnosis.	Model has not been rerun.	Please see response to comment 16 in Section A.
Sunrise - Fabien Crespo	6	We confirm that in the event of a device failure with Sunrise, we will replace the device at no additional cost to the NHS. It seems there was a misunderstanding with the information provided earlier. This policy aligns with the practices considered for other single-use devices in the external assessment report and is a	Cost of a failed sleep study with Sunrise device: • Posted = £4.42 + £8.43 + £5.67 + £2.99 = £21.91 • Collected = £4.42 + £8.43 + £5.67 = £18.92	ICER vs oximetry = £6,946 (before £7,020) ICER vs home RP = £69,862 (before £67,426)	Please see response to comment 38 in section A.

		standard for Sunrise in the UK market. Could you please make the following update on page 149, review the table 31, the table 74 and the economic model accordingly? Thank you. • Replace 'In base case analyses, we assume that should a sleep study fail, the full cost of a new device would be incurred to undertake a second sleep study.' with 'In base case analyses, we assume that there is no additional sleep study or device cost to the NHS for a failed sleep study with the Sunrise device.'			