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

ADDENDUM

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:
Date completed	19TH June 2024

1. Background

This is an addendum to the external assessment group (EAG) report produced by SHTAC and the Exeter Test Group for the NICE diagnostic assessment DG70.

This addendum has been produced in response to comments made by consultees on the draft NICE guidance published in 2024. It was commented that one of the study publications assessed for inclusion in the systematic review of test accuracy and clinical effectiveness (Martinot et al 2022) had erroneously been classified as a secondary publication of an existing included study (Pepin et al 2020). The company pointed out that the Martinot et al 2022 publication relates to a completely separate study, and its findings should therefore be included in the synthesis of study findings for consideration by the diagnostic advisory committee. Below we present a narrative review of the study and its results, and a critical appraisal using the QUADAS2 instrument.

2. Study design and characteristics

The primary focus of the publication was to explore the approach of near boundary labelling (NBL). The authors postulated that the risk of AHI-based severity mis-classification due to inter-human PSG rating could be reduced when considering borderline zones around the traditional fixed AHI thresholds. They applied the NBL approach to a clinical study aiming to validate a machine learning–based algorithm for mandibular movement signals (Sunrise, Namur, Belgium).

Since the issue of NBL is not central to the scope of this diagnostic assessment, and for brevity, we focus below on the diagnostic performance of Sunrise in terms of sensitivity, specificity and other metrics. These estimates are presented without the use of NBL, for comparability with the results of other studies in our systematic review

The study included 289 participants presenting with obstructive sleep apnea (OSA) suspicion. No details are given of the socio-demographic or health characteristics of participants.
The participants underwent an in-laboratory PSG coupled with simultaneous MM recordings using the Sunrise device. The PSG data were then manually scored by two experienced and blinded investigators. The collected MM data were automatically analysed by a machine learning algorithm developed by Sunrise.

3. Study results

The study reports that, based on the conventional rules for severity grading, the participants were categorized into non-OSA (n = 14; 4.8%), mild (n = 109; 37.7%), moderate (n = 113; 39.1%), and severe OSA (n = 53; 18.4%).

Table 1 below is a confusion matrix showing the distribution of participants classified across severity groupings by Sunrise and PSG.

Table 1 - distribution of PSG-AHI scores within four conventional severity levels for PSG scoring and sunrise classification (NB. EAG converted proportions presented in study publication Figure 1 to numbers of patients)

	OSA Severity Sunrise				
OSA Severity PSG	**Normal**	**Mild**	**Moderate**	**Severe**	**Total**
Normal	12	2	0	0	14
Mild	2	100	7	0	109
Moderate	0	13	97	3	113
Severe	0	1	9	43	53
Total	14	116	113	46	289

Table 2 below gives diagnostic accuracy estimates based on the figures given in table 1 above. This is based on the threshold for test positivity incorporating mild, moderate and severe groupings combined.

Table 1 Diagnostic accuracy based on figures from above table (true positive = mild, moderate or severe OSA; true negative = not mild, moderate or severe OSA)

	Reference standard positive	**Reference standard negative**	**Total**
Index test positive	273	2	275
Index test negative	2	12	14
Total	275	14	289
Accuracy	98.62% (95% CI 96.49% to 99.62%)		

Diagnosis	**Value**	**95% CI**
Clinical sensitivity a / (a + c)	99.27%	97.40% to 99.91%
Clinical specificity d / (b + d)	85.71%	57.19% to 98.22%
PPV a / (a + b)	99.27%	97.42% to 99.80%
NPV d / (c + d)	85.71%	59.73% to 96.04%
Positive likelihood ratio [sensitivity/(1-specificity)]	6.95	1.93 to 25.07
Negative likelihood ratio [(1-sensitivity)/specificity]	0.01	0.00 to 0.03
Disease prevalence	95.16%	92.01% to 97.33%

4. Critical appraisal

Appendix 1 gives the EAG's critical appraisal of the study. Based on the available information we judged the study to be at low risk of bias for some of the domains, and unclear risk of bias for others. Overall the limited available information from this study does not permit a full critical appraisal of the risk of bias and our overall judgement is that this is unclear.

References

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

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.

Appendix 1. DAP70: QUADAS- 2 Risk of bias and applicability study assessments

Study - First Author: Martinot (study in adults)	**Year:2022**	**Rayyan No: 566581088**
DOMAIN 1: PATIENT SELECTION	**Assessment (delete as appropriate)**	**Comments**
A. Risk of Bias		
Signalling question 1: Was a consecutive or random sample of patients enrolled?	Yes	"Consecutive participants presenting with obstructive sleep apnea (OSA) suspicion"
Signalling question 2: Was a case-control design avoided?	Yes	"Consecutive participants presenting with obstructive sleep apnea (OSA) suspicion"
Signalling question 3: Did the study avoid inappropriate exclusions? *(Note: Remember that the device may be contraindicated in certain patient populations)*	Unclear	Exclusions were not reported. Inclusion criteria were not reported
Judgment: Could the selection of patients have introduced bias?	**RISK:** UNCLEAR	Unclear as exclusion criteria were not reported
B. Concerns regarding applicability		
Judgment: Is there concern that the included patients do not match the review question?	**CONCERN:** UNCLEAR	"Consecutive participants presenting with obstructive sleep apnea (OSA) suspicion" but unclear as exclusion criteria were not reporte
DOMAIN 2: INDEX TEST(S)	**Assessment (delete as appropriate)**	**Comments**
A. Risk of Bias		
Signalling question 1: Were the index test results interpreted without knowledge of	Yes	Data were automatically analysed

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)*		
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	Unclear what thresholds were used
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	No comment
DOMAIN 3: REFERENCE STANDARD		
A. Risk of Bias		
Signalling question 1: Is the reference standard likely to correctly classify the target condition?	Yes	In Lab PSG
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	"The PSG data were then manually scored by two experienced and blinded investigators"

Judgment: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK:** 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	
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 testing of Sunrise and PSG
Signalling question 2: Did all patients receive a reference standard?	Yes	("Based on the conventional rules for severity grading, the participants could be categorized into non-OSA (n = 14; 4.8%), mild (n = 109; 37.7%), moderate (n = 113; 39.1%), and severe OSA (n = 53; 18.4%). Corresponding proportions of the seven categories in the NBL classification are presented in Table 1" – if you add the number of participants in each category the total is 289, which is the total sample of enrolled participants)
Signalling question 3: Did patients receive the same reference standard?	Yes	

Signalling question 4: Were all patients included in the analysis?	Yes	("Based on the conventional rules for severity grading, the participants could be categorized into non-OSA (n = 14; 4.8%), mild (n = 109; 37.7%), moderate (n = 113; 39.1%), and severe OSA (n = 53; 18.4%). Corresponding proportions of the seven categories in the NBL classification are presented in Table 1" – if you add the number of participants in each category the total is 289, which is the total sample of enrolled participants)
Judgment: Could the patient flow have introduced bias?	**RISK:** LOW	