National Institute for Health and Care Excellence

1

Draft

Thyroid cancer

[H] Evidence review for the clinical and cost effectiveness of initial treatments for people with differentiated thyroid cancer

NICE guideline <number>

Evidence reviews underpinning recommendations 1.3.1 to 1.3.10 and research recommendations in the NICE guideline

June 2022

Draft for Consultation

These evidence reviews were developed by the National Guideline Centre

1

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN:

Contents

1

1 The clinical and cost effectiveness of initial treatments for people with differentiated thyroid cancer

1.1 Review question

1.1.1 For people with differentiated thyroid cancer, what is the clinical and cost effectiveness of active surveillance, hemi-thyroidectomy (with or without prophylactic or therapeutic node dissection) or total thyroidectomy (with or without prophylactic or therapeutic node dissection)?

1.1.2 Introduction

Recent years have seen a significant increase in the incidence of differentiated thyroid cancer (SEER Cancer Statistics Review 1975-2916, NIH., Lim et al, JAMA 2017; 317; 1338-1348) notwithstanding that post-mortem series show a prevalence of undiagnosed, differentiated thyroid cancer of up to 11%, which has been stable over decades (DOI: 10.1200/JCO.2016.67.7419 Furuya-Kanamari et al Journal of Clinical Oncology 34, no. 30 (October 20, 2016) 3672-3679.) The absence of a corresponding increase in mortality suggests that the increased incidence may be driven by an increase in thyroid imaging identifying indolent and low risk cancers which may not require treatment at all. (https://doi.org/10.1089/thy.2016.0100)

For most small thyroid cancers, debate has focused on the extent of surgery required (total versus hemithyroidectomy) as initial treatment. There has also been a great deal of debate regarding the management of cervical lymph nodes in patients with differentiated thyroid cancer with studies showing a high prevalence of lymph node metastases at the time of surgery, but without evidence for improvement in mortality in groups undergoing lymph node dissection. This review seeks to determine the best initial surgical treatment strategy for differentiated thyroid cancer.

1.1.3 Summary of the protocol

For full details see the review protocol in Appendix A.

Table 1: PICO characteristics of review question

Population	Inclusion: People aged 16 or over with a diagnosis of differentiated thyroid cancer Exclusion: Children and young people under 16 years.
Interventions	• Active surveillance • Hemithyroidectomy + node dissection (level 6) • Total thyroidectomy + node dissection (level 6) (including completion) • Total thyroidectomy + lateral neck (+ level 6) (including completion) • Hemithyroidectomy only • Total thyroidectomy only (including completion) • Other types of thyroidectomy / subtotal
Comparisons	• Each other • However, note that other types of thyroidectomy / subtotal are only to be compared to active surveillance

Outcomes	**Critical:** • Mortality • Quality of life • Cost effectiveness • Local cancer progression • Incidence of distant metastases • Cancer recurrence • Postoperative dysphagia • Recurrent nerve palsy • Hypoparathyroidism • Need for further treatment Time of follow up: longest available
Study design	• Published NMAs and IPDs will be considered for inclusion. • Systematic reviews of RCTs • RCTs • PROTOCOL AMENDMENT: Because of the importance of obtaining data on active surveillance versus other comparators, together with the likelihood of there being no RCTs looking at active surveillance, we will drop down to observational studies (if there are zero RCTs) for comparisons involving active surveillance ONLY. These observational studies MUST either have a propensity-score matching methodology, or a multivariable analysis that accounts for biologically plausible confounders. This decision on a protocol change has been made before any literature has been searched or extracted and is therefore not influenced by analysis findings. The decision is the result of further reflection by the development team and has the aim of generating the most useful set of data possible.

1.1.4 Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A and the methods document.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.1.5 Effectiveness evidence

1.1.5.1 Included studies

A search was conducted for randomised controlled trials (RCTs) of initial treatments for differentiated thyroid cancer. Trials were included if they compared any of the following initial treatments: active surveillance, hemithyroidectomy (with and without prophylactic or therapeutic node dissection) and total thyroidectomy (with and without prophylactic or therapeutic node dissection). The search for RCT evidence yielded five studies.[1, 5, 61, 103, 124] Ali[5] compared total thyroidectomy to hemithyroidectomy plus isthmusectomy (both without prophylactic lymph node dissection). Viola[124] compared total thyroidectomy accompanied by prophylactic central lymph node dissection to total thyroidectomy alone. Ahn[1] and Sippel[103] both compared total thyroidectomy accompanied by (non-prophylactic) central lymph node dissection to total thyroidectomy alone. Finally, Kim[61] compared hemithyroidectomy with prophylactic central neck dissection to hemithyroidectomy alone.

A search was also conducted for observational studies evaluating active surveillance against other protocol interventions. This was because no RCTs were available that covered active surveillance, and it had been decided pre-hoc to search for observational studies if no RCTs on active surveillance were found. The drop-down to observational studies for active

surveillance only was to ensure that the health economic model was informed by evidence about active surveillance. The search for observational evidence comparing active surveillance to other protocol interventions yielded three studies.[60, 70, 71] Megwalu[70] compared surgery (either total or hemithyroidectomy) to active surveillance, whilst Jeon[60] compared hemithyroidectomy to active surveillance. In contrast, Moon[71] compared both total and hemithyroidectomy to active surveillance.

Of the eight included studies, the stage of disease was deemed to be stage 1 in five.[1, 61, 70, 71, 103] In the remaining studies, the stage was either unclear or mixed. The review therefore employed two strata: (1) Stage I disease and (2) unclear/mixed stage.

Evidence from these studies is summarised in the clinical evidence summary below (Table 3).

See also the study selection flow chart in Appendix C, study evidence tables in Appendix D, forest plots in Appendix E and GRADE tables in Appendix F.

1.5.2 Excluded studies

See the excluded studies list in Appendix I.

1.1.6 Summary of studies included in the effectiveness evidence

Table 2: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Ali 2011[5] Pakistan RCT	TT (n=30) versus lobectomy with isthmusectomy (n=30) Assume no LN dissection (not described)	Adults with histologically proven differentiated carcinoma of thyroid.	Hoarseness Cancer recurrence	Disease stage unclear
Jeon 2019[60] Korea Observational	Active surveillance (n=43) versus HT (conventional or robotic, n=148) Assume no LN dissection (not described)	Adults with cytologically confirmed papillary microcarcinoma.	Quality of life: Physical component summary Mental component summary Neuromuscular Voice Concentration Sympathetic symptoms Throat/mouth Psychological Sensory Problems with scar Felt chilly Tingling hands/feet Gained weight Headache Less interest in sex	Disease stage unclear
Megwalu 2017[70] USA Observational	Active surveillance (n=15) versus surgery (TT or HT, n=2308) Assume no LN dissection (not described)	Aged >65 years, early-stage papillary thyroid carcinoma <1cm.	Overall survival	Disease stage I (TNM 8 classification)
Viola 2015[124] Italy RCT	TT + prophylactic central node dissection (n=93) versus TT only (n=88)	Adults with FNA-diagnosed papillary carcinoma and no evidence of lymph node or distant metastases.	Persistence of disease at end of study (median follow-up of 5 years). Requirement for additional course(s) of radioactive iodine ablation.	Disease stage mixed

Study	Intervention and comparison	Population	Outcomes	Comments
			Permanent hypoparathyroidism. Recurrent laryngeal nerve palsy.	
Ahn 2022[1] South Korea RCT	Total thyroidectomy with CND (n=56) versus total thyroidectomy only (n=56)	Patients aged 20-70; small non invasive PTC; scheduled to receive total thyroidectomy	Cancer recurrence Recurrent laryngeal nerve palsy Hypoparathyroidism Need for further treatment	Stage 1 - clinically node-negative (cN0) papillary thyroid cancer
Sippel, 2020[103] USA RCT	Total thyroidectomy with CND (n=31) versus total thyroidectomy only (n=30)	Confirmed diagnosis of PTC or a fine needle aspirate (FNA) and/or ultrasound (US) that were suspicious for PTC; between the ages of 21–70,	Quality of life Recurrent laryngeal nerve injury Hypoparathyroidism (in terms of Calcium and PTH levels)	Stage 1 - clinically node-negative (cN0) papillary thyroid cancer
Kim, 2020[61] South Korea RCT	Hemithyroidectomy with pCND (n=94) versus hemithyroidectomy only (n=90)	(1) age between 18 and 70 years, (2) cytologically proven PTMC, (3) no evidence of clinically positive lymph node and (4) acquisition of informed consent from patient.	Cancer recurrence Need for further treatment Recurrent laryngeal nerve injury hypoparathyroidism	Stage 1 - clinically node-negative (cN0) papillary thyroid microcarcinoma (PTMC).
Moon, 2021[71] South Korea Observational study	Active surveillance vs Hemi (lobectomy or isthmusectomy) vs TT	Patients diagnosed with PTMC	Overall health score on the Korean version of the thyroid-specific QoL questionnaire (Dow)	Stage 1 – low risk PTMC

1 See Appendix D for full evidence tables.
2

1.1.7 Summary of the effectiveness evidence

Stratum 1: Stage I disease

Table 3: Clinical evidence summary: thyroid surgery (total thyroidectomy alone or hemithyroidectomy alone) versus active surveillance

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active surveillance	Risk difference between surgery versus active surveillance (95% CI)
Overall mortality	2323 (1 study) 5 years	LOW[1] due to risk of bias	HR 0.11 (0.09 to 0.13)	Not available	RD not calculable

1. *Downgraded due to risk of bias secondary to probable selection bias*

Table 4: Clinical evidence summary: total thyroidectomy versus active surveillance

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active surveillance	Risk difference between surgery versus active surveillance (95% CI)
Quality of life (Korean version of the thyroid-specific QoL questionnaire – overall score). Higher score better.	184 (1 study) 2 years	VERY LOW[1,2] due to risk of bias and imprecision	MD: -0.354 (-0.529 to -0.179)	Not available	RD not calculable

1. *Downgraded due to risk of bias secondary to probable selection bias.*
2. *Downgraded for imprecision on the basis of optimal information size <350. It was not possible to assess on the basis of 0.5 x sd in the control group as such data were not provided in the paper*

Table 5: Clinical evidence summary: Hemithyroidectomy (lobectomy or isthmusectomy) versus active surveillance

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active surveillance	Risk difference between surgery versus active surveillance (95% CI)
Quality of life (Korean version of the thyroid-specific QoL questionnaire – overall score). Higher score better.	184 (1 study) 2 years	VERY LOW[1,2] due to risk of bias and imprecision	MD: -0.141 (-0.248 to -0.141)	Not available	RD not calculable

1. *Downgraded for very serious risk of bias due to selection bias and incomplete outcome data*
2. *Downgraded for imprecision on the basis of optimal information size <350. It was not possible to assess on the basis of 0.5 x sd in the control group as such data were not provided in the paper*

Table 6: Clinical evidence summary: Total thyroidectomy + central neck dissection versus total thyroidectomy alone

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with total thyroidectomy alone	Risk difference between Total thyroidectomy + CND versus total thyroidectomy alone (95% CI)
Cancer recurrence	101 (1) 46 months	VERY LOW[1,2]	RR 0.86 (0.34 to 2.19)	160 per 1000	22 fewer per 1000 (from 106 fewer to 190 more)
Recurrent laryngeal nerve palsy	101 (1) 46 months	VERY LOW[1,2]	RR 1.63 (0.41 to 6.48)	60 per 1000	38 more per 1000 (from 35 fewer to 329 more)
Hypoparathyroidism	101 (1) 46 months	LOW[1,2]	RR 0.53 (0.23 to 1.21)	260 per 1000	122 fewer per 1000 (from 200 fewer to 55 more)

Need for further treatment	101 (1) 46 months	VERY LOW[1,2]	RR 0.98 (0.47 to 2.05)	220 per 1000	4 fewer per 1000 (from 117 fewer to 231 more)
EAT-10 swallowing score (higher worse)	60 (1) 12 months	LOW[1,2]	MD 0.46 (-1.37 to 2.29)	-	The mean EAT-10 swallowing score in the intervention groups was 0.46 higher (1.37 lower to 2.29 higher)
Calcium levels	60 (1) 12 months	LOW[1,2]	MD: 0.1(-0.18 to 0.38)	-	The mean calcium levels in the intervention groups was 0.1 higher (0.18 lower to 0.38 higher)
Parathyroid hormone (PTH) levels	60 (1) 12 months	LOW[1,2]	MD: 1.5 (-11.95 to 14.95)	-	The mean PTH levels in the intervention groups was 1.5 higher (11.95 lower to 14.95 higher)

1. *Downgraded for serious risk of bias due to selection bias secondary to no reports of allocation concealment*
2. *Serious imprecision if the 95% CIs crossed one MID and very serious if they cross two MIDs. For binary outcomes the MIDs were defined as a RR/HR or OR of 0.8 and 1.2, and for continuous outcomes the MIDs were defined as ± half the standard deviation of the control group. For the EAT-10 swallowing score, the MID was ±2.08, based on the control group sd of 4.16. For Calcium levels, the MID was ±0.2738, based on the control group sd of 0.5477. For PTH, the MID was ±12.87, based on the control group sd of 25.87.*

Table 7: Clinical evidence summary: Hemithyroidectomy + pCND versus Hemithyroidectomy alone

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Hemithyroidectomy alone	Risk difference between Hemithyroidectomy + pCND versus Hemithyroidectomy alone (95% CI)
Cancer recurrence	164 (1) 60 months	VERY LOW[1,2]	RR 3 (0.32 to 28.25)	12 per 1000	24 more per 1000 (from 8 fewer to 327 more)
Recurrent laryngeal nerve palsy	164 (1) 60 months	VERY LOW[1,2]	Peto OR 0.14 (0.00 to 6.82)	12 per 1000	10 less per 1000 (from 50 fewer to 20 more)

Hypoparathyroidism	164 (1) 60 months	VERY LOW[1,2]	RD: 0.00 (-0.02 to +0.02)	0 per 1000	0 less per 1000 (from 20 fewer to 20 more)
Need for further treatment	164 (1) 60 months	VERY LOW[1,2]	Peto OR 8.31 (2.32 to 29.73)	0 per 1000	120 more per 1000 (from 50 more to 200 more)

1. *Downgraded for serious risk of bias due to selection bias secondary to no reports of allocation concealment*
2. *Serious imprecision if the95% CIs crossed one MID and very serious if they cross two MIDs. The MIDs were defined as a RR/HR or OR of 0.8 and 1.2. For the outcome with a risk difference, the imprecision was based on the calculation of total information size.*

Stratum 2: Disease stage mixed or unclear

Table 8: Clinical evidence summary: total thyroidectomy alone versus hemithyroidectomy plus isthmusectomy alone

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with HT + isthmusectomy	Risk difference between TT alone versus HT + isthmusectomy (95% CI)
Cancer recurrence	60 (1 study) 3 months	LOW[1] due to risk of bias	Peto OR 0.11 (0.02 to 0.6)	Not available	RD not calculable
Hoarseness	60 (1 study) 3 months	VERY LOW[2,3,4] due to risk of bias, indirectness, imprecision	RR 1 (0.15 to 6.64)	67 per 1000	0 fewer per 1000 (from 57 fewer to 376 more)

[1] *No description of sequence generation or allocation concealment.*
[2] *No description of sequence generation or allocation concealment. No blinding. No description of how the outcome was measured.*
[3] *Outcome was hoarseness whereas the protocol outcome was recurrent laryngeal nerve palsy.*
[4] *Downgraded by 2 increments as the confidence interval crossed both default MIDs. The MIDs were defined as a RR/HR or OR of 0.8 and 1.25.*

Table 9: Clinical evidence summary: total thyroidectomy plus prophylactic central lymph node dissection versus total thyroidectomy alone.

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with TT alone	Risk difference between TT+PCNC versus TT alone (95% CI)
Disease persistence	181 (1 study) 5 years	VERY LOW[1,2] due to risk of bias, imprecision	RR 0.95 (0.35 to 2.59)	80 per 1000	4 fewer per 1000 (from 52 fewer to 126 more)
Need for additional ^{131}I ablation	181 (1 study) 5 years	LOW[1] due to risk of bias	RR 0.19 (0.06 to 0.64)	174 per 1000	141 fewer per 1000 (from 63 fewer to 164 fewer)
Hypoparathyroidism	181 (1 study) 5 years	VERY LOW[1,3] due to risk of bias, imprecision	RR 2.43 (1.07 to 5.54)	80 per 1000	114 more per 1000 (from 6 more to 361 more)
Recurrent laryngeal nerve palsy	181 (1 study) 5 years	VERY LOW[1,2] due to risk of bias, imprecision	RR 0.54 (0.16 to 1.78)	80 per 1000	37 fewer per 1000 (from 67 fewer to 62 more)

[1] *No description of sequence generation or allocation concealment.*
[2] *Downgraded by 2 increments as the confidence interval crossed both default MIDs. The MIDs were defined as a RR/HR or OR of 0.8 and 1.25*
[3] *Downgraded by one increment as the confidence interval crossed one default MID. The MIDs were defined as a RR/HR or OR of 0.8 and 1.25*

Table 10: Clinical evidence summary: hemithyroidectomy alone versus active surveillance

Outcomes	No of Participants Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active surveillance	Risk difference with hemithyroidectomy versus active surveillance (95% CI)
QoL (physical component summary) SF-12. Transformed scale from 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (physical component summary) in the intervention groups was 1.31 lower (3.66 lower to 1.04 higher).
QoL (mental component summary) SF-12. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (mental component summary) in the intervention groups was 1.41 lower (4.16 lower to 1.34 higher)
QoL (neuromuscular) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (neuromuscular) in the intervention groups was 4.99 higher (0.63 to 9.35 higher)
QoL (voice) THYCA-QoL. Scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (voice) in the intervention groups was 3.02 higher (1.95 lower to 7.99 higher)

QoL (concentration) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (concentration) in the intervention groups was 5.25 higher (0.45 lower to 10.95 higher)
QoL (sympathetic symptoms) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (sympathetic symptoms) in the intervention groups was 4.64 higher (1.73 lower to 11.01 higher)
QoL (throat/mouth) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (throat/mouth) in the intervention groups was 5.28 higher (0.18 to 10.38 higher)
QoL (psychological) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	LOW[1] due to risk of bias	Not estimable	Not estimable	The mean qol (psychological) in the intervention groups was 2.29 higher (3.29 lower to 7.87 higher)
QoL (sensory) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	LOW[1] due to risk of bias	Not estimable	Not estimable	The mean qol (sensory) in the intervention groups was 0.4 higher (5.88 lower to 6.68 higher)
QoL (scar) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	LOW[1] due to risk of bias	Not estimable	Not estimable	The mean qol (scar) in the intervention groups was 9.34 higher (4.38 to 14.3 higher)
QoL (felt chilly) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (felt chilly) in the intervention groups was 4.61 higher (1.01 lower to 10.23 higher)

QoL (tingling hands/feet) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (tingling hands/feet) in the intervention groups was 3.27 higher (4.29 lower to 10.83 higher)
QoL (weight gain) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (weight gain) in the intervention groups was 5.25 higher (2.42 lower to 12.92 higher)
QoL (headache) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study)	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (headache) in the intervention groups was 2.22 higher (4.2 lower to 8.64 higher)
QoL (less interest in sex) THYCA-QoL. Transformed scale from: 0 to 100.	191 (1 study) J	VERY LOW[1,2] due to risk of bias, imprecision	Not estimable	Not estimable	The mean qol (less interest in sex) in the intervention groups was 5.81 lower (11.65 lower to 0.03 higher)

[1] *Downgraded for very serious risk of bias due to selection bias and blinding (Group allocation was determined by disease severity criteria. No blinding. Groups dissimilar for baseline characteristics).*
[2] *Downgraded by 1 increment as the confidence interval crossed one default MID. The MIDs were as follows. PCS: ±2.83, based on sd of control group sd of 5.66; MCS: ±4.09, based on sd of control group sd of 8.18; neuromuscular: ±7.03, based on sd of control group sd of 14.07; voice: ±6.41, based on sd of control group sd of 12.82; concentration: ±6.61, based on sd of control group sd of 13.22; sympathetic: ±8.55, based on sd of control group sd of 17.09; throat/mouth: ±5.79, based on sd of control group sd of 11.58; psychological: ±7.92, based on sd of control group sd of 15.84; sensory: ±6.89, based on sd of control group sd of 13.79; PCS: ±2.83, based on sd of control group sd of 5.66; scar: ±0, based on sd of control group sd of 0; Felt chilly: ±5.93, based on sd of control group sd of 11.85; tingling hands and feet: ±9.68, based on sd of control group sd of 19.35; gained weight: ±8.20, based on sd of control group sd of 16.39; PCS: ±2.83, based on sd of control group sd of 5.66; headache: ±7.07, based on sd of control group sd of 14.14; less interest in sex: ±10.9, based on sd of control group sd of 21.79;*

See Appendix F for full GRADE table

1.1.8 Economic evidence

1.1.8.1 Included studies

Three health economic studies with relevant comparisons were included in this review: 1 comparing total thyroidectomy with hemithyroidectomy with or without neck dissection[62]; 2 comparing active surveillance with immediate treatment[67, 76]. These are summarised in the health economic evidence profiles below (**Table 11** and **Table 12: Health economic evidence profile: Active surveillance (AS) vs immediate treatmentTable 12**) and the health economic evidence tables in Appendix H.

1.1.8.2 Excluded studies

No relevant health economic studies were excluded due to assessment of limited applicability or methodological limitations.

See also the health economic study selection flow chart in Appendix G

1.1.9 Summary of included economic evidence

Table 11: Health economic evidence profile: Total thyroidectomy vs hemithyroidectomy (with or without neck dissection)

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Kim 2019[62] ([South Korea])	Partially applicable [(a)]	Potentially serious limitations [(b)]	• Retrospective database analysis • Cost-comparison (no QALYs) • Population: Adults with low-risk, advanced and recurrent differentiated thyroid cancer (papillary carcinoma) • Comparators: 1. Hemithyroidectomy 2. Total thyroidectomy 3. Total thyroidectomy with ipsilateral radical neck dissection 4. Total thyroidectomy with bilateral radical neck dissection and mediastinal dissection • Time horizon: 5 years	Total thyroidectomy costs £ £3,220 [(c)] more than hemithyroidectomy Total thyroidectomy with ipsilateral radical neck dissection costs £3,699 [(c)] than total thyroidectomy Total thyroidectomy with bilateral radical neck dissection and mediastinal dissection costs	NA	NA	No exploration of uncertainty

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
				£31,417 (c) than total thyroidectomy			

Abbreviations: ICER = incremental cost-effectiveness ratio; NA = not applicable; NR = not reported; RCT = randomized controlled trial; QALYs= quality-adjusted life years.

(a) Population included advanced and recurrent thyroid cancer patients who require radical neck dissection or mediastinal dissection and therefore the cost of high-dose RAI therapy should be considered in addition to the cost of surgery and outpatient follow-up. Active surveillance was not included as a comparator as it was not performed at the study site. Patients were retrospectively categorized in each comparator group according to surgical extent which was also determined by disease severity. Korean health system context. Clinical outcomes were not included. Discounting was not reported QALYs were not included.

(b) Retrospective study that only included data from a single hospital of unclear representativeness or generalizability. Surgery choice is not random but depended on disease severity. Resource use and unit cost sources were not reported. Sensitivity analyses were not conducted and parameter uncertainty was not reported.

(c) 2015 South Korean won converted to 2015 UK pounds.[78]. Cost components incorporated: Intervention (surgical admission), outpatient follow-up (outpatient visit, thyroid function test, thyroid ultrasound, neck CT, PET-CT), Radioiodine therapy (131-Iodine therapy, 131-Iodine full body scan).

Table 12: Health economic evidence profile: Active surveillance (AS) vs immediate treatment

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Oda 2017[76] ([Japan])	Partially applicable (a)	Potentially serious limitations (b)	• Patient flow model based on Oda 2016[77764] • Cost-comparison (no QALYs) • Population: Adults with low-risk differentiated thyroid cancer (papillary microcarcinoma) • Comparators: 1. Active surveillance (followed-up by ultrasound and blood tests at 6-months and 1-year)	Immediate surgery costs £4,821(c) more than immediate surgery	NA	NA	No exploration of uncertainty.

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			2. Intervention 2: Immediate surgery (total thyroidectomy with central node dissection or hemithyroidectomy with paratracheal dissection) • Follow-up: 10 years				
Lin 2020[67] ([Australian])	Partially applicable[(e)]	Potentially serious limitations [(f)]	• Retrospective database analysis • Cost-comparison (no QALYs) • Population: Adults with low-risk differentiated thyroid cancer (papillary microcarcinoma) • Comparators: 1. Active surveillance 2. Immediate surgery • Time horizon: 3 years	Immediate surgery costs £3,653[(g)] more than active surveillance	NA	NA	Decreasing the follow-up interval for active surveillance from twice to once a year halved the annual cost of active surveillance. Age has a big impact on the results of the analysis as younger patients have a higher risk of disease progression than older patients and therefore active surveillance is a much more costly strategy for people in their 20s, 30s and 40s than 50s, 60s and 70s).

Abbreviations: ICER = incremental cost-effectiveness ratio; NA = not applicable; NR = not reported; RCT = randomized controlled trial; QALYs= quality-adjusted life years

(a) Japanese healthcare context. Outcomes were not included. Discounting was not applied. QALYs were not included.

(b) Transient and permanent vocal cord paralysis (both potential outcomes of the surgery) were not included in the model. Resource use and unit costs were obtained from one hospital with unclear representativeness or generalizability. Patients chose active surveillance or immediate surgery, so their baseline characteristics are likely unbalanced; descriptive statistics not reported. Cost year was not reported and assumed to be the same as the completion date of clinical trial. Sensitivity analyses were not conducted. Source of funding was not reported.

(c) 2013 Japanese yen converted to 2013 UK pounds.[78]. Cost components incorporated: Initial diagnosis (physician consultation, blood test, ultrasound, fine needle aspiration cytology), surgery (pre-operative examinations, surgery, anaesthesia, pathologic examination, and inpatient stay), follow-up care (physician consultation, blood test, ultrasound), medication (l-thyroxine, vitamin D supplements).

(d) Australian healthcare context. QALYs were not included. Discount rate was not reported.

(e) Outcomes were obtained from a single hospital of unclear representativeness or generalizability. Resource use estimates for active surveillance were based on resource use for the program proposed by Oda 2017[76] because active surveillance was not offered as a treatment option to patients at the Endocrine Surgery Unit at the University of Sydney (therefore they are subject to the same limitations of Oda 2017). Descriptive statistics were not reported. Cost year was not reported and assumed to be the final year of the database analysis.

(f) 2017 Australian dollars converted to 2017 UK pounds.[78]. Cost components incorporated: Initial diagnosis (endocrinologist consultation, blood test, ultrasound, fine needle aspiration cytology), surgery (pre-operative examinations, surgery, anaesthesia, pathologic examination, and inpatient stay), follow-up care (endocrinologist consultation, blood test, ultrasound), medication (l-thyroxine, calcium, and vitamin D supplements).

1.1.10 Economic model

This area was not prioritised for new cost-effectiveness analysis.

1.1.11 Economic evidence statements

One cost-comparison analysis found hemithyroidectomy to be less costly than thyroidectomy with or without neck dissection. The analysis was assessed as partially applicable with potentially serious limitations.

Two cost-comparison analyses found active surveillance to be less costly than immediate intervention. The analyses were assessed as partially applicable with potentially serious limitations.

1.1.12 The committee's discussion and interpretation of the evidence

1.1.12.1 The outcomes that matter most

Protocol-specified outcomes of mortality, quality of life, cost effectiveness, local cancer progression, incidence of distant metastases, cancer recurrence, postoperative dysphagia, recurrent laryngeal nerve palsy, hypoparathyroidism and need for further treatment were all deemed critical and were therefore of equal importance in decision-making.

1.1.12.2 The quality of the evidence

The quality of evidence ranged from very low to low, with most of the downgrading resulting from risk of bias and imprecision. The risk of bias was derived from inadequately described sequence generation and allocation concealment in the randomised controlled trials (RCTs), and from selection bias and incomplete outcome data in the observational studies.

1.1.12.3 Benefits and harms

Three main treatment comparisons were discussed by the committee, and these have been separated below for clarity.

Total thyroidectomy versus hemithyroidectomy

Randomised control trial (RCT) evidence favoured total thyroidectomy over hemithyroidectomy among patients aged 20 and above, in terms of cancer recurrence three months after the initial operation. However, the committee noted that 'recurrence' had not been adequately defined, and that the six events (all occurring in the hemithyroidectomy group) were more likely to represent persistence of disease present but undetected in the contralateral lobe prior to surgery. In addition, the short 3 month follow up was considered inadequate to draw any meaningful conclusions about oncological outcomes. In terms of harms, the evidence suggested that the treatments had similar effects in terms of hoarseness. In the absence of other evidence, and weighing up the available benefits and harms, the committee agreed that total thyroidectomy should be offered over hemithyroidectomy where there were definite indications for post-operative radioiodine (RAI). These include gross extra thyroidal extension and nodal disease. In such patients, where gross extra thyroidal extension or nodal metastases (for example) make RAI vital to prevent further spread or progression, a total thyroidectomy is required because RAI can only be used if all macroscopic thyroid tissue has been removed. Additionally, the committee agreed in the treatment recommendations for RAI (evidence report J) that anyone with an T3 or T4 stage primary tumour or adverse pathological features should be offered RAI and therefore should be offered a total thyroidectomy. However, where the risk of recurrence is lower, the committee agreed that a hemithyroidectomy would be as beneficial and potentially less harmful by leading to lower rates of recurrent laryngeal nerve injury and hypoparathyroidism. It might also allow people to maintain normal thyroid function. Finally, the committee agreed

that those patients having a hemithyroidectomy might later require a completion thyroidectomy if indicated by the histological review or during subsequent surveillance.

Surgery versus active surveillance

Observational evidence for active surveillance showed surgery led to lower overall mortality than active surveillance in patients with stage I disease. A propensity score was used to control for baseline characteristics, however, the committee were aware of the lack of adjustment for likely confounding by comorbidity.

In an observational study of a population with papillary thyroid microcarcinoma (PTMC), active surveillance was shown to lead to a better quality of life than either hemithyroidectomy or total thyroidectomy. Furthermore, observational evidence in another study among adults with PTMC favoured active surveillance over hemithyroidectomy in terms of fewer surgical scar problems, neuromuscular symptoms, loss of interest in sex and throat and mouth symptoms. However, other quality of life outcomes were largely inconclusive. The committee agreed that whilst the evidence base provided relatively good evidence that active surveillance led to better quality of life than surgery in people with PTMC, this could not be extrapolated to people with higher levels of disease.

The committee agreed that the evidence base for active surveillance suggested it should not be used for the majority of patients with thyroid cancer. They agreed that it could be considered in people with a small solitary (<1cm) microcarcinoma because in their experience there is a low risk of the tumour adversely affecting the person's quality of life. This would only apply to solitary microcarcinomas because multifocal microcarcinomas may be associated with greater risk. Therefore, they agreed that either a hemithyroidectomy or active surveillance could be considered for people with a solitary microcarcinoma.

Given the lack of RCT evidence and low quality of the observational data for active surveillance the committee also made a research recommendation comparing active surveillance to surgery.

Surgery for existing nodal disease

Limited evidence was found for treatment of *existing* nodal disease, comparing total thyroidectomy with central neck dissection to total thyroidectomy alone in patients with stage 1 disease. Apart from a trend suggesting that total thyroidectomy with central neck dissection reduced the risk of hypoparathyroidism, which the committee agreed was counterintuitive, there were no differences observed between approaches. The committee therefore drew upon their clinical experience to form recommendations. The committee thought that any nodal disease should be dealt with at the time of the total thyroidectomy to prevent unnecessary delays and reduce the increased risks of harms from repeated surgery. The committee agreed that if nodal disease is present in the lateral neck, a compartment orientated lateral neck dissection should be offered in order to remove the cancer and improve the person's quality of life. They also discussed that carrying out an ipsilateral central neck dissection at the same time may also be of benefit. As the cancer has already spread to the neck carrying out this procedure at the same time may help avoid the need for future surgery. For disease in the central neck the committee agreed that if nodal disease is present a compartment orientated central neck dissection should be offered.

Prophylactic surgery for nodal disease

In terms of a preventative approach to nodal disease, RCT evidence in a sample where the stage of disease was unclear suggested that total thyroidectomy patients undergoing prophylactic central lymph node dissection (PCCND) required fewer additional radioiodine treatments. In terms of harms, those undergoing PCCND had a higher risk of permanent hypoparathyroidism, but evidence was inconclusive in terms of recurrent laryngeal nerve palsy. In low risk (stage 1) patients evidence suggested that prophylactic central neck dissection did not affect cancer recurrence, recurrent laryngeal never injury or

hypoparathyroidism, but might be associated with the need for further treatment. Overall, the committee thought that the evidence of benefit was poor for both strata: for the mixed/unclear stratum, benefits accrued from the requirement for fewer additional ablations were offset by the risks of permanent hypoparathyroidism, whilst for the low risk stratum, there was only evidence of harm. In view of this the committee agreed that prophylactic central lymph node dissection should not be recommended. No evidence was found for prophylactic *lateral* lymph node dissection, but the committee agreed that the harms might exceed those observed for central lymph node dissection alongside no improvement in benefits, and so prophylactic lateral lymph node dissection was also not recommended. The only exception to this is when a lateral lymph node dissection was being carried out to treating

Surgery during pregnancy

The committee used consensus to make recommendations in relation to pregnant women. They agreed that there could be risks to the foetus if operating on pregnant women although this risk is unclear. The concern in the first trimester is largely about preventing birth defects from the anaesthetic drugs. The risk later is about loss of the pregnancy. Therefore, the committee agreed that it would be better to defer any surgical treatment during pregnancy. They also agreed that the pregnant woman should be reassured and that plans for treatment after birth should still be started. However, the committee also noted that in the rare event of there being clinical or radiological evidence of progression (local invasion or regional disease development) then they would consider surgery after discussion with the mother and an obstetrician.

21.12.4 Cost effectiveness and resource use

Three health economics studies were included for this review question. One was on thyroidectomy versus hemithyroidectomy while the other two were on active surveillance versus surgery.

A cost comparison analysis compared hemithyroidectomy and total thyroidectomy using a retrospective database. The analysis was assessed partially applicable and with potential serious limitation as patients were retrospectively categorized according to the surgical extent, which is determined by disease severity, and therefore likely to be significantly different in the baseline. The analysis found patients undergoing hemithyroidectomy to be associated with the lowest cost and patients undergoing thyroidectomy with ipsilateral or bilateral neck dissection to be associated with the highest costs.

Two other cost comparison analyses compared active surveillance with immediate surgery and were assessed to be partially applicable with potential serious limitations as outcomes were taken from a single hospital with unclear representability and the enrolled patients chose themselves whether to receive immediate surgery or active surveillance. The analyses concluded that active surveillance costs less than immediate surgery in people with low-risk differentiated thyroid cancer.

The committee noted that the economics studies were based on observational studies with no control for differences between people receiving the different treatments. In the cost comparison of hemi and total thyroidectomy, patients were categorized retrospectively based on the extent of the surgery they received. It was noted that people undergoing hemithyroidectomy generally have a lower disease severity than people receiving total thyroidectomy with or without neck dissection, as the latter have usually advanced or recurrent thyroid cancer. Therefore, it was already expected that people undergoing hemi-thyroidectomy had a lower healthcare cost as their disease was likely to be milder compared to people undergoing more complex procedures. The committee concluded that the study failed to estimate the real cost differences between the interventions and that was poorly informative. The committee were aware that there is an ongoing RCT comparing hemithyroidectomy versus total thyroidectomy (HoT). The trial will be the first randomised controlled trial on hemi and total thyroidectomy and, as such, will not suffer from selection

bias common in observational studies. Future health economics analyses based on the HoT trial will hopefully shed light on the cost effectiveness of hemi-thyroidectomy and full thyroidectomy for people with low-risk thyroid cancer. Due to the limitation of available evidence, the committee made recommendations on thyroidectomy in line with their experience and expertise. Total thyroidectomy should be offered to people showing clear sign of needing post-operative RAI. In the other cases, a choice of total or hemi-thyroidectomy was recommended. This reflects current practice where total or hemi-thyroidectomy are offered based on cancer characteristics and predicted need of post-surgery treatments and therefore, these recommendations are unlikely to change practice or require additional NHS resource.

Likewise, economic evidence on active surveillance were not robust. The two health economics studies were based on an observational study on a single hospital, where people chose whether undergoing immediate surgery or be in active surveillance. Individual choice is clearly an endogenous variable correlated with age, gender, health condition, social status and lifestyle; all these characteristics are expected to affect the outcomes but were not controlled in either study. This casts doubts on the validity of the results of the studies which found active surveillance to be cheaper. It is possible, for instance, that the patients who chose active surveillance are healthier or younger and therefore associated with a lower cost regardless of the intervention. This was impossible to assess as descriptive statistics were not provided in any of the two studies. Unfortunately, no randomized data is available on active surveillance and the few controlled clinical evidence included were assessed to be of poor quality by the committee. The committee concluded that the evidence was insufficient to make any positive recommendation on active surveillance and decided offer recommendation for surgery limiting active surveillance to only in a small group of low-risk cases taking into account patient's preferences. This reflects current practice of surgery being the standard of care and is also consistent with the observational evidence collected where active surveillance was offered only as a personal choice of the patients. Therefore, this recommendation is not expected to cause any additional use of NHS resources.

There was no clinical evidence for treatment of existing nodal disease, so the committee made a recommendation to offer central and/or lateral neck dissection during the thyroidectomy drawing from their clinical experience. The strong recommendation reflects the lack of alternative treatments for people with nodal disease and highlights the importance of removing all the cancerous tissue in a single surgery when possible. This is likely to be cost-effective as it prevents further surgeries on people with evidence of persistent disease while improving quality of life.

1.1.12.5 Other factors the committee took into account

The equality considerations for this recommendation related to pregnant women and surgery are discussed in section 1.1.12.3 benefits and harms.

In writing the recommendations, the committee was mindful to avoid pre-empting the results of an upcoming trial of hemithyroidectomy versus total thyroidectomy (the HoT trial). The committee felt it was important to define active surveillance. It was agreed that active surveillance involves monitoring the person's thyroid cancer with periodic appointments that include investigations such as blood tests and ultrasound.

1.1.13 Recommendations supported by this evidence review

This evidence review supports recommendations 1.3.1 to 1.3.10 and the research recommendation on active surveillance compared with surgery.

References

1. Ahn JH, Kwak JH, Yoon SG, Yi JW, Yu HW, Kwon H et al. A prospective randomized controlled trial to assess the efficacy and safety of prophylactic central compartment lymph node dissection in papillary thyroid carcinoma. Surgery. 2022; 171(1):182-189

2. Ahn SH, Kim WS. The effect of prophylactic central neck dissection during hemithyroidectomy on locoregional recurrence in patients with papillary thyroid carcinoma: A meta-analysis. Clinical and Experimental Otorhinolaryngology. 2020; Epub

3. Alabdrabalnabi HA, Bajafar OA, Alsuwaidan MF, Shaikh AA, Alsayed FA, Alshammari YT et al. An overview on thyroid cancer diagnosis and management approach, literature review. International Journal of Pharmaceutical and Phytopharmacological Research. 2020; 10(6):47-50

4. Alhashemi A, Goldstein DP, Sawka AM. A systematic review of primary active surveillance management of low-risk papillary carcinoma. Current Opinion in Oncology. 2016; 28(1):11-17

5. Ali I, Muhammad G, Ashraf M, Tariq AR, Khalid RS, Abid J. Comparison of total thyroidectomy and lobectomy with isthmusectomy in treatment of well-differentiated carcinoma of thyroid gland. Pakistan Journal of Medical and Health Sciences. 2011; 5(1):113-115

6. Altedlawi Albalawi IA, 2nd, Altidlawi AI, Mirghani H. Radioactive iodine following total thyroidectomy is comparable to lobectomy in low/intermediate-risk differentiated thyroid carcinoma: A meta-analysis. Cureus. 2020; 12(12):e12332

7. Attene F, Pisano IP, Pala C, Marrosu A, Bozzo C, Scognamillo F. Surgical management of primary and recurrent carcinoma showing thymous-like elements (CASTLE). Annali Italiani di Chirurgia. 2017; 88:247-252

8. Bai B, Chen Z, Chen W. Risk factors and outcomes of incidental parathyroidectomy in thyroidectomy: A systematic review and meta-analysis. PLoS ONE [Electronic Resource]. 2018; 13(11):e0207088

9. Bojoga A, Koot A, Bonenkamp J, de Wilt J, IntHout J, Stalmeier P et al. The impact of the extent of surgery on the long-term outcomes of patients with low-risk differentiated non-medullary thyroid cancer: A systematic meta-analysis. Journal of Clinical Medicine. 2020; 9(7):21

10. Bononi M, Tocchi A, Cangemi V, Vecchione A, Giovagnoli MR, De Cesare A et al. Lymph node dissection in papillary or follicular thyroid carcinoma. Anticancer Research. 2004; 24(4):2439-2442

11. Burch HB. Evaluation and management of the solid thyroid nodule. Endocrinology and Metabolism Clinics of North America. 1995; 24(4):663-710

12. Cabrera RN, Chone CT, Zantut-Wittmann D, Matos P, Ferreira DM, Pereira PSG et al. Value of sentinel lymph node biopsy in papillary thyroid cancer: initial results of a prospective trial. European Archives of Oto-Rhino-Laryngology. 2015; 272(4):971-979

13. Caglia P, Puglisi S, Buffone A, Bianco SL, Okatyeva V, Veroux M et al. Post-thyroidectomy hypoparathyroidism, what should we keep in mind? Annali Italiani di Chirurgia. 2017; 6:371-381

14. Caglia P, Zappulla E, Costa S, Tracia M, Veroux M, Russo V et al. Differentiated thyroid cancer: role of the lymph node dissection. Giornale di Chirurgia. 2010; 31(6-7):293-295

15. Carling T, Long WD, 3rd, Udelsman R. Controversy surrounding the role for routine central lymph node dissection for differentiated thyroid cancer. Current Opinion in Oncology. 2010; 22(1):30-34

16. Carling T, Ocal IT, Udelsman R. Special variants of differentiated thyroid cancer: does it alter the extent of surgery versus well-differentiated thyroid cancer? World Journal of Surgery. 2007; 31(5):916-923

17. Caron NR, Clark OH. Papillary thyroid cancer. Current Treatment Options in Oncology. 2006; 7(4):309-319

18. Chan S, Karamali K, Kolodziejczyk A, Oikonomou G, Watkinson J, Paleri V et al. Systematic review of recurrence rate after hemithyroidectomy for low-risk well-differentiated thyroid cancer. European Thyroid Journal. 2020; 9(2):73-84

19. Chaukar DA, Deshmukh AD, Dandekar MR. Management of thyroid cancers. Indian Journal of Surgical Oncology. 2010; 1(2):151-162

20. Chen L, Wu YH, Lee CH, Chen HA, Loh EW, Tam KW. Prophylactic central neck dissection for papillary thyroid carcinoma with clinically uninvolved central neck lymph nodes: A systematic review and meta-analysis. World Journal of Surgery. 2018; 42(9):2846-2857

21. Chisholm EJ, Kulinskaya E, Tolley NS. Systematic review and meta-analysis of the adverse effects of thyroidectomy combined with central neck dissection as compared with thyroidectomy alone. Laryngoscope. 2009; 119(6):1135-1139

22. Cho SJ, Suh CH, Baek JH, Chung SR, Choi YJ, Chung KW et al. Active surveillance for small papillary thyroid cancer: A systematic review and meta-analysis. Thyroid. 2019; 29(10):1399-1408

23. Cohen O, Tzelnick S, Lahav Y, Schindel D, Halperin D, Yehuda M. Selection of atypia/follicular lesion of unknown significance patients for surgery versus active surveillance, without using genetic testing: A single institute experience, prospective analysis, and recommendations. Thyroid. 2017; 27(7):928-935

24. Colombo C, De Leo S, Di Stefano M, Trevisan M, Moneta C, Vicentini L et al. Total thyroidectomy versus lobectomy for thyroid cancer: Single-center data and literature review. Annals of Surgical Oncology. 2021; 28(8):4334-4344

25. Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ et al. Management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid. 2006; 16(2):109-141

26. De Ceulaer J, De Clercq C, Swennen GR. Robotic surgery in oral and maxillofacial, craniofacial and head and neck surgery: a systematic review of the literature. International Journal of Oral and Maxillofacial Surgery. 2012; 41(11):1311-1324

27. Ding B, Yu JF, Sun W, Ma NF. Surgical safety analysis of retaining the glands in papillary thyroid microcarcinoma. European Review for Medical and Pharmacological Sciences. 2017; 21(2):234-238

28. Dionigi G, Kraimps JL, Schmid KW, Hermann M, Sheu-Grabellus SY, De Wailly P et al. Minimally invasive follicular thyroid cancer (MIFTC)--a consensus report of the European Society of Endocrine Surgeons (ESES). Langenbecks Archives of Surgery. 2014; 399(2):165-184

29. Dong L, Wang X, Wang Q, Cai H, Zhang Y, Wang S et al. Total thyroidectomy versus lobectomy for recurrence and complications of papillary thyroid microcarcinoma: A meta-analysis. International Journal of Clinical and Experimental Medicine. 2018; 11(1):23-32

30. El-Labban GM. Minimally invasive video-assisted thyroidectomy versus conventional thyroidectomy: A single-blinded, randomized controlled clinical trial. Journal of Minimal Access Surgery. 2009; 5(4):97-102

31. Fan LJ, Jiang J. Present and future of robot-assisted endoscopic thyroid surgery. Chinese Medical Journal. 2012; 125(5):926-931

32. Friedman M, Ibrahim H. Total versus subtotal thyroidectomy: Arguments, approaches, and recommendations. Operative Techniques in Otolaryngology - Head and Neck Surgery. 2002; 13(3):196-202

33. Friedman M, Pacella BL, Jr. Total versus subtotal thyroidectomy. Arguments, approaches, and recommendations. Otolaryngologic Clinics of North America. 1990; 23(3):413-427

34. Fu H, Cheng L, Jin Y, Chen L. Thyrotoxicosis with concomitant thyroid cancer. Endocrine-Related Cancer. 2019; 26(7):R395-R413

35. Gambale C, Elisei R, Matrone A. Management and follow-up of differentiated thyroid cancer not submitted to radioiodine treatment: a systematic review. Minerva Endocrinologica. 2020; 45(4):306-317

36. Gambardella C, Tartaglia E, Nunziata A, Izzo G, Siciliano G, Cavallo F et al. Clinical significance of prophylactic central compartment neck dissection in the treatment of clinically node-negative papillary thyroid cancer patients. World Journal of Surgical Oncology. 2016; 14(1):247

37. Gartland RM, Lubitz CC. Impact of extent of surgery on tumor recurrence and survival for papillary thyroid cancer patients. Annals of Surgical Oncology. 2018; 25(9):2520-2525

38. Geramizadeh B, Maleki Z. Non-invasive follicular thyroid neoplasm with papillary-like nuclearfeatures (NIFTP): a review and update. Endocrine. 2019; 64(3):433-440

39. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L et al. American association of Clinical Endocrinologists, American college of endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules - 2016 update. Endocrine Practice. 2016; 22(Suppl 1):A001

40. Giuffrida D, Prestifilippo A, Scarfia A, Martino D, Marchisotta S. New treatment in advanced thyroid cancer. Journal of Oncology Print. 2012:391629

41. Grani G, Sponziello M, Pecce V, Ramundo V, Durante C. Contemporary thyroid nodule evaluation and management. Journal of Clinical Endocrinology and Metabolism. 2020; 105(9):2869-2883

42. Grant CS. Papillary thyroid cancer: strategies for optimal individualized surgical management. Clinical Therapeutics. 2014; 36(7):1117-1126

43. Griffin A, Brito JP, Bahl M, Hoang JK. Applying criteria of active surveillance to low-risk papillary thyroid cancer over a decade: How many surgeries and complications can be avoided? Thyroid. 2017; 27(4):518-523

44. Grossman RF, Clark OH. Hurthle Cell Carcinoma. Cancer Control. 1997; 4(1):13-17

45. Guo MY, Wiseman JJ, Wiseman SM. Current surgical treatment of intermediate risk differentiated thyroid cancer: a systematic review. Expert Review of Anticancer Therapy. 2020:1-16

46. Hall SF, Irish J, Groome P, Griffiths R, Hurlbut D. Do lower-risk thyroid cancer patients who live in regions with more aggressive treatments have better outcomes? Thyroid. 2017; 27(10):1246-1257

47. Hassanain M, Wexler M. Conservative management of well-differentiated thyroid cancer. Canadian Journal of Surgery. 2010; 53(2):109-118

48. Hewitt J, Srivatsa S. Management of thyroid nodules. Journal of Clinical Outcomes Management. 2006; 13(8):456-462

49. Hu YL, Cao XY, Zhou YR, Ye XH, Wang JX, Li X et al. Management of sonographically suspicious thyroid nodules 1 cm or smaller and candidacy for active surveillance: Experience of a tertiary center in china. Endocrine Practice. 2021; 27(9):903-911

50. Huang TW, Lai JH, Wu MY, Chen SL, Wu CH, Tam KW. Systematic review of clinical practice guidelines in the diagnosis and management of thyroid nodules and cancer. BMC Medicine. 2013; 11(1):191

51. Hughes DT, Doherty GM. Central neck dissection for papillary thyroid cancer. Cancer Control. 2011; 18(2):83-88

52. Hughes DT, Rosen JE, Evans DB, Grubbs E, Wang TS, Solorzano CC. Prophylactic central compartment neck dissection in papillary thyroid cancer and effect on locoregional recurrence. Annals of Surgical Oncology. 2018; 25(9):2526-2534

53. Husson O, Haak HR, Oranje WA, Mols F, Reemst PHM, Van De Poll-Franse LV. Health-related quality of life among thyroid cancer survivors: A systematic review. Clinical Endocrinology. 2011; 75(4):544-554

54. Ito Y, Miyauchi A. Active surveillance of low-risk papillary thyroid microcarcinomas in Japan and other countries: a review. Expert Review of Endocrinology & Metabolism. 2020; 15(1):5-12

55. Ito Y, Miyauchi A, Inoue H, Fukushima M, Kihara M, Higashiyama T et al. An observational trial for papillary thyroid microcarcinoma in Japanese patients. World Journal of Surgery. 2010; 34(1):28-35

56. Ito Y, Miyauchi A, Oda H. Low-risk papillary microcarcinoma of the thyroid: A review of active surveillance trials. European Journal of Surgical Oncology. 2018; 44(3):307-315

57. Ito Y, Tsushima Y, Masuoka H, Yabuta T, Fukushima M, Inoue H et al. Significance of prophylactic modified radical neck dissection for patients with low-risk papillary thyroid carcinoma measuring 1.1-3.0 cm: first report of a trial at Kuma Hospital. Surgery Today. 2011; 41(11):1486-1491

58. Ito Y, Uruno T, Nakano K, Takamura Y, Miya A, Kobayashi K et al. An observation trial without surgical treatment in patients with papillary microcarcinoma of the thyroid. Thyroid. 2003; 13(4):381-387

59. Jackson NR, Yao L, Tufano RP, Kandil EH. Safety of robotic thyroidectomy approaches: meta-analysis and systematic review. Head and Neck. 2014; 36(1):137-143

60. Jeon MJ, Lee YM, Sung TY, Han M, Shin YW, Kim WG et al. Quality of life in patients with papillary thyroid microcarcinoma managed by active surveillance or lobectomy: A cross-sectional study. Thyroid. 2019; 29(7):956-962

61. Kim BY, Choi N, Kim SW, Jeong HS, Chung MK, Son YI. Randomized trial of prophylactic ipsilateral central lymph node dissection in patients with clinically node negative papillary thyroid microcarcinoma. European Archives of Oto-Rhino-Laryngology. 2020; 277(2):569-576

62. Kim SY, Kim SM, Chang H, Kim BW, Lee YS, Kwon SS et al. Cost for treatment and follow-up of thyroid cancer increases according to the severity of disease. Head and Neck. 2019; 41(7):2376-2379

63. Kong SH, Ryu J, Kim MJ, Cho SW, Song YS, Yi KH et al. Longitudinal assessment of quality of life according to treatment options in low-risk papillary thyroid microcarcinoma patients: Active surveillance or immediate surgery (interim analysis of maestro). Thyroid. 2019; 29(8):1089-1096

64. Kuo EJ, Wu JX, Li N, Zanocco KA, Yeh MW, Livhits MJ. Nonoperative management of differentiated thyroid cancer in california: A population-level analysis of 29,978 patients. Endocrine Practice. 2017; 23(10):1262-1269

65. Lee DH, Kim YK, Yu HW, Choi JY, Park SY, Moon JH. Computed tomography for detecting cervical lymph node metastasis in patients who have papillary thyroid microcarcinoma with tumor characteristics appropriate for active surveillance. Thyroid. 2019; 29(11):1653-1659

66. Li Y, Liu Y, Huang Y, Liu J, Chu J. Total versus subtotal thyroidectomy for differentiated thyroid carcinoma and their influence on related indexes. International Journal of Clinical and Experimental Medicine. 2020; 13(10):8007-8013

67. Lin JF, Jonker PKC, Cunich M, Sidhu SB, Delbridge LW, Glover AR et al. Surgery alone for papillary thyroid microcarcinoma is less costly and more effective than long term active surveillance. Surgery. 2020; 167(1):110-116

68. Lin JK, Sakoda LC, Darbinian J, Socarras M, Chiao W, Calixto N et al. Risk of mortality between untreated and treated papillary thyroid cancer: A matched cohort analysis. Annals of Otology, Rhinology and Laryngology. 2020; 129(3):265-272

69. Liu Z, Zeng W, Zhou L, Zhou W, Chen D, Feng H et al. Active surveillance for young patients with insular thyroid cancer: An initial and novel finding. American Journal of Translational Research. 2019; 11(1):176-187

70. Megwalu UC. Observation versus thyroidectomy for papillary thyroid microcarcinoma in the elderly. Journal of Laryngology and Otology. 2017; 131(2):173-176

71. Moon JH, Ryu CH, Cho SW, Choi JY, Chung EJ, Hah JH et al. Effect of initial treatment choice on 2-year quality of life in patients with low-risk papillary thyroid microcarcinoma. Journal of Clinical Endocrinology and Metabolism. 2021; 106(3):724-735

72. Nagarkatti SS, Faquin WC, Lubitz CC, Garcia DM, Barbesino G, Ross DS et al. Management of thyroid nodules with atypical cytology on fine-needle aspiration biopsy. Annals of Surgical Oncology. 2013; 20(1):60-65

73. Nakamura T, Miyauchi A, Ito Y, Ito M, Kudo T, Tanaka M et al. Quality of life in patients with low-risk papillary thyroid microcarcinoma: Active surveillance versus immediate surgery. Endocrine Practice. 2020; 26(12):1451-1457

74. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

75. Neagoe RM, Cvasciuc IT, Muresan M, Sala DT. Incidental parathyroidectomy during thyroid surgery - risk, prevention and controversies; an evidence-based review. Acta Endocrinologica. 2017; 13(4):467-475

76. Oda H, Miyauchi A, Ito Y, Sasai H, Masuoka H, Yabuta T et al. Comparison of the costs of active surveillance and immediate surgery in the management of low-risk papillary microcarcinoma of the thyroid. Endocrine Journal. 2017; 64(1):59-64

77. Oda H, Miyauchi A, Ito Y, Yoshioka K, Nakayama A, Sasai H et al. Incidences of unfavorable events in the management of low-risk papillary microcarcinoma of the thyroid by active surveillance versus immediate surgery. Thyroid. 2016; 26(1):150-155

78. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2021. Available from: http://www.oecd.org/std/ppp Last accessed: 24/03/2022.

79. Pan JH, Zhou H, Zhao XX, Ding H, Wei L, Qin L et al. Robotic thyroidectomy versus conventional open thyroidectomy for thyroid cancer: a systematic review and meta-analysis. Surgical Endoscopy. 2017; 31(10):3985-4001

80. Parker WA, Edafe O, Balasubramanian SP. Long-term treatment-related morbidity in differentiated thyroid cancer: a systematic review of the literature. Pragmatic & Observational Research. 2017; 8:57-67

81. Paschke R, Lincke T, Muller SP, Kreissl MC, Dralle H, Fassnacht M. The treatment of well-differentiated thyroid carcinoma. Deutsches Arzteblatt International. 2015; 112(26):452-458

82. Pisanu A, Podda M, Reccia I, Porceddu G, Uccheddu A. Systematic review with meta-analysis of prospective randomized trials comparing minimally invasive video-assisted thyroidectomy (MIVAT) and conventional thyroidectomy (CT). Langenbecks Archives of Surgery. 2013; 398(8):1057-1068

83. Qu H, Sun GR, Liu Y, He QS. Clinical risk factors for central lymph node metastasis in papillary thyroid carcinoma: A systematic review and meta-analysis. Clinical Endocrinology. 2015; 83(1):124-132

84. Qu N, Zhang L, Lu ZW, Ji QH, Yang SW, Wei WJ et al. Predictive factors for recurrence of differentiated thyroid cancer in patients under 21 years of age and a meta-analysis of the current literature. Tumour Biology. 2016; 37(6):7797-7808

85. Raffaelli M, De Crea C, Sessa L, Giustacchini P, Revelli L, Bellantone C et al. Prospective evaluation of total thyroidectomy versus ipsilateral versus bilateral central neck dissection in patients with clinically node-negative papillary thyroid carcinoma. Surgery (United States). 2012; 152(6):957-964

86. Ren X, Dai Z, Sha H, Wu J, Hong X, Xiu Z. Comparative study of endoscopic thyroidectomy via a breast approach versus conventional open thyroidectomy in papillary thyroid microcarcinoma patients. Biomedical Research (India). 2017; 28(12):5315-5320

87. Rodriguez-Martin AM, Zacharopoulou P, Hassan AB, Tsiachristas A. Cost-effectiveness of healthcare interventions for rare cancers: Evidence from a systematic literature review and meta-analysis. Journal of Cancer Policy. 2018; 18:1-10

88. Rosario PW, Mourao GF, Calsolari MR. Active surveillance in adults with low-risk papillary thyroid microcarcinomas: A prospective study. Hormone and Metabolic Research. 2019; 51(11):703-708

89. Roti E, degli Uberti EC, Bondanelli M, Braverman LE. Thyroid papillary microcarcinoma: A descriptive and meta-analysis study. European Journal of Endocrinology. 2008; 159(6):659-673

90. Ruggiero FP, Fedok FG. Outcomes in reoperative thyroid cancer. Otolaryngologic Clinics of North America. 2008; 41(6):1261-1268, xii

91. Sakai T, Sugitani I, Ebina A, Fukuoka O, Toda K, Mitani H et al. Active surveillance for t1bn0m0 papillary thyroid carcinoma. Thyroid. 2019; 29(1):59-63

92. Sanabria A, Kowalski LP, Nixon IJ, Simo R. Microscopic positive surgical margins in thyroid carcinoma: a proposal for thyroid oncology teams. Langenbecks Archives of Surgery. 2021; 406(3):563-569

93. Saravana-Bawan B, Bajwa A, Paterson J, McMullen T. Active surveillance of low-risk papillary thyroid cancer: A meta-analysis. Surgery. 2020; 167(1):46-55

94. Schmidbauer B, Menhart K, Hellwig D, Grosse J. Differentiated thyroid cancer-treatment: State of the art. International Journal of Molecular Sciences. 2017; 18(6):1292

95. Sgourakis G, Sotiropoulos GC, Neuhauser M, Musholt TJ, Karaliotas C, Lang H. Comparison between minimally invasive video-assisted thyroidectomy and conventional thyroidectomy: is there any evidence-based information? Thyroid. 2008; 18(7):721-727

96. Shan CX, Zhang W, Jiang DZ, Zheng XM, Liu S, Qiu M. Routine central neck dissection in differentiated thyroid carcinoma: a systematic review and meta-analysis. Laryngoscope. 2012; 122(4):797-804

97. Shan L, Liu J. Meta-analysis comparison of bilateral axillo-breast approach robotic thyroidectomy and conventional thyroidectomy. Surgical Innovation. 2019; 26(1):112-123

98. Shan YZ, Zhou LM, Yu ZF, Wang SG, Gao GL, Shen Y et al. Comparison between transareola singlesite endoscopic thyroidectomy and minimally invasive video-assisted thyroidectomy. Journal of International Medical Research. 2012; 40(6):2213-2219

99. Shen H, Shan C, Qiu M. Systematic review and meta-analysis of transaxillary robotic thyroidectomy versus open thyroidectomy. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques. 2014; 24(3):199-206

100. Sieda B, Tawfik MM, Khatur H. Is routine dissection of central lymph node and radio-active iodine therapy, necessary for papillary thyroid carcinoma, T1-2 N0? A randomized controlled trial. International Journal of Surgery Open. 2020; 24:117-124

101. Singer PA, Cooper DS, Daniels GH, Ladenson PW, Greenspan FS, Levy EG et al. Treatment guidelines for patients with thyroid nodules and well- differentiated thyroid cancer. Archives of Internal Medicine. 1996; 156(19):2165-2172

102. Sipos JA, Mazzaferri EL. The therapeutic management of differentiated thyroid cancer. Expert Opinion on Pharmacotherapy. 2008; 9(15):2627-2637

103. Sippel RS, Robbins SE, Poehls JL, Pitt SC, Chen H, Leverson G et al. A randomized controlled clinical trial: No clear benefit to prophylactic central neck dissection in patients with clinically node negative papillary thyroid cancer. Annals of Surgery. 2020; 272(3):496-503

104. Smulever A, Pitoia F. Active surveillance in papillary thyroid carcinoma: not easily accepted but possible in Latin America. Archives of Endocrinology & Metabolism. 2019; 63(5):462-469

105. Son SK, Kim JH, Bae JS, Lee SH. Surgical safety and oncologic effectiveness in robotic versus conventional open thyroidectomy in thyroid cancer: a systematic review and meta-analysis. Annals of Surgical Oncology. 2015; 22(9):3022-3032

106. Sonkar AA, Rajamanickam S, Singh D. Papillary thyroid carcinoma: debate at rest. Indian Journal of Cancer. 2010; 47(2):206-216

107. Su H, Li Y. Factors related to lymph node recurrence in patients with n1b papillary thyroid carcinoma after unilateral therapeutic modified radical neck dissection: A meta-analysis. Zentralblatt für Chirurgie. 2018; 143(4):373-379

108. Su H, Li Y. Prophylactic central neck dissection and local recurrence in papillary thyroid microcarcinoma: a meta-analysis. Revista Brasileira de Otorrinolaringologia. 2019; 85(2):237-243

109. Sugitani I, Ito Y, Miyauchi A, Imai T, Suzuki S. Active surveillance versus immediate surgery: Questionnaire survey on the current treatment strategy for adult patients with low-risk papillary thyroid microcarcinoma in japan. Thyroid. 2019; 29(11):1563-1571

110. Sun GH, Peress L, Pynnonen MA. Systematic review and meta-analysis of robotic vs conventional thyroidectomy approaches for thyroid disease. Otolaryngology - Head & Neck Surgery. 2014; 150(4):520-532

111. Sun W, Lan X, Zhang H, Dong W, Wang Z, He L et al. Risk factors for central lymph node metastasis in cn0 papillary thyroid carcinoma: A systematic review and meta-analysis. PLoS ONE [Electronic Resource]. 2015; 10(10):e0139021

112. Sun W, Zheng B, Wang Z, Dong W, Qin Y, Zhang H. Meta-analysis of risk factors for CCLNM in patients with unilateral cN0 PTC. Endocrine Connections. 2020; 9(5):387-395

113. Sywak M, Pasieka JL, Ogilvie T. A review of thyroid cancer with intermediate differentiation. Journal of Surgical Oncology. 2004; 86(1):44-54

114. Sywak MS, Yeh MW, McMullen T, Stalberg P, Low H, Alvarado R et al. A randomized controlled trial of minimally invasive thyroidectomy using the lateral direct approach versus conventional hemithyroidectomy. Surgery. 2008; 144(6):1016-1021; discussion 1021-1012

115. Tan CT, Cheah WK, Delbridge L. "Scarless" (in the neck) endoscopic thyroidectomy (SET): an evidence-based review of published techniques. World Journal of Surgery. 2008; 32(7):1349-1357

116. Tunca F, Giles Y, Terzioglu T, Mudun A, Adalet I, Salmaslioglu A et al. Does intraoperative radioguided surgery influence the complication rates and completeness of completion thyroidectomy? American Journal of Surgery. 2008; 196(1):40-46

117. Tuttle RM, Zhang L, Shaha A. A clinical framework to facilitate selection of patients with differentiated thyroid cancer for active surveillance or less aggressive initial surgical management. Expert Review of Endocrinology & Metabolism. 2018; 13(2):77-85

118. Udelsman R, Lakatos E, Ladenson P. Optimal surgery for papillary thyroid carcinoma. World Journal of Surgery. 1996; 20(1):88-93

119. van Gerwen M, Alsen M, Lee E, Sinclair C, Genden E, Taioli E. Recurrence-free survival after total thyroidectomy and lobectomy in patients with papillary thyroid microcarcinoma. Journal of Endocrinological Investigation. 2021; 44(4):725-734

120. Vargas-Pinto S, Romero Arenas MA. Lobectomy compared to total thyroidectomy for low-risk papillary thyroid cancer: A systematic review. Journal of Surgical Research. 2019; 242:244-251

121. Vasileiadis I, Boutzios G, Karalaki M, Misiakos E, Karatzas T. Papillary thyroid carcinoma of the isthmus: Total thyroidectomy or isthmusectomy? American Journal of Surgery. 2018; 216(1):135-139

122. Venkat R, Guerrero MA. Recent advances in the surgical treatment of differentiated thyroid cancer: a comprehensive review. Thescientificworldjournal. 2013:425136

123. Venkatesh S, Pasternak JD, Beninato T, Drake FT, Kluijfhout WP, Liu C et al. Cost-effectiveness of active surveillance versus hemithyroidectomy for micropapillary thyroid cancer. Surgery. 2017; 161(1):116-126

124. Viola D, Materazzi G, Valerio L, Molinaro E, Agate L, Faviana P et al. Prophylactic central compartment lymph node dissection in papillary thyroid carcinoma: clinical implications derived from the first prospective randomized controlled single institution study. Journal of Clinical Endocrinology and Metabolism. 2015; 100(4):1316-1324

125. Vuong HG, Tran TTK, Bychkov A, Jung CK, Nakazawa T, Kakudo K et al. Clinical impact of non-invasive follicular thyroid neoplasm with papillary-like nuclear features on the risk of malignancy in the Bethesda System for reporting thyroid cytopathology: A meta-analysis of 14,153 resected thyroid nodules. Endocrine Practice. 2019; 25(5):491-502

126. Walgama E, Sacks WL, Ho AS. Papillary thyroid microcarcinoma: optimal management versus overtreatment. Current Opinion in Oncology. 2020; 32(1):1-6

127. Wang LY, Roman BR, Migliacci JC, Palmer FL, Tuttle RM, Shaha AR et al. Cost-effectiveness analysis of papillary thyroid cancer surveillance. Cancer. 2015; 121(23):4132-4140

128. Wang TS, Cheung K, Farrokhyar F, Roman SA, Sosa JA. A meta-analysis of the effect of prophylactic central compartment neck dissection on locoregional recurrence rates in patients with papillary thyroid cancer. Annals of Surgical Oncology. 2013; 20(11):3477-3483

129. Wang Y, Liu K, Xiong J, Zhu J. Total endoscopic versus conventional open thyroidectomy for papillary thyroid microcarcinoma. Journal of Craniofacial Surgery. 2015; 26(2):464-468

130. Wang Y, Zhou S, Liu X, Rui S, Li Z, Zhu J et al. Transoral endoscopic thyroidectomy vestibular approach vs conventional open thyroidectomy: Meta-analysis. Head and Neck. 2021; 43(1):345-353

131. Wang YC, Liu K, Xiong JJ, Zhu JQ. Robotic thyroidectomy versus conventional open thyroidectomy for differentiated thyroid cancer: meta-analysis. Journal of Laryngology and Otology. 2015; 129(6):558-567

132. White ML, Doherty GM. Level VI lymph node dissection for papillary thyroid cancer. Minerva Chirurgica. 2007; 62(5):383-393

133. White ML, Gauger PG, Doherty GM. Central lymph node dissection in differentiated thyroid cancer. World Journal of Surgery. 2007; 31(5):895-904

134. Witt RL, Ferris RL, Pribitkin EA, Sherman SI, Steward DL, Nikiforov YE. Diagnosis and management of differentiated thyroid cancer using molecular biology. Laryngoscope. 2013; 123(4):1059-1064

135. Wojtczak B, Domoslawski P, Dawiskiba J, Lukienczuk T. State-of-the-art treatment of thyroid disorders with special emphasis on surgery including minimally invasive techniques. Family Medicine and Primary Care Review. 2010; 12(4):1059-1063

136. Won HR, Chang JW, Kang YE, Kang JY, Koo BS. Optimal extent of lateral neck dissection for well-differentiated thyroid carcinoma with metastatic lateral neck lymph nodes: A systematic review and meta-analysis. Oral Oncology. 2018; 87:117-125

137. Wong CKH, Lang BHH. A randomized trial comparing health-related quality-of-life and utility measures between routine fine-needle aspiration cytology (FNAC) and surveillance alone in patients with thyroid incidentaloma measuring 1-2 cm. Endocrine. 2020; 67(2):397-405

138. Wong CKH, Liu X, Lang BHH. Cost-effectiveness of fine-needle aspiration cytology (FNAC) and watchful observation for incidental thyroid nodules. Journal of Endocrinological Investigation. 2020; 43(11):1645-1654

139. Yang J, Wang C, Li J, Yang W, Cao G, Wong HM et al. Complete endoscopic thyroidectomy via oral vestibular approach versus areola approach for treatment of thyroid diseases. Journal of Laparoendoscopic and Advanced Surgical Techniques. 2015; 25(6):470-476

140. Yi D, Song P, Huang T, Tang X, Sang J. A meta-analysis on the effect of operation modes on the recurrence of papillary thyroid microcarcinoma. Oncotarget. 2017; 8(4):7148-7156

141. Yip L, Sosa JA. Molecular-directed treatment of differentiated thyroid cancer: Advances in diagnosis and treatment. JAMA Surgery. 2016; 151(7):663-670

142. Yuk-Wah Liu S, Hung-Hin Lang B. Revisiting robotic approaches to endocrine neoplasia: Do the data support their continued use? Current Opinion in Oncology. 2016; 28(1):26-36

143. Zhan S, Luo D, Ge W, Zhang B, Wang T. Clinicopathological predictors of occult lateral neck lymph node metastasis in papillary thyroid cancer: A meta-analysis. Head and Neck. 2019; 41(7):2441-2449

144. Zhang C, Li Y, Li J, Chen X. Total thyroidectomy versus lobectomy for papillary thyroid cancer: A systematic review and meta-analysis. Medicine. 2020; 99(6):e19073

145. Zhang Z, Xia F, Wang W, Jiang B, Yao L, Huang Y et al. Ambulatory thyroidectomy is safe and beneficial in papillary thyroid carcinoma: Randomized controlled trial. Head and Neck. 2021; 43(4):1116-1121

146. Zhao W, You L, Hou X, Chen S, Ren X, Chen G et al. The effect of prophylactic central neck dissection on locoregional recurrence in papillary thyroid cancer after

total thyroidectomy: A systematic review and meta-analysis : Pcnd for the locoregional recurrence of papillary thyroid cancer. Annals of Surgical Oncology. 2017; 24(8):2189-2198

147. Zhao WJ, Luo H, Zhou YM, Dai WY, Zhu JQ. Evaluating the effectiveness of prophylactic central neck dissection with total thyroidectomy for cN0 papillary thyroid carcinoma: An updated meta-analysis. European Journal of Surgical Oncology. 2017; 43(11):1989-2000

148. Zheng W, Li J, Lv P, Chen Z, Fan P. Treatment efficacy between total thyroidectomy and lobectomy for patients with papillary thyroid microcarcinoma: A systemic review and meta-analysis. European Journal of Surgical Oncology. 2018; 44(11):1679-1684

149. Zhu W, Zhong M, Ai Z. Systematic evaluation of prophylactic neck dissection for the treatment of papillary thyroid carcinoma. Japanese Journal of Clinical Oncology. 2013; 43(9):883-888

Appendices

Appendix A – Review protocols

A.1 Review protocol for initial treatment of differentiated thyroid cancer

Field	Content
PROSPERO registration number	CRD42021233669
Review title	The clinical and cost effectiveness of initial treatments for people with differentiated thyroid cancer: active surveillance, hemi-thyroidectomy with or without prophylactic and/or therapeutic node dissection, or total thyroidectomy with or without prophylactic and/or therapeutic node dissection.
Review question	For people with differentiated thyroid cancer, what is the clinical and cost effectiveness of active surveillance, hemi-thyroidectomy (with or without prophylactic or therapeutic node dissection) or total thyroidectomy (with or without prophylactic or therapeutic node dissection)?
Objective	To determine the best initial management strategy for differentiated thyroid cancer
Searches	The following databases (from inception) will be searched: • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR)

	• Embase • MEDLINE Searches will be restricted by: • English language • Human studies • Letters and comments are excluded. Other searches: • Inclusion lists of relevant systematic reviews will be checked by the reviewer. The searches may be re-run 6 weeks before the final committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review. Medine search strategy to be quality assured using the PRESS evidence-based checklist (see methods chapter for full details).
Condition or domain being studied	Thyroid cancer

Population	Inclusion: People aged 16 or over with a diagnosis of differentiated thyroid cancer Exclusion: Children and young people under 16 years.
Intervention/Exposure/Test	• Active surveillance • Hemi -thyroidectomy + node dissection (level 6) • total -thyroidectomy + node dissection (level 6) (including completion) • total thyroidectomy + lateral neck (+ level 6) (including completion) • Hemi -thyroidectomy only • total -thyroidectomy only (including completion) • Other types of thyroidectomy / subtotal
Comparator/Reference standard/Confounding factors	• Each other • However, note that other types of thyroidectomy / subtotal are only to be compared to active surveillance
Types of study to be included	Published NMAs and IPDs will be considered for inclusion. • Systematic reviews of RCTs • RCTs

	Non-randomised trials will not be considered because the committee agreed that spurious data from non-randomised trials would risk potentially harmful recommendations
Other exclusion criteria	Non-English language studies. Abstracts will be excluded as it is expected there will be sufficient full text published studies available.
Context	There is currently some uncertainty as to the optimum surgical methods. In addition, there is a belief that active surveillance may be an appropriate approach in some patients
Primary outcomes (critical outcomes)	All outcomes are considered equally important for decision making and therefore have all been rated as critical: • mortality • quality of life • cost effectiveness • local cancer progression • incidence of distant metastases • cancer recurrence • postoperative dysphagia • recurrent nerve palsy • hypoparathyroidism

	• need for further treatment Time of follow up: longest available
Data extraction (selection and coding)	The full text of potentially eligible studies will be retrieved and will be assessed for eligibility in line with the criteria outlined above. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. An in-house developed database; EviBase, will be used for data extraction. A standardised form is followed to extract data from studies (see Developing NICE guidelines: the manual section 6.4) and for undertaking assessment of study quality. Summary evidence tables will be produced including information on: study setting; study population and participant demographics and baseline characteristics; details of the intervention and control interventions; study methodology' recruitment and missing data rates; outcomes and times of measurement; critical appraisal ratings. 10% of all evidence reviews are quality assured by a senior research fellow. This includes checking: • papers were included /excluded appropriately • a sample of the data extractions • correct methods are used to synthesise data • a sample of the risk of bias assessments Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary.
Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual.

	For Intervention reviews the following checklist will be used according to study design being assessed: • Systematic reviews: Risk of Bias in Systematic Reviews (ROBIS) • Randomised Controlled Trial: Cochrane RoB (2.0)
Strategy for data synthesis	• Where possible, data will be meta-analysed. Pairwise meta-analyses will be performed using Cochrane Review Manager (RevMan5). Fixed-effects (Mantel-Haenszel) techniques will be used to calculate risk ratios for the binary outcomes where possible. Continuous outcomes will be analysed using an inverse variance method for pooling weighted mean differences. Heterogeneity between the studies in effect measures will be assessed using the I^2 statistic and visually inspected. We will consider an I^2 value greater than 50% indicative of substantial heterogeneity. Sensitivity analyses will be conducted based on pre-specified subgroups using stratified meta-analysis to explore the heterogeneity in effect estimates. If this does not explain the heterogeneity, the results will be presented using random-effects. GRADE pro will be used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) will be appraised for each outcome. Publication bias is tested for when there are more than 5 studies for an outcome. Other bias will only be taken into consideration in the quality assessment if it is apparent. Where meta-analysis is not possible, data will be presented and quality assessed individually per outcome.

	If sufficient data is available to make a network of treatments, WinBUGS will be used for network meta-analysis.
Analysis of sub-groups	Stratification The meta-analysis will be stratified for different stages of disease severity: according to categories if TNM 8 scheme (see Appendix A). Also, probably will need another stratum for 'unclear staging' and one for 'mixed stages'. These will be of limited use for decision making but may be the strata that cover most of the papers. Subgroups that will be investigated if heterogeneity is present: • None
Type and method of review	☒ Intervention ☐ Diagnostic ☐ Prognostic ☐ Qualitative ☐ Epidemiologic ☐ Service Delivery ☐ Other (please specify)
Language	English

Country	England
Named contact	**Named contact** National Guideline Centre **Organisational affiliation of the review** National Institute for Health and Care Excellence (NICE) and the National Guideline Centre
Review team members	From the National Guideline Centre: Carlos Sharpin, Guideline lead Mark Perry, Senior systematic reviewer Vimal Bedia, Systematic reviewer Alfredo Mariani, Health economist Lina Gulhane, Head of Information specialists Giulia Zuodar, Project manager
Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.
Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's

	declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10150/documents
Other registration details	N/A
Reference/URL for published protocol	https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=233669
Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.
Keywords	N/A
Details of existing review of same topic by same authors	N/A
Additional information	N/A

Details of final publication	www.nice.org.uk

A.2 Review protocol health economic evidence

Review question	All questions – health economic evidence
Objectives	To identify health economic studies relevant to any of the review questions.
Search criteria	• Populations, interventions and comparators must be as specified in the clinical review protocol above. • Studies must be of a relevant health economic study design (cost–utility analysis, cost-effectiveness analysis, cost–benefit analysis, cost–consequences analysis, comparative cost analysis). • Studies must not be a letter, editorial or commentary, or a review of health economic evaluations. (Recent reviews will be ordered although not reviewed. The bibliographies will be checked for relevant studies, which will then be ordered.) • Unpublished reports will not be considered unless submitted as part of a call for evidence. • Studies must be in English.
Search strategy	A health economic study search will be undertaken using population-specific terms and a health economic study filter – see Appendix B below.
Review strategy	Studies not meeting any of the search criteria above will be excluded. Studies published before 2005, abstract-only studies and studies from non-OECD countries or the USA will also be excluded. Each remaining study will be assessed for applicability and methodological limitations using the NICE economic evaluation checklist which can be found in appendix H of Developing NICE guidelines: the manual (2014).[74] **Inclusion and exclusion criteria** • If a study is rated as both 'Directly applicable' and with 'Minor limitations', then it will be included in the guideline. A health economic evidence table will be completed, and it will be included in the health economic evidence profile. • If a study is rated as either 'Not applicable' or with 'Very serious limitations', then it will usually be excluded from the guideline. If it is excluded, then a health economic evidence table will not be completed, and it will not be included in the health economic evidence profile. • If a study is rated as 'Partially applicable', with 'Potentially serious limitations' or both then there is discretion over whether it should be included.

Where there is discretion

The health economist will make a decision based on the relative applicability and quality of the available evidence for that question, in discussion with the guideline committee if required. The ultimate aim is to include health economic studies that are helpful for decision-making in the context of the guideline and the current NHS setting. If several studies are considered of sufficiently high applicability and methodological quality that they could all be included, then the health economist, in discussion with the committee if required, may decide to include only the most applicable studies and to selectively exclude the remaining studies. All studies excluded on the basis of applicability or methodological limitations will be listed with explanation in the excluded health economic studies appendix below.

The health economist will be guided by the following hierarchies.

Setting:

- UK NHS (most applicable).
- OECD countries with predominantly public health insurance systems (for example, France, Germany, Sweden).
- OECD countries with predominantly private health insurance systems (for example, Switzerland).
- Studies set in non-OECD countries or in the USA will be excluded before being assessed for applicability and methodological limitations.

Health economic study type:

- Cost–utility analysis (most applicable).
- Other type of full economic evaluation (cost–benefit analysis, cost-effectiveness analysis, cost–consequences analysis).
- Comparative cost analysis.
- Non-comparative cost analyses including cost-of-illness studies will be excluded before being assessed for applicability and methodological limitations.

Year of analysis:

- The more recent the study, the more applicable it will be.
- Studies published in 2005 or later but that depend on unit costs and resource data entirely or predominantly from before 2005 will be rated as ‘Not applicable’.
- Studies published before 2005 will be excluded before being assessed for applicability and methodological limitations.

Quality and relevance of effectiveness data used in the health economic analysis:

- The more closely the clinical effectiveness data used in the health economic analysis match with the outcomes of the studies included in the clinical review the more useful the analysis will be for decision-making in the guideline.

Appendix B – Literature search strategies

The literature searches for these reviews are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual, 2014 (updated 2020) https://www.nice.org.uk/process/pmg20/chapter/identifying-the-evidence-literature-searching-and-evidence-submission.

For more information, please see the Methodology review published as part of the accompanying documents for this guideline.

Clinical literature search strategy

This literature search strategy was used for the following review:

- For people with differentiated thyroid cancer, what is the clinical and cost effectiveness of active surveillance, hemi-thyroidectomy (with or without prophylactic or therapeutic node dissection) or total thyroidectomy (with or without prophylactic or therapeutic node dissection)?

Searches were constructed using a PICO framework where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies for interventions as these concepts may not be well described in title, abstract or indexes and therefore difficult to retrieve. Search filters were applied to the search where appropriate.

Table 13: Database parameters, filters and limits applied

Database	Dates searched	Search filters and limits applied
Medline (OVID)	1946 – 13 January 2022	Randomised controlled trials Systematic review studies Observational studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, children) English language
Embase (OVID)	1974 – 13 January 2022	Randomised controlled trials Systematic review studies Observational studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts, children) English language
The Cochrane Library (Wiley)	Cochrane Database of Systematic Reviews to Issue 12 of 12, December 2021	Exclusions (clinical trials, conference abstracts)

Database	Dates searched	Search filters and limits applied
	Cochrane Central Register of Controlled Trials to Issue 12 of 12, December 2021	
Epistemonikos (The Epistemonikos Foundation)	Inception – 13 January 2022	Systematic review Exclusions (Cochrane reviews) English language

1 **Medline (Ovid) search terms**

1.	exp Thyroid Neoplasms/
2.	(thyroid and (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or follicul* or lymphoma* or anaplastic or sarcoma* or medullar* or cyst* or malignan*)).ti,ab.
3.	DTC.ti,ab.
4.	((papillar* or follicul* or medullar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump* or lymphoma*)).ti,ab.
5.	or/1-4
6.	letter/
7.	editorial/
8.	news/
9.	exp historical article/
10.	Anecdotes as Topic/
11.	comment/
12.	case report/
13.	(letter or comment*).ti.
14.	or/6-13
15.	randomized controlled trial/ or random*.ti,ab.
16.	14 not 15
17.	animals/ not humans/
18.	exp Animals, Laboratory/
19.	exp Animal Experimentation/
20.	exp Models, Animal/
21.	exp Rodentia/
22.	(rat or rats or mouse or mice or rodent*).ti.
23.	or/16-22
24.	5 not 23
25.	limit 24 to english language
26.	(exp child/ or exp pediatrics/ or exp infant/) not (exp adolescent/ or exp adult/ or exp middle age/ or exp aged/)
27.	25 not 26
28.	Watchful Waiting/
29.	(active adj2 (surveillanc* or monitor* or observ*)).ti,ab.
30.	"watchful* wait*".ti,ab.
31.	((conservative or expect*) adj2 manag*).ti,ab.

32.	or/28-31
33.	27 and 32
34.	randomized controlled trial.pt.
35.	controlled clinical trial.pt.
36.	randomi#ed.ab.
37.	placebo.ab.
38.	randomly.ab.
39.	clinical trials as topic.sh.
40.	trial.ti.
41.	or/34-40
42.	Meta-Analysis/
43.	Meta-Analysis as Topic/
44.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
45.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
46.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
47.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
48.	(search* adj4 literature).ab.
49.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
50.	cochrane.jw.
51.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
52.	or/42-51
53.	Epidemiologic studies/
54.	Observational study/
55.	exp Cohort studies/
56.	(cohort adj (study or studies or analys* or data)).ti,ab.
57.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
58.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
59.	Controlled Before-After Studies/
60.	Historically Controlled Study/
61.	Interrupted Time Series Analysis/
62.	(before adj2 after adj2 (study or studies or data)).ti,ab.
63.	exp case control study/
64.	case control*.ti,ab.
65.	Cross-sectional studies/
66.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
67.	or/54-67
68.	33 and (41 or 52 or 67)

1 **Embase (Ovid) search terms**

1.	exp Thyroid Cancer/
2.	(thyroid and (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or follicul* or lymphoma* or anaplastic or sarcoma* or medullar* or cyst* or malignan*)).ti,ab.

3.	DTC.ti,ab.
4.	((papillar* or follicul* or medullar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump* or lymphoma*)).ti,ab.
5.	or/1-4
6.	letter.pt. or letter/
7.	note.pt.
8.	editorial.pt.
9.	case report/ or case study/
10.	(letter or comment*).ti.
11.	(conference abstract or conference paper).pt.
12.	or/6-11
13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice or rodent*).ti.
22.	or/14-21
23.	5 not 22
24.	limit 23 to english language
25.	(exp child/ or exp pediatrics/) not (exp adult/ or exp adolescent/)
26.	24 not 25
27.	Watchful Waiting/
28.	*Conservative treatment/
29.	(active adj2 (survellianc* or monitor* or observ*)).ti,ab.
30.	"watchful* wait*".ti,ab.
31.	((conservative or expect*) adj2 manag*).ti,ab.
32.	or/27-31
33.	26 and 32
34.	random*.ti,ab.
35.	factorial*.ti,ab.
36.	(crossover* or cross over*).ti,ab.
37.	((doubl* or singl*) adj blind*).ti,ab.
38.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
39.	crossover procedure/
40.	single blind procedure/
41.	randomized controlled trial/
42.	double blind procedure/
43.	or/34-42
44.	systematic review/
45.	Meta-Analysis/
46.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
47.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.

48.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
49.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
50.	(search* adj4 literature).ab.
51.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
52.	cochrane.jw.
53.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
54.	or/44-53
55.	Clinical study/
56.	Observational study/
57.	family study/
58.	longitudinal study/
59.	retrospective study/
60.	prospective study/
61.	cohort analysis/
62.	follow-up/
63.	cohort*.ti,ab.
64.	63 and 64
65.	(cohort adj (study or studies or analys* or data)).ti,ab.
66.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
67.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
68.	(before adj2 after adj2 (study or studies or data)).ti,ab.
69.	exp case control study/
70.	case control*.ti,ab.
71.	cross-sectional study/
72.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
73.	or/56-62,65-73
74.	33 and (43 or 54 or 73)

1 **Cochrane Library (Wiley) search terms**

#1.	MeSH descriptor: [Thyroid Neoplasms] explode all trees
#2.	thyroid and (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or follicul* or lymphoma* or anaplastic or sarcoma* or medullar* or cyst* or malignan*):ti,ab
#3.	DTC:ti,ab
#4.	((papillar* or follicul* or medullar* or anaplastic) near/2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump* or lymphoma*)):ti,ab
#5.	#1 or #2 or #3 or #4
#6.	MeSH descriptor: [Thyroidectomy] explode all trees
#7.	(thyroidectom* or lobectom* or hemithyroidectom* or isthmusectom* or isthmectom*):ti,ab
#8.	(thyroid near/3 (surg* or operat* or remov*)):ti,ab
#9.	MeSH descriptor: [Watchful Waiting] explode all trees
#10.	(active near/2 (survellianc* or monitor* or observ*)):ti,ab

#11.	watchful* wait*:ti,ab
#12.	((conservative or expect*) near/2 manag*):ti,ab
#13.	((initial or first) near/3 (care or caring or manag* or treatment* or therap*)):ti,ab
#14.	(or #6-#13)
#15.	#5 and #14
#16.	conference:pt or (clinicaltrials or trialsearch):so
#17.	#15 not #16

Epistemonikos search terms

1.	(title:((title:((title:(active surveillance) OR abstract:(active surveillance)) OR (title:(thyroidectomy) OR abstract:(thyroidectomy)) OR (title:(hemithyroidectomy) OR abstract:(hemithyroidectomy))) OR abstract:((title:(active surveillance) OR abstract:(active surveillance)) OR (title:(thyroidectomy) OR abstract:(thyroidectomy)) OR (title:(hemithyroidectomy) OR abstract:(hemithyroidectomy)))) AND (title:((title:(thyroid cancer*) OR abstract:(thyroid cancer*)) OR (title:(thyroid neoplasm*) OR abstract:(thyroid neoplasm*))) OR abstract:((title:(thyroid cancer*) OR abstract:(thyroid cancer*)) OR (title:(thyroid neoplasm*) OR abstract:(thyroid neoplasm*))))) OR abstract:((title:((title:(active surveillance) OR abstract:(active surveillance)) OR (title:(thyroidectomy) OR abstract:(thyroidectomy)) OR (title:(hemithyroidectomy) OR abstract:(hemithyroidectomy))) OR abstract:((title:(active surveillance) OR abstract:(active surveillance)) OR (title:(thyroidectomy) OR abstract:(thyroidectomy)) OR (title:(hemithyroidectomy) OR abstract:(hemithyroidectomy)))) AND (title:((title:(thyroid cancer*) OR abstract:(thyroid cancer*)) OR (title:(thyroid neoplasm*) OR abstract:(thyroid neoplasm*))) OR abstract:((title:(thyroid cancer*) OR abstract:(thyroid cancer*)) OR (title:(thyroid neoplasm*) OR abstract:(thyroid neoplasm*))))))

Health Economics literature search strategy

Health economic evidence was identified by conducting searches using terms for a broad Thyroid Cancer population. The following databases were searched: NHS Economic Evaluation Database (NHS EED - this ceased to be updated after 31st March 2015), Health Technology Assessment database (HTA - this ceased to be updated from 31st March 2018) and The International Network of Agencies for Health Technology Assessment (INAHTA). Searches for recent evidence were run on Medline and Embase from 2014 onwards for health economics, and all years for quality-of-life studies.

Table 2: Database parameters, filters and limits applied

Database	Dates searched	Search filters and limits applied
Medline (OVID)	Health Economics 1 January 2014 – 16 December 2021	Health economics studies Quality of life studies
	Quality of Life 1946 – 16 December 2021	Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
Embase (OVID)	Health Economics 1 January 2014 – 16 December 2021	Health economics studies Quality of life studies

Database	Dates searched	Search filters and limits applied
	Quality of Life 1974 – 16 December 2021	Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
NHS Economic Evaluation Database (NHS EED) (Centre for Research and Dissemination - CRD)	Inception –31st March 2015	
Health Technology Assessment Database (HTA) (Centre for Research and Dissemination – CRD)	Inception – 31st March 2018	
The International Network of Agencies for Health Technology Assessment (INAHTA)	Inception - 16 December 2021	English language

1 **Medline (Ovid) search terms**

1.	exp Thyroid Neoplasms/
2.	(thyroid adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)).ti,ab.
3.	((papillar* or follicul* or medullary or anaplastic) adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)).ti,ab.
4.	or/1-3
5.	letter/
6.	editorial/
7.	news/
8.	exp historical article/
9.	Anecdotes as Topic/
10.	comment/
11.	case report/
12.	(letter or comment*).ti.
13.	or/5-12
14.	randomized controlled trial/ or random*.ti,ab.
15.	13 not 14
16.	animals/ not humans/
17.	exp Animals, Laboratory/
18.	exp Animal Experimentation/
19.	exp Models, Animal/
20.	exp Rodentia/
21.	(rat or rats or mouse or mice).ti.
22.	or/15-21
23.	4 not 22
24.	limit 23 to english language

25.	economics/
26.	value of life/
27.	exp "costs and cost analysis"/
28.	exp Economics, Hospital/
29.	exp Economics, medical/
30.	Economics, nursing/
31.	economics, pharmaceutical/
32.	exp "Fees and Charges"/
33.	exp budgets/
34.	budget*.ti,ab.
35.	cost*.ti.
36.	(economic* or pharmaco?economic*).ti.
37.	(price* or pricing*).ti,ab.
38.	(cost* adj2 (effectiv* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
39.	(financ* or fee or fees).ti,ab.
40.	(value adj2 (money or monetary)).ti,ab.
41.	or/25-40
42.	24 and 41
43.	quality-adjusted life years/
44.	sickness impact profile/
45.	(quality adj2 (wellbeing or well being)).ti,ab.
46.	sickness impact profile.ti,ab.
47.	disability adjusted life.ti,ab.
48.	(qal* or qtime* or qwb* or daly*).ti,ab.
49.	(euroqol* or eq5d* or eq 5*).ti,ab.
50.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
51.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
52.	(hui or hui1 or hui2 or hui3).ti,ab.
53.	(health* year* equivalent* or hye or hyes).ti,ab.
54.	discrete choice*.ti,ab.
55.	rosser.ti,ab.
56.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
57.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
58.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
59.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
60.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
61.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
62.	or/52-70
63.	24 and 62

1 **Embase (Ovid) search terms**

1.	exp Thyroid Cancer/
2.	(thyroid adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)).ti,ab.
3.	((papillar* or follicul* or medullary or anaplastic) adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)).ti,ab.

4.	or/1-3
5.	letter.pt. or letter/
6.	note.pt.
7.	editorial.pt.
8.	case report/ or case study/
9.	(letter or comment*).ti.
10.	or/5-9
11.	randomized controlled trial/ or random*.ti,ab.
12.	10 not 11
13.	animal/ not human/
14.	nonhuman/
15.	exp Animal Experiment/
16.	exp Experimental Animal/
17.	animal model/
18.	exp Rodent/
19.	(rat or rats or mouse or mice).ti.
20.	or/12-19
21.	4 not 20
22.	limit 21 to english language
23.	health economics/
24.	exp economic evaluation/
25.	exp health care cost/
26.	exp fee/
27.	budget/
28.	funding/
29.	budget*.ti,ab.
30.	cost*.ti.
31.	(economic* or pharmaco?economic*).ti.
32.	(price* or pricing*).ti,ab.
33.	(cost* adj2 (effectiv* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
34.	(financ* or fee or fees).ti,ab.
35.	(value adj2 (money or monetary)).ti,ab.
36.	or/23-35
37.	22 and 36
38.	quality-adjusted life years/
39.	"quality of life index"/
40.	short form 12/ or short form 20/ or short form 36/ or short form 8/
41.	sickness impact profile/
42.	(quality adj2 (wellbeing or well being)).ti,ab.
43.	sickness impact profile.ti,ab.
44.	disability adjusted life.ti,ab.
45.	(qal* or qtime* or qwb* or daly*).ti,ab.
46.	(euroqol* or eq5d* or eq 5*).ti,ab.
47.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
48.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
49.	(hui or hui1 or hui2 or hui3).ti,ab.

50.	(health* year* equivalent* or hye or hyes).ti,ab.
51.	discrete choice*.ti,ab.
52.	rosser.ti,ab.
53.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
54.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
55.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
56.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
57.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
58.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
59.	or/37-58
60.	22 and 59

NHS EED and HTA (CRD) search terms

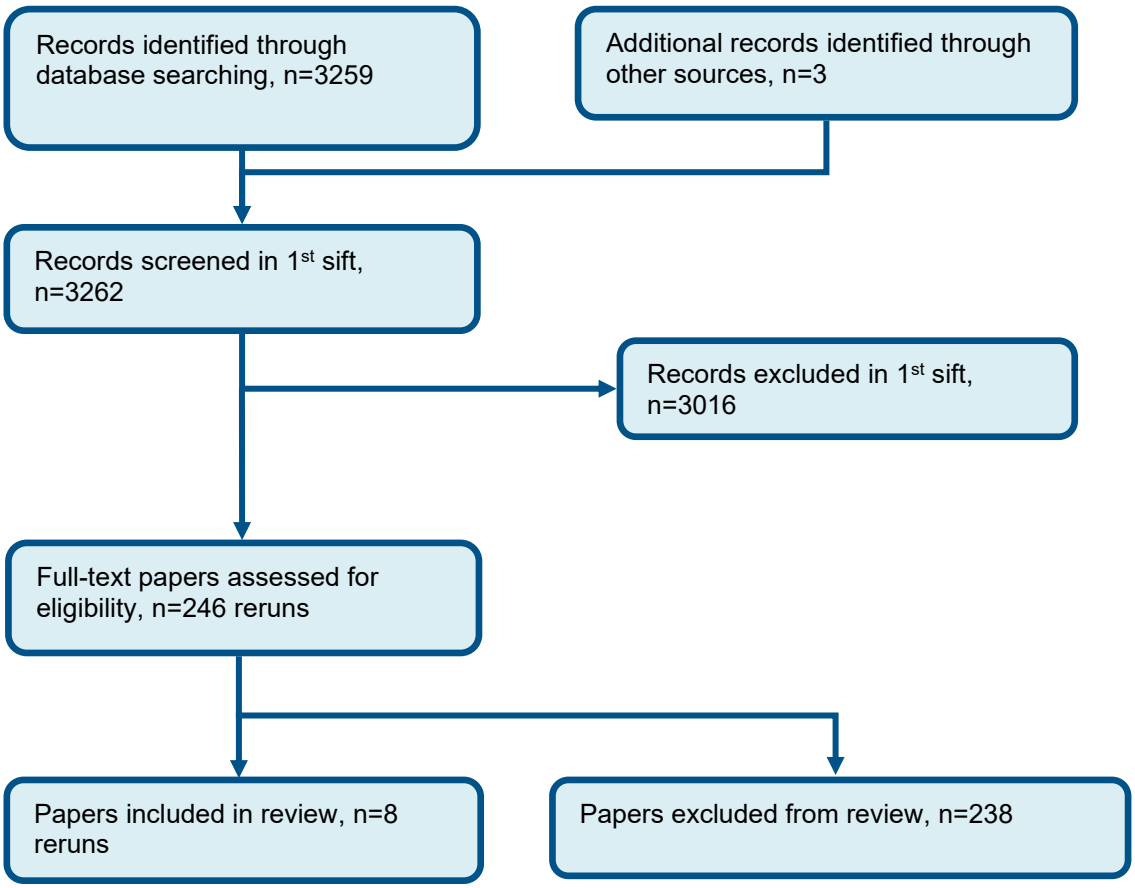
#1.	MeSH DESCRIPTOR Thyroid Neoplasms EXPLODE ALL TREES
#2.	((thyroid NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)))
#3.	(((papillar* or follicul* or medullary or anaplastic) NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)))
#4.	#1 OR #2 OR #3

INHATA search terms

1.	(Thyroid Neoplasms)[mh] OR (thyroid neoplasms) AND (thyroid cancers)

Appendix C – Effectiveness evidence study selection

Figure 1: Flow chart of clinical study selection for the review of initial treatment



Appendix D – Effectiveness evidence

Study	Ahn 2022[1]
Study type	Randomised comparative study
Number of studies (number of participants)	1 (n=112)
Countries and setting	South Korea
Line of therapy	Not applicable
Duration of study	4 years
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	clinically node-negative (cN0) papillary thyroid cancer
Subgroup analysis within study	
Inclusion criteria	Patients aged 20-70; small non invasive PTC; scheduled to receive total thyroidectomy
Exclusion criteria	Suspected advanced PTC and a history of cervical surgery or radiation exposure
Recruitment/selection of patients	Enrolled from 2015-2020. No other recruitment details provided.
Age, gender and ethnicity	Age – mean(sd): 52.7(9.4). Gender (M:F): Male/female ratio: 24/76. Ethnicity: Not reported

Further population details	
Indirectness of population	No indirectness
Interventions	(n=56) Intervention 1: Total thyroidectomy with central node dissection. Duration NA. Concurrent medication/care: not described. Indirectness: No indirectness (n=56) Intervention 2: Total thyroidectomy only. Duration NA. Concurrent medication/care: not described: No indirectness
Funding	No funding received

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: TOTAL THYROIDECTOMY with CND versus TOTAL THYROIDECTOMY ONLY

Protocol outcome 1: cancer recurrence
- Actual outcome: local cancer recurrence: 46 months; Total with CND: 7/51; TOTAL only: 8/50
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 5; Group 2 Number missing: 6; Reasons were similar across groups and not necessarily related to outcome

Protocol outcome 2: Recurrent laryngeal nerve palsy
- Actual outcome: transient recurrent laryngeal nerve injury; Total with CND: 5/51; TOTAL only: 3/50; Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 5; Group 2 Number missing: 6; Reasons were similar across groups and not necessarily related to outcome

Protocol outcome 3: Hypoparathyroidism
- Actual outcome: transient hypoparathyroidism; Total with CND: 7/51; TOTAL only: 13/50
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement -

Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 5; Group 2 Number missing: 6; Reasons were similar across groups and not necessarily related to outcome Protocol outcome 4: Need for further treatment - Actual outcome: treatment with RAI; Total with CND: 11/51; TOTAL only: 11/50 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 5; Group 2 Number missing: 6; Reasons were similar across groups and not necessarily related to outcome	
Protocol outcomes not reported by the study	Mortality; quality of life; local cancer progression; incidence of distant mets; post operative dysphagia;

Study	**Ali 2011[5]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=60)
Countries and setting	Conducted in Pakistan; Setting: Secondary care
Line of therapy	Not applicable
Duration of study	Intervention + follow up: 3 months post discharge from hospital
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Histological confirmation of differentiated thyroid cancer
Stratum	Unclear
Subgroup analysis within study	Not applicable
Inclusion criteria	Histologically proven differentiated carcinoma of thyroid; above 12 years of age.
Exclusion criteria	Patients with toxic goitre; pregnant women.
Recruitment/selection of patients	Selected from the Department of Surgery, Mayo Hospital Lahore.
Age, gender and ethnicity	Age - Other: Mean age (SD) in group receiving total thyroidectomy: 35.6 (6.7); and in group receiving lobectomy + isthmusectomy: 37.6 (7.3). All patients were age 20 years or above. Gender (M:F): Define. Ethnicity:
Further population details	
Indirectness of population	No indirectness

Interventions	(n=30) Intervention 1: Surgery - Total -thyroidectomy only (including completion). dose/quantity, brand name, extra details. Duration From surgery until 3 months post hospital discharge. Concurrent medication/care: Prophylactic antibiotic given at induction of anaesthesia (same antibiotic for each group).. Indirectness: No indirectness (n=30) Intervention 2: Surgery - Hemi -thyroidectomy only. One lobe of the thyroid gland removed with isthmusectomy. Duration From surgery until 3 months post hospital discharge. Concurrent medication/care: Prophylactic antibiotic given at induction of anaesthesia (same antibiotic for each group). Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: TOTAL -THYROIDECTOMY ONLY (INCLUDING COMPLETION) versus HEMI -THYROIDECTOMY ONLY

Protocol outcome 1: Cancer recurrence at Define
- Actual outcome for Unclear: Recurrence. From surgery until 3 months post hospital discharge.; Group 1: 0/30, Group 2: 6/30
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Recurrence was not clearly defined. Measurement method was not described; however, it is likely that recurrence was objectively determined by clinical and histological methods.; Indirectness of outcome: No indirectness; Baseline details: Groups were comparable for age distribution and proportions with papillary/follicular cancer. No other baseline details were given.; Blinding details: No description of blinding; however the outcome is likely to have been objectively determined.; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 2: Recurrent laryngeal nerve palsy at Define
- Actual outcome for Unclear: Hoarseness at From surgery until 3 months post hospital discharge.; Group 1: 2/30, Group 2: 2/30; Comments: Hoarseness was not clearly described as having resulted from recurrent laryngeal nerve palsy, but this can probably be safely assumed to be the case.
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Very high, Crossover - Low, Comments - Hoarseness was not defined. There was no description of vocal cord examination.; Indirectness of outcome: Serious indirectness, Comments: Hoarseness was not defined. There was no description of vocal cord examination.; Baseline details: Groups were comparable for age distribution and proportions with papillary/follicular cancer. No other baseline details were given.; Group 1 Number missing; Group 2 Number missing:

Protocol outcomes not reported by the study	Quality of life at Define; Hypoparathyroidism at Define; incidence of distant metastases at Define; local cancer progression at Define; mortality at Define; Need for further treatment at Define; post-operative dysphagia at Define; do not use at Define

Study	**Jeon 2019[60]**
Study type	Non-randomised comparative study
Number of studies (number of participants)	1 (n=191)
Countries and setting	
Line of therapy	Not applicable
Duration of study	Other: Cross-sectional Time interval from initial diagnosis: Active surveillance group: 29.6 months Surgery group: 38.0 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Unclear
Subgroup analysis within study	Not applicable
Inclusion criteria	Cytologically diagnosed papillary thyroid microcarcinoma,
Exclusion criteria	Evidence of disease progression or recurrent/persistent disease during follow-up, uncontrolled chronic disease or other malignancies, or a history of other malignancies.
Recruitment/selection of patients	Enrolled from June 2016 to October 2017. No other recruitment details provided.

Age, gender and ethnicity	Age - Median (IQR): Active surveillance: 50.3(+/-10.57); lobectomy: 51.0(+/-10.38). Gender (M:F): Male/female ratio: 36/155. Ethnicity:
Further population details	
Indirectness of population	No indirectness
Interventions	(n=43) Intervention 1: No active treatment - Active surveillance. No description. Duration Median 29.6 months (IQR 14.2 to 37.5). Concurrent medication/care: No description. Indirectness: No indirectness (n=148) Intervention 2: Surgery - Hemi -thyroidectomy only. No details were given. Duration Median 38 months (IQR 25.4 to 53). Concurrent medication/care: No details were given. Indirectness: Serious indirectness; Indirectness comment: It was unclear whether lymph node dissection was performed in addition to hemithyroidectomy.
Funding	Academic or government funding (Grant of the Korean Health Technology R&D Project, Ministry of Health and Welfare, Republic of Korea (HC15C3372).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: ACTIVE SURVEILLANCE versus HEMI -THYROIDECTOMY ONLY

Protocol outcome 1: Quality of life at Define
- Actual outcome for Unclear: SF-12 Physical Component Summary at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: SF-12 Mental Component Summary at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;

Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (neuromuscular) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (voice) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (concentration) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (sympathetic symptoms) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5);

hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (throat/mouth) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (psychological) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (sensory) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:

- Actual outcome for Unclear: THYCA-QOL (problems with scar) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (felt chilly) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (tingling hands/feet) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome for Unclear: THYCA-QOL (gained weight) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).;
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the

questionnaires.; Group 1 Number missing: ; Group 2 Number missing: - Actual outcome for Unclear: THYCA-QOL (headache) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).; Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing: - Actual outcome for Unclear: THYCA-QOL (less interest in sex) at Active surveillance: median 29.6 months (IQR 14.2 to 37.5); hemithyroidectomy: median 38 months (IQR 25.4 to 53).; Risk of bias: All domain - Very high, Selection - Very high, Blinding - Very high, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Group allocation determined by disease severity (active surveillance only if tumour </= 1cm, Bethesda category 5 or 6, no lateral cervical lymph node or distant metastasis, macroscopic invasion or invasion to trachea or recurrent laryngeal nerve). Groups dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires. ; Indirectness of outcome: No indirectness ; Baseline details: Groups were similar for age, marital status, education level, employment status and socioeconomic status, and dissimilar for sex, LT4 supplementation, TSH level and time from initial diagnosis to completion of the questionnaires.; Group 1 Number missing: ; Group 2 Number missing:
Protocol outcomes not reported by the study	do not use at Define; Hypoparathyroidism at Define; incidence of distant metastases at Define; local cancer progression at Define; mortality at Define; Need for further treatment at Define; post-operative dysphagia at Define; Recurrent laryngeal nerve palsy at Define; Cancer recurrence at Define

Study	**Kim 2020[61]**

Study type	Randomised comparative study
Number of studies (number of participants)	1 (n=184)
Countries and setting	South Korea
Line of therapy	Not applicable
Duration of study	5 years
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	clinically node-negative (cN0) papillary thyroid microcarcinoma (PTMC).
Subgroup analysis within study	
Inclusion criteria	(1) age between 18 and 70 years, (2) cytologically proven PTMC, (3) no evidence of clinically positive lymph node and (4) acquisition of informed consent from patient.
Exclusion criteria	Previous history of irradiation or other malignancy
Recruitment/selection of patients	Enrolled from November 2011 to June 2014. No other recruitment details provided.
Age, gender and ethnicity	Age - Median (IQR): Hemithyroidectomy with pCND: 47.9(+/-9.1); Hemithyroidectomy alone: 48.5(+/-9.4). Gender (M:F): Male/female ratio: 54/110. Ethnicity: Not reported
Further population details	
Indirectness of population	No indirectness

Interventions	(n=94) Intervention 1: Hemithyroidectomy with prophylactic central node dissection. Duration NA. Concurrent medication/care: Total thyroidectomy used if clinically indicated. Indirectness: No indirectness (n=90) Intervention 2: Hemithyroidectomy only. Duration NA. Concurrent medication/care: Total thyroidectomy used if clinically indicated Indirectness: No indirectness
Funding	Not reported

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: HEMITHYROIDECTOMY with prophylactic CND versus HEMITHYROIDECTOMY ONLY

Protocol outcome 1: cancer recurrence
- Actual outcome: Regional recurrence: 60 months; Hemi with pCND: 3/82; HEMI only 1/82
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups differed for gender [more males in pCND group], and size of tumour and presence of ETE [bias favouring HEMI only group]; Group 1 Number missing: 8; Group 2 Number missing: 12; Reasons were loss to follow up or FU < 24 months.

Protocol outcome 2: need for further treatment
- Actual outcome: Need for conversion to total thyroidectomy: post operative; Hemi with pCND: 10/82; HEMI only 0/82
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups differed for gender [more males in pCND group], and size of tumour and presence of ETE [bias favouring HEMI only group]; Group 1 Number missing: 8; Group 2 Number missing: 12; Reasons were loss to follow up or FU < 24 months.

Protocol outcome 3: Recurrent laryngeal nerve palsy
- Actual outcome: vocal cord paralysis: post operative; Hemi with pCND: 0/82; HEMI only 1/82
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups differed for gender [more males in pCND group], and size of tumour and presence of ETE [bias favouring HEMI only group]; Group 1 Number missing: 8; Group 2 Number missing: 12; Reasons were loss to follow up or FU < 24 months.

Protocol outcome 4: Hypoparathyroidism - Actual outcome: Hypocalcaemia: post operative; Hemi with pCND: 0/82; HEMI only 0/82 Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups differed for gender [more males in pCND group], and size of tumour and presence of ETE [bias favouring HEMI only group]; Group 1 Number missing: 8; Group 2 Number missing: 12; Reasons were loss to follow up or FU < 24 months.	
Protocol outcomes not reported by the study	Mortality; quality of life; local cancer progression; incidence of distant mets; post operative dysphagia;

Study	**Megwalu 2017[70]**
Study type	Non-randomised comparative study
Number of studies (number of participants)	(n=2323)
Countries and setting	Conducted in USA; Setting:
Line of therapy	Not applicable
Duration of study	Other: Data were extracted from registrations made between 1988 and 2009 in the Surveillance, Epidemiology, and End Results ('SEER') 18 database of the National Cancer Institute.
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Stage I
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 65years and over, diagnosed with early-stage papillary thyroid carcinoma between 1988 and 2009, with a tumour size of 1 cm or less.
Exclusion criteria	Regional or distant metastasis, multiple primaries and patients in whom surgery was contraindicated.
Recruitment/selection of patients	Data were extracted from a population-based cancer database in the US.
Age, gender and ethnicity	Age - Range: 65 to 97. Gender (M:F): 477:1846. Ethnicity: Black, other.
Further population details	

Indirectness of population	No indirectness
Interventions	(n=15) Intervention 1: No active treatment - Active surveillance. Patients received no surgical therapy in the immediate treatment period. Duration 11 years of observation recorded from database. Concurrent medication/care: No co-interventions were described. Indirectness: No indirectness (n=2308) Intervention 2: Surgery - Other types of thyroidectomy / subtotal. Patients received underwent either thyroid lobectomy or total thyroidectomy in the immediate four-month treatment period. Duration 11 years of . Concurrent medication/care: No con-interventions were described. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: ACTIVE SURVEILLANCE versus OTHER TYPES OF THYROIDECTOMY / SUBTOTAL Protocol outcome 1: mortality at Define - Actual outcome for Stage I: Overall cumulative survival at 5 years; Group 1: n=15; Group 2: n=2308; HR 0.11; Lower CI 0.09 to Upper CI 0.13; Advantage to research or control? Research; Actuarial or Kaplan Meier curves reported? Yes Risk of bias: All domain – Very High, Selection – Very High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Risk of selection bias regarded as 'very high' because of failure to include comorbidity as a covariate in propensity analysis.; Indirectness of outcome: No indirectness; Baseline details: Comparable for multifocal disease, extrathyroidal invasion (none in either group), female, mean age. Higher proportion 'other' race in the active surveillance group (20% versus 8.8%). ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Quality of life at Define; do not use at Define; Hypoparathyroidism at Define; incidence of distant metastases at Define; local cancer progression at Define; Need for further treatment at Define; post-operative dysphagia at Define; Recurrent laryngeal nerve palsy at Define; Cancer recurrence at Define
Study	**Moon 2021[71]**

Study type	Non-randomised comparative study
Number of studies (number of participants)	1 (n=1055)
Countries and setting	South Korea
Line of therapy	Not applicable
Duration of study	2 years
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Low risk PTMC.
Subgroup analysis within study	
Inclusion criteria	Patients diagnosed with PTMC
Exclusion criteria	Suspected major organ involvement; poorly differentiated histology or cytology; variant with a poor prognosis
Recruitment/selection of patients	Enrolled from 2016-2020. No other recruitment details provided.
Age, gender and ethnicity	Age - Mean (sd): AS: 48.8(11.9); Lobectomy or Isthmusectomy 45.7(10.4). Gender (M:F): Male/female ratio: 236/819. Ethnicity: Not reported
Further population details	
Indirectness of population	No indirectness

Interventions	(n=94) Intervention 1: Active surveillance. Duration NA. Concurrent medication/care: none. Indirectness: No indirectness (n=90) Intervention 2: Lobectomy or isthmusectomy or total thyroidectomy. Duration NA. Concurrent medication/care: none; Indirectness: No indirectness
Funding	Not reported
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: Active surveillance versus Lobectomy or isthmusectomy or total thyroidectomy Protocol outcome 1: quality of life - Actual outcome: Overall health score on the Korean version of the thyroid-specific QoL questionnaire (Dow): 24 months; AS versus Lobectomy/isthmusectomy: standardised beta coefficient 0.141 (0.034-0.248); AS versus TT: standardised beta co-efficient 0.354 (0.179-0.529). Note: a positive value denotes a benefit to AS. Risk of bias: All domain - Very high, Selection - High, Blinding - NA, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups were based on patient choice and the AS group were older and had smaller tumours. However the analysis is adjusted through the use of GEE; Group 1 Number missing: 57; Group 2 Number missing: 24; Reasons were loss to follow up or FU < 24 months.	
Protocol outcomes not reported by the study	Mortality; local cancer progression; incidence of distant mets; post operative dysphagia; recurrent laryngeal nerve palsy, hypothyroidism, need for further treatment

Study	**Sippel 2020[103]**

Study type	Randomised comparative study
Number of studies (number of participants)	1 (n=60)
Countries and setting	USA
Line of therapy	Not applicable
Duration of study	1 year
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	clinically node-negative (cN0) papillary thyroid cancer
Subgroup analysis within study	
Inclusion criteria	Confirmed diagnosis of PTC or a fine needle aspirate (FNA) and/or ultrasound (US) that were suspicious for PTC; between the ages of 21–70,
Exclusion criteria	Other malignancy; evidence of nodal disease or distant metastasis at their initial presentation.
Recruitment/selection of patients	Enrolled from June 2014. No other recruitment details provided.
Age, gender and ethnicity	Age - Median (IQR): Total thyroidectomy with CND: 50.1(+/-2.4); Total thyroidectomy alone: 46.1(+/-2.5). Gender (M:F): Male/female ratio: 24/76. Ethnicity: Not reported
Further population details	
Indirectness of population	No indirectness

Interventions	(n=31) Intervention 1: Total thyroidectomy with central node dissection. Duration NA. Concurrent medication/care: not described. Indirectness: No indirectness (n=30) Intervention 2: Total thyroidectomy only. Duration NA. Concurrent medication/care: not described: No indirectness
Funding	Support for this research included the University of Wisconsin Carbone Cancer Center Support Grant P30CA014520 and the National Cancer Institute of the National Institutes of Health award number R01CA176911.

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: TOTAL THYROIDECTOMY with CND versus TOTAL THYROIDECTOMY ONLY

Protocol outcome 1: quality of life
- Actual outcome: ThyCA QoL: 12 months; No group data provided but p=0.96
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 1; Group 2 Number missing: 0; Reasons were patient excluded as found not to have cancer

Protocol outcome 2: Recurrent laryngeal nerve palsy
- Actual outcome: EAT-10 swallowing score: 12 months; Total with CND: 1.22(0.54); TOTAL only: 0.76(0.31)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 1; Group 2 Number missing: 0; Reasons were patient excluded as found not to have cancer

Protocol outcome 3: Hypoparathyroidism
- Actual outcome: Calcium: 6 months; Total with CND: 9.1(0.1); TOTAL only: 9.0(0.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 1; Group 2 Number missing: 0; Reasons were patient excluded as found not to have cancer

Protocol outcome 4: Hypoparathyroidism - Actual outcome: PTH: 6 months; Total with CND: 46.5(5); TOTAL only: 45(4.7) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Groups similar for most characteristics; Group 1 Number missing: 1; Group 2 Number missing: 0; Reasons were patient excluded as found not to have cancer *In the main analysis shown in GRADE tables and Forest plots the variance values above have been multiplied by the square root of the sample size. This is on the premise that the variance values in the published article are actually standard errors of the mean and not standard deviations. Multiplying by the square root of the sample size converts these figures to standard deviations. Although the paper did not state that standard errors had been used, this was assumed based on the lack of coherence between the authors' statements of statistical significance and the results gained when assuming that the variance values were standard deviations. When values were converted to standard errors, the loss of coherence disappeared.*	
Protocol outcomes not reported by the study	Mortality; local cancer progression; incidence of distant mets; cancer recurrence; post operative dysphagia; need for further treatment

Study	**Viola 2015[124]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=196)
Countries and setting	Conducted in Italy; Setting: Hospital: Department of Surgery and Department of Clinical and Experimental Medicine of Pisa University.
Line of therapy	Not applicable
Duration of study	Follow up (post intervention): 5 years
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: PTC documented by fine needle aspiration cytology. The histological diagnosis and staging was made according to the standard classification.
Stratum	Mixed (no group >75%)
Subgroup analysis within study	Not applicable
Inclusion criteria	1) PTC documented by fine needle aspiration cytology, 2) no evidence of lymph node metastases (cN0) at palpation and neck ultrasound (US); 3) no clinical evidence of distant metastases at diagnosis; 4) at least 18 years of age
Exclusion criteria	1) Histotypes other than PTC, 2) evidence of lymph node metastases during surgery even if not previously visualized at neck ultrasound.
Recruitment/selection of patients	Consecutive patients with PTC with no preoperative clinical evidence of lymph node metastases (cN0) were invited to participate.
Age, gender and ethnicity	Age - Mean (range): 44.5 (18 to 80). Gender (M:F): 46:135. Ethnicity:

Further population details	
Indirectness of population	No indirectness
Interventions	(n=98) Intervention 1: Surgery - Total -thyroidectomy + node dissection (level 6) (including completion). Node dissection comprised removal of the nodes of the prelaryngeal, pretracheal, and both the right and left paratracheal basins. Duration 5 years. Concurrent medication/care: After surgical treatment, all the patients were treated with low radioiodine (131I) activities (1.1 GBq/30 mCi) for postsurgical thyroid remnant ablation, when indicated (10). Subsequent treatments of 131I ranging from 100–150 mCi (3.7–4.05 GBq) were administered when required. Patients who were not considered free of disease underwent subsequent 131I and/or other surgical treatments if necessary. Indirectness: No indirectness (n=98) Intervention 2: Surgery - Total -thyroidectomy only (including completion). dose/quantity, brand name, extra details. Duration 5 years. Concurrent medication/care: After surgical treatment, all the patients were treated with low radioiodine (131I) activities (1.1 GBq/30 mCi) for postsurgical thyroid remnant ablation, when indicated (10). Subsequent treatments of 131I ranging from 100–150 mCi (3.7–4.05 GBq) were administered when required. Patients who were not considered free of disease underwent subsequent 131I and/or other surgical treatments if necessary. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: TOTAL -THYROIDECTOMY + NODE DISSECTION (LEVEL 6) (INCLUDING COMPLETION) versus TOTAL -THYROIDECTOMY ONLY (INCLUDING COMPLETION)

Protocol outcome 1: Cancer recurrence at Define
- Actual outcome for Unclear: Disease persistence at 5 years follow-up; Group 1: 7/93, Group 2: 7/88
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Patients were considered free of disease when sTg levels after recombinant human TSH were less than 1 ng/mL, neck ultrasound was negative, and TgAb undetectable. The patients who did not undergo 131I remnant ablation were considered to be free of disease when neck ultrasound was negative and sTg and TgAb were undetectable and/or stable during follow-up.; Indirectness of outcome: No indirectness ; Baseline details: Comparable for age, sex, familiarity, tumoral capsule infiltration, aggressive variant,

multifocality, bilaterality, extrathyroid extension, nodule size, tumour size, T1 vs T>1, T3+T4 vs T1+T2, advanced stage (II and III), and BRAF V600E. ; Group 1 Number missing: , Reason: Unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3). ; Group 2 Number missing: , Reason: Did not receive allocated intervention (evidence of suspicious lymph nodes at surgery - n=5); unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3).

Protocol outcome 2: Recurrent laryngeal nerve palsy at Define
- Actual outcome for Unclear: Recurrent laryngeal nerve palsy at 5 years follow-up; Group 1: 4/93, Group 2: 7/88
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Patients were considered free of disease when sTg levels after recombinant human TSH were less than 1 ng/mL, neck ultrasound was negative, and TgAb undetectable. The patients who did not undergo 131I remnant ablation were considered to be free of disease when neck ultrasound was negative and sTg and TgAb were undetectable and/or stable during follow-up.; Indirectness of outcome: No indirectness ; Baseline details: Comparable for age, sex, familiarity, tumoral capsule infiltration, aggressive variant, multifocality, bilaterality, extrathyroid extension, nodule size, tumour size, T1 vs T>1, T3+T4 vs T1+T2, advanced stage (II and III), and BRAF V600E. ; Group 1 Number missing: , Reason: Unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3). ; Group 2 Number missing: , Reason: Did not receive allocated intervention (evidence of suspicious lymph nodes at surgery - n=5); unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3).

Protocol outcome 3: Hypoparathyroidism at Define
- Actual outcome for Unclear: Permanent hypoparathyroidism at 5 years follow-up; Group 1: 18/93, Group 2: 7/88
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Patients were considered free of disease when sTg levels after recombinant human TSH were less than 1 ng/mL, neck ultrasound was negative, and TgAb undetectable. The patients who did not undergo 131I remnant ablation were considered to be free of disease when neck ultrasound was negative and sTg and TgAb were undetectable and/or stable during follow-up.; Indirectness of outcome: No indirectness ; Baseline details: Comparable for age, sex, familiarity, tumoral capsule infiltration, aggressive variant, multifocality, bilaterality, extrathyroid extension, nodule size, tumour size, T1 vs T>1, T3+T4 vs T1+T2, advanced stage (II and III), and BRAF V600E. ; Group 1 Number missing: 5, Reason: Unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3). ; Group 2 Number missing: 10, Reason: Did not receive allocated intervention (evidence of suspicious lymph nodes at surgery - n=5); unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3).

Protocol outcome 4: Need for further treatment at Define
- Actual outcome for Unclear: Requirement for additional course(s) of radioactive iodine at 5 years follow-up; Group 1: 3/89, Group 2: 15/86
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - Patients were considered free of disease when sTg levels after recombinant human TSH were less than 1 ng/mL, neck ultrasound was negative, and TgAb undetectable. The patients who did not undergo 131I remnant ablation were considered to be free of disease when neck ultrasound was negative and sTg and TgAb were undetectable and/or stable during follow-up.;

Indirectness of outcome: No indirectness ; Baseline details: Comparable for age, sex, familiarity, tumoral capsule infiltration, aggressive variant, multifocality, bilaterality, extrathyroid extension, nodule size, tumour size, T1 vs T>1, T3+T4 vs T1+T2, advanced stage (II and III), and BRAF V600E. ; Group 1 Number missing: , Reason: Unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3). ; Group 2 Number missing: , Reason: Did not receive allocated intervention (evidence of suspicious lymph nodes at surgery - n=5); unwilling to continue scheduled visits (n=2); unwilling to perform radioiodine treatment in location of study (n=3).	
Protocol outcomes not reported by the study	Quality of life at Define; incidence of distant metastases at Define; local cancer progression at Define; mortality at Define; post-operative dysphagia at Define; do not use at Define

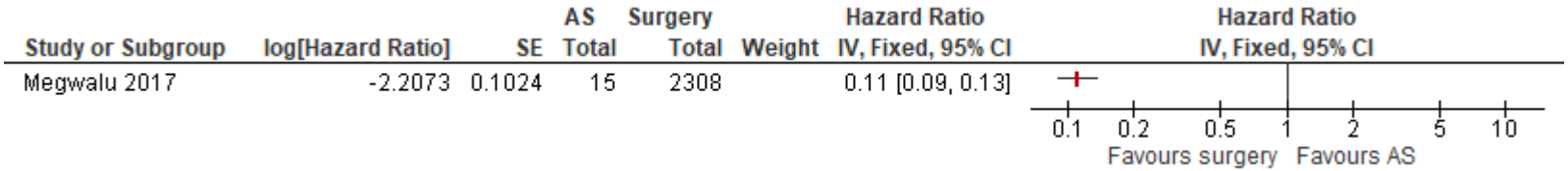
1

Appendix E – Forest plots

Stage 1 disease

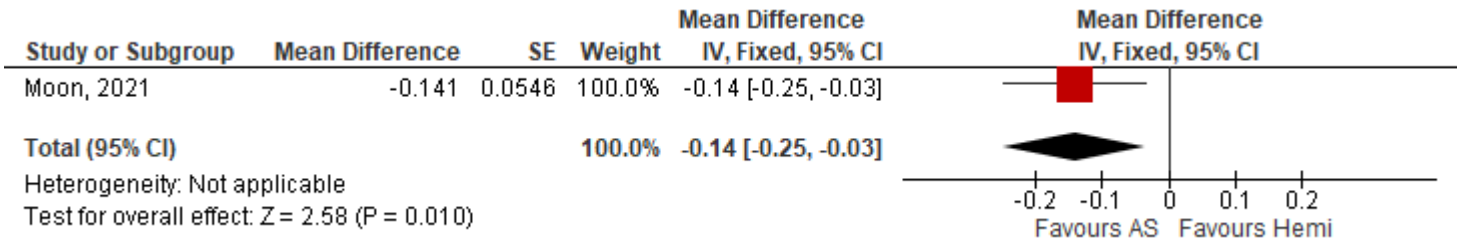
Surgery (total thyroidectomy alone or hemithyroidectomy alone) versus active surveillance

Figure 1: Overall mortality



Hemithyroidectomy alone versus active surveillance

Figure 2: Quality of life



Total thyroidectomy alone versus active surveillance

Figure 3: Quality of life

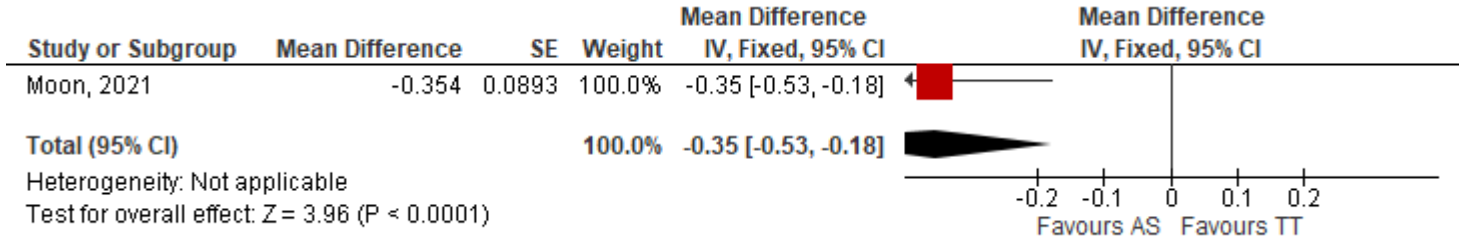


Figure 4: Cancer recurrence

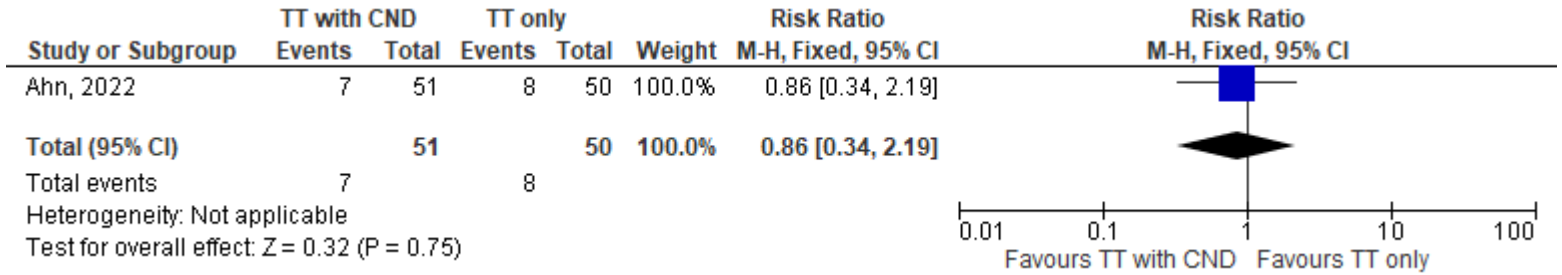


Figure 5: Recurrent laryngeal nerve palsy

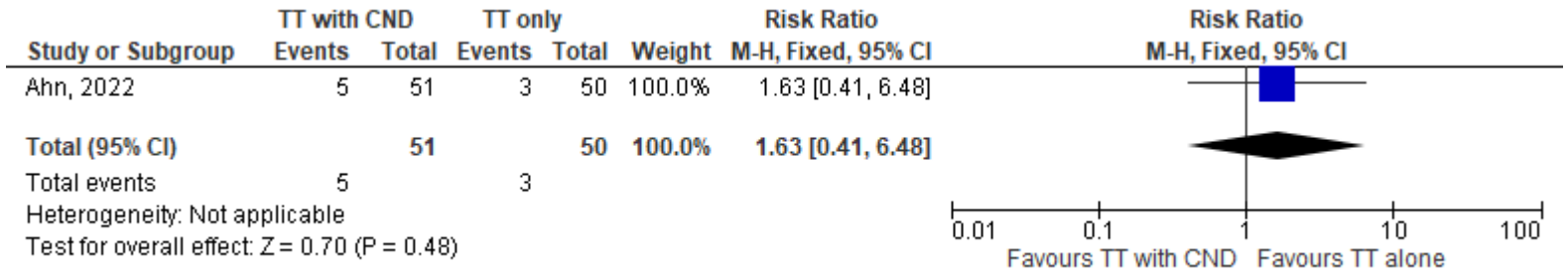


Figure 6: Hypoparathyroidism

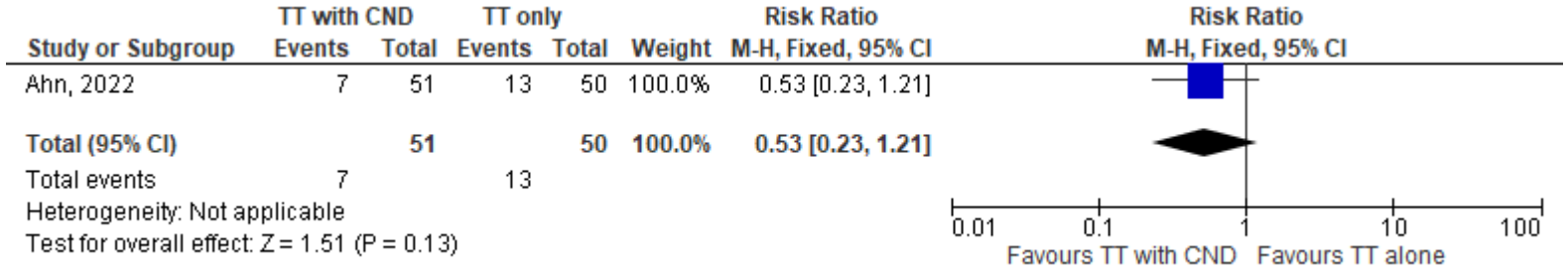


Figure 7: Need for further treatment

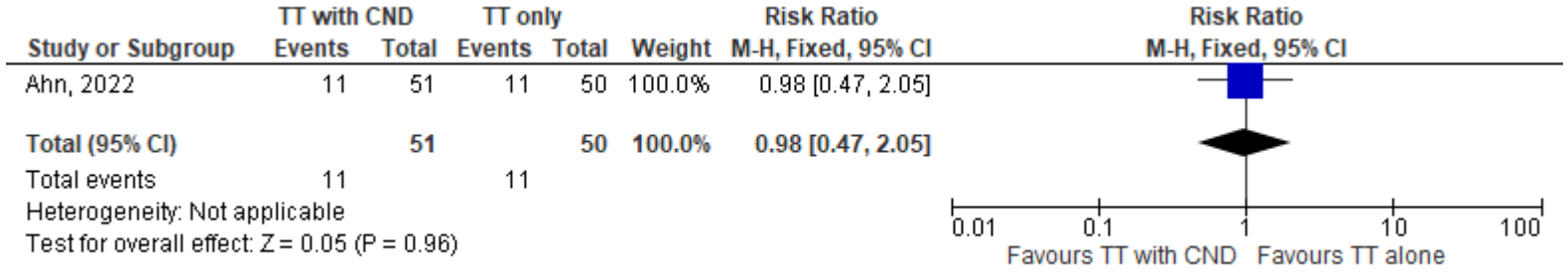


Figure 8: EAT-10 swallowing score

Study or Subgroup	TT with CND Mean	SD	Total	TT only Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Sippel, 2020	1.22	2.96	30	0.76	4.16	30	100.0%	0.46 [-1.37, 2.29]
Total (95% CI)			**30**			**30**	**100.0%**	**0.46 [-1.37, 2.29]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.49 (P = 0.62)

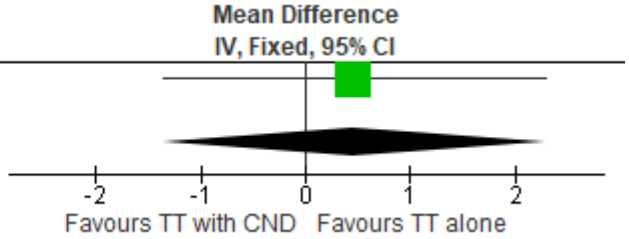


Figure 9: Calcium levels

Study or Subgroup	TT with CND Mean	SD	Total	TT only Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Sippel, 2020	9.1	0.5477	30	9	0.5477	30	100.0%	0.10 [-0.18, 0.38]
Total (95% CI)			**30**			**30**	**100.0%**	**0.10 [-0.18, 0.38]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.71 (P = 0.48)

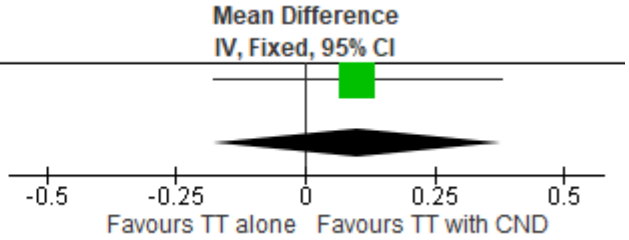
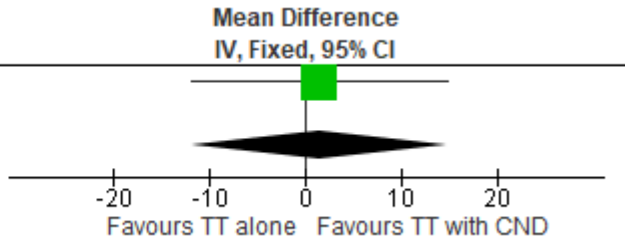


Figure 10: PTH levels

Study or Subgroup	TT with CND Mean	SD	Total	TT only Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Sippel, 2020	46.5	27.38	30	45	25.74	30	100.0%	1.50 [-11.95, 14.95]
Total (95% CI)			**30**			**30**	**100.0%**	**1.50 [-11.95, 14.95]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.22 (P = 0.83)



Hemithyroidectomy with pCND versus Hemithyroidectomy alone

Figure 11: Cancer recurrence

Study or Subgroup	Hemi with pCN Events	Total	Hemi alone Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Kim, 2020	3	82	1	82	100.0%	3.00 [0.32, 28.25]
Total (95% CI)		**82**		**82**	**100.0%**	**3.00 [0.32, 28.25]**
Total events	3		1			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.96 (P = 0.34)

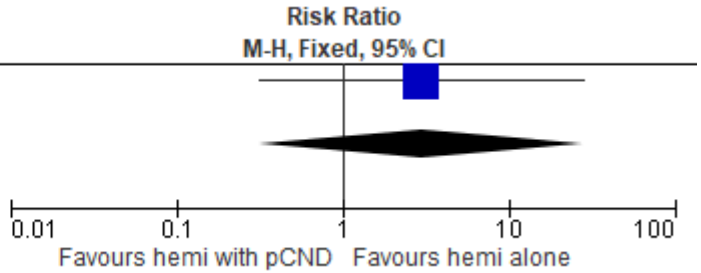


Figure 12: Need for further treatment

Study or Subgroup	Hemi with pCN Events	Total	Hemi alone Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Kim, 2020	10	82	0	82	100.0%	8.31 [2.32, 29.73]
Total (95% CI)		**82**		**82**	**100.0%**	**8.31 [2.32, 29.73]**
Total events	10		0			

Heterogeneity: Not applicable
Test for overall effect: Z = 3.25 (P = 0.001)

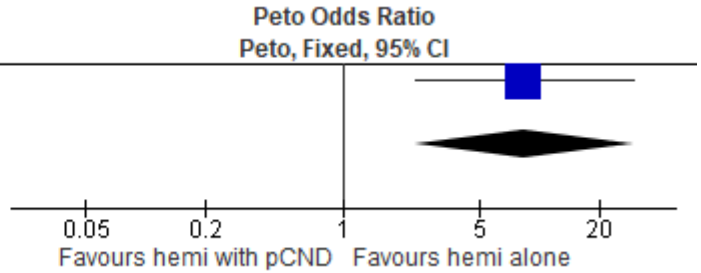


Figure 13: Recurrent laryngeal nerve plasy

Study or Subgroup	Hemi with pCN Events	Total	Hemi alone Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Kim, 2020	0	82	1	82	100.0%	0.14 [0.00, 6.82]
Total (95% CI)		**82**		**82**	**100.0%**	**0.14 [0.00, 6.82]**
Total events	0		1			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.00 (P = 0.32)

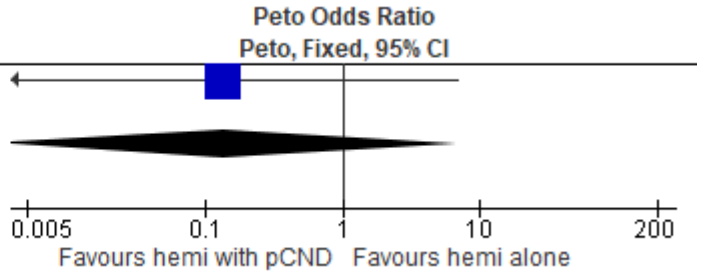
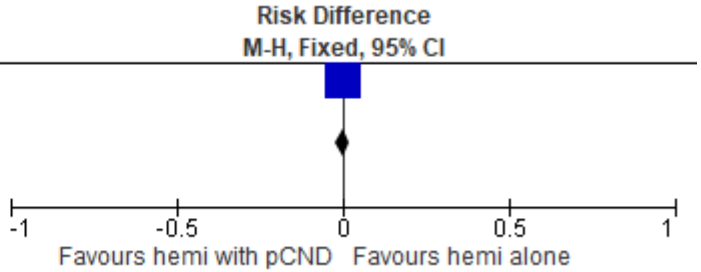


Figure 14: Hypoparathyroidism

Study or Subgroup	Hemi with pCN Events	Total	Hemi alone Events	Total	Weight	Risk Difference M-H, Fixed, 95% CI
Kim, 2020	0	82	0	82	100.0%	0.00 [-0.02, 0.02]
Total (95% CI)		**82**		**82**	**100.0%**	**0.00 [-0.02, 0.02]**
Total events	0		0			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.00 (P = 1.00)



Disease stage unclear or mixed

Total thyroidectomy alone (TT) versus hemithyroidectomy plus isthmusectomy alone (HT+I)

Figure 15: Cancer recurrence

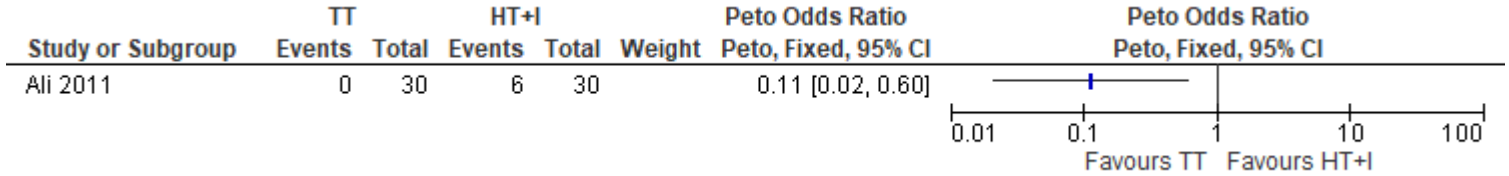
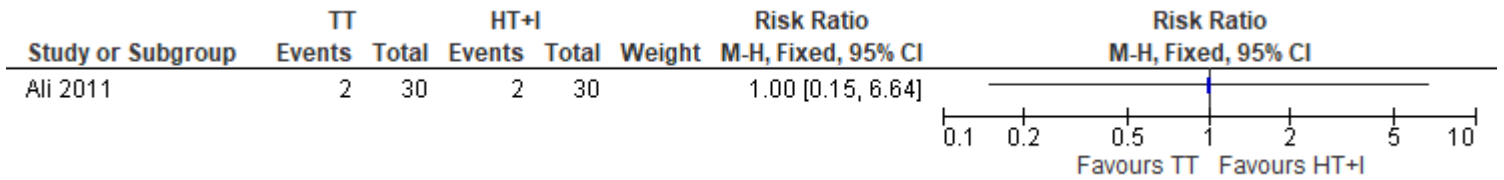


Figure 16: Hoarseness



Total thyroidectomy plus prophylactic central lymph node dissection (TT+PCND) versus total thyroidectomy only (TT)

Figure 17: Disease persistence

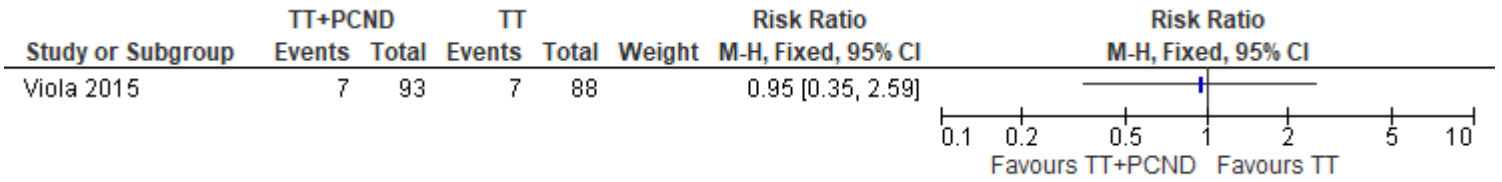


Figure 18: Need for additional ^{131}Iodine ablation

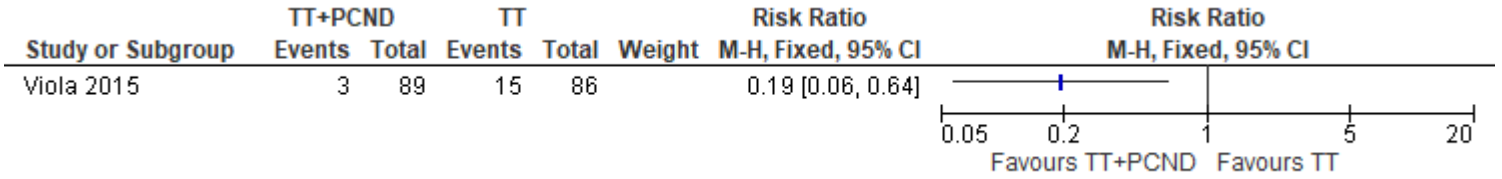


Figure 19: Hypoparathyroidism

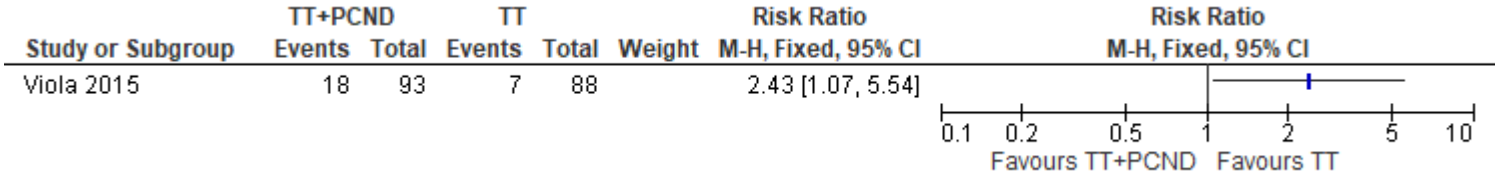
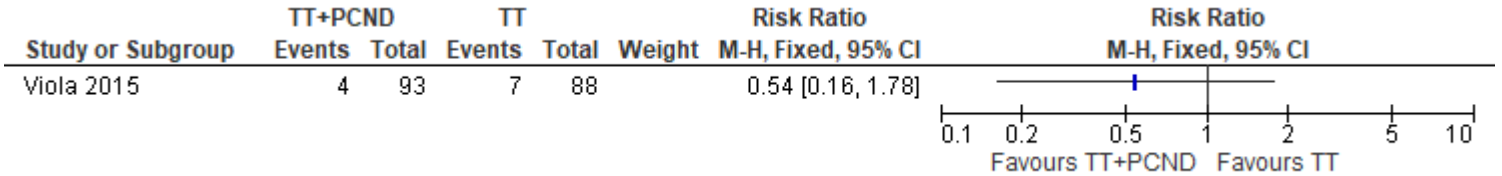


Figure 20: Recurrent laryngeal nerve palsy



Hemithyroidectomy alone (HT) versus active surveillance (AS)

Figure 21: Quality of life: SF-12 physical component summary

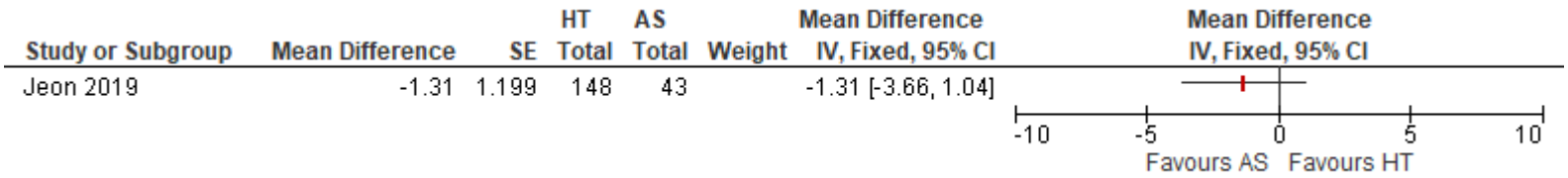
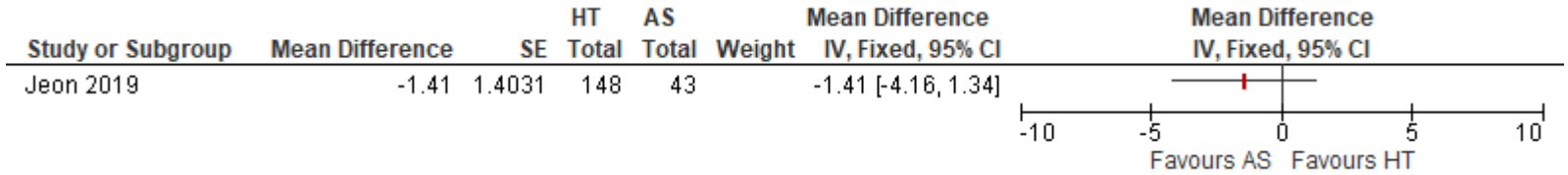
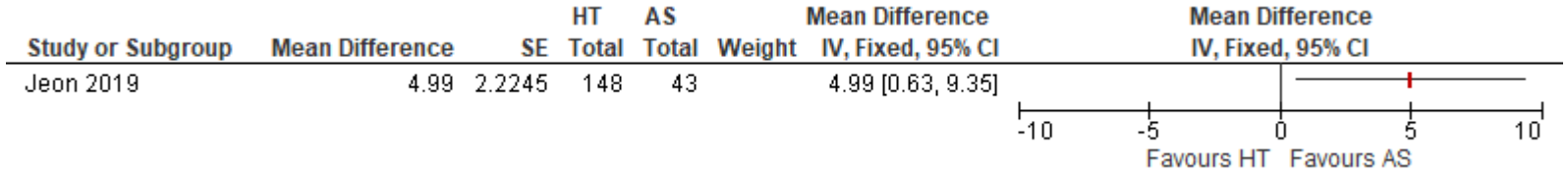


Figure 22: Quality of life: SF-12 mental component summary



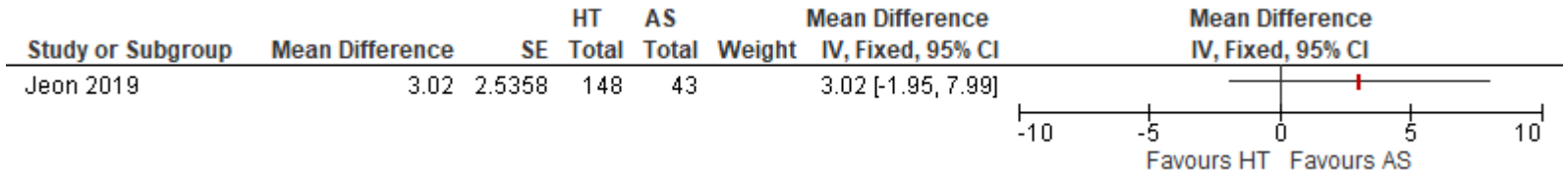
1

Figure 23: Quality of life: THYCA-QoL neuromuscular



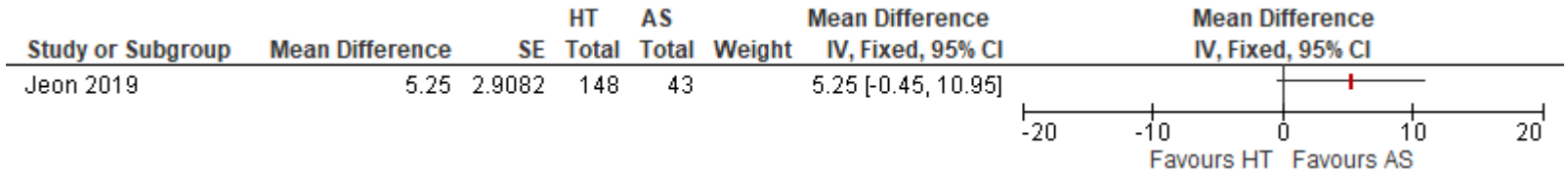
2

Figure 24: Quality of life: THYCA-QoL voice



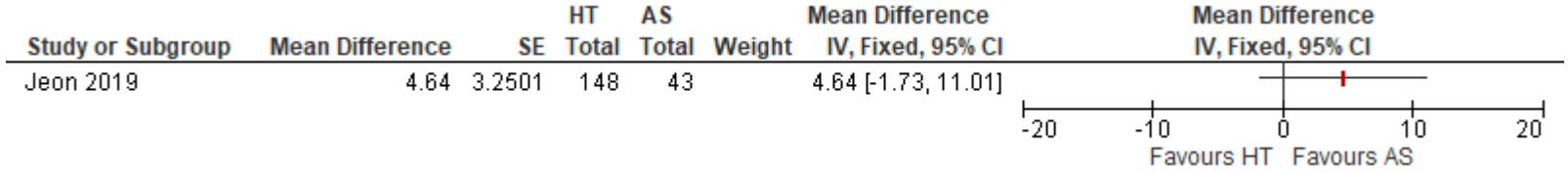
3

Figure 25: Quality of life: THYCA-QoL concentration



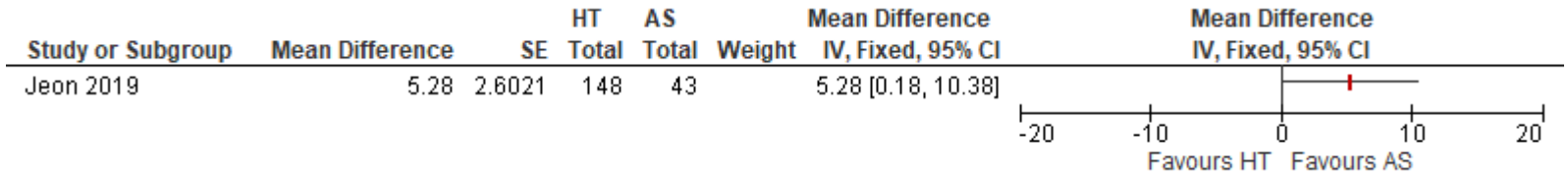
1

Figure 26: Quality of life: THYCA-QoL sympathetic symptoms



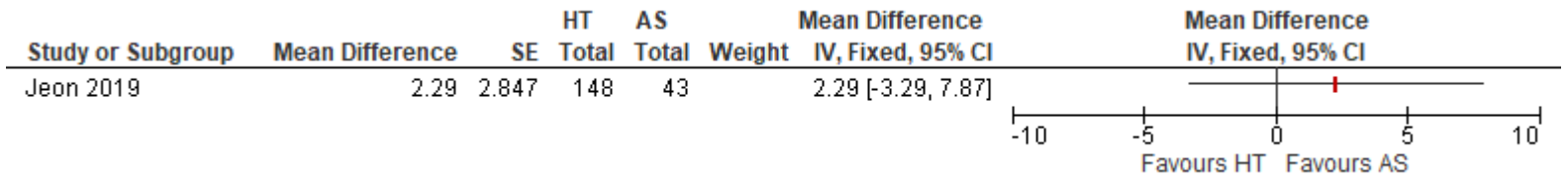
2

Figure 27: Quality of life: THYCA-QoL throat/mouth symptoms



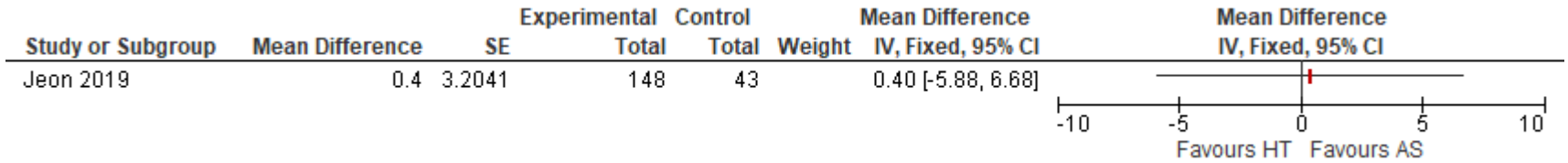
3

Figure 28: Quality of life: THYCA-QoL psychological symptoms



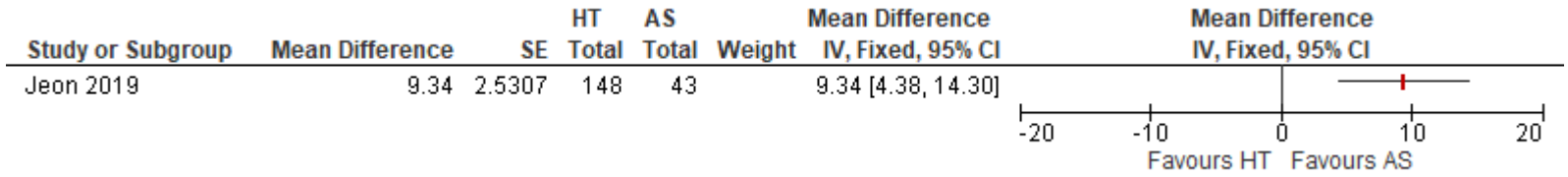
1

Figure 29: Quality of life: THYCA-QoL sensory



2

Figure 30: Quality of life: THYCA-QoL scar



3

Figure 31: Quality of life: THYCA-QoL symptoms of feeling chilly

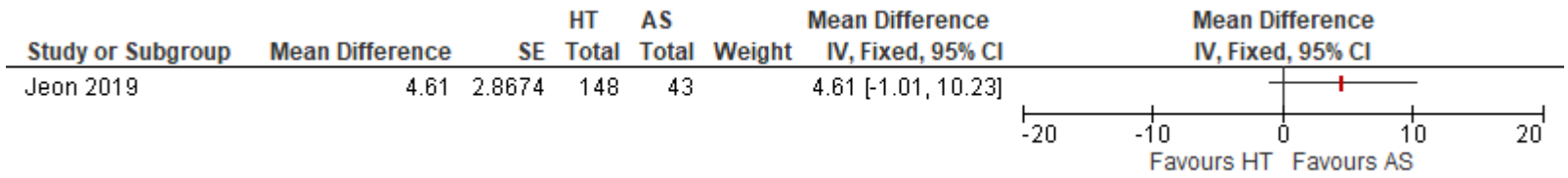


Figure 32: Quality of life: THYCA-QoL tingling hands/feet

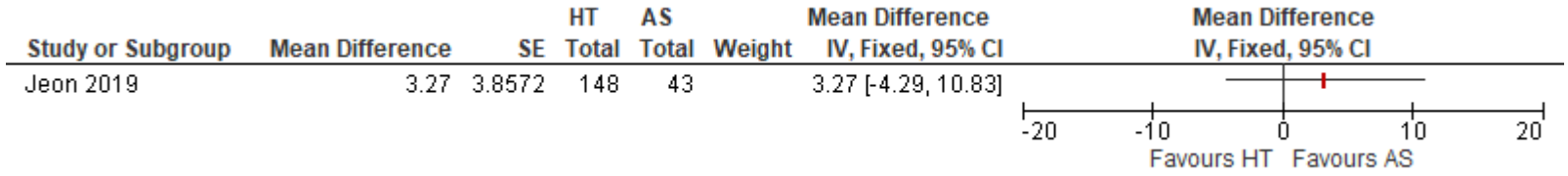


Figure 33: Quality of life: THYCA-QoL weight gain

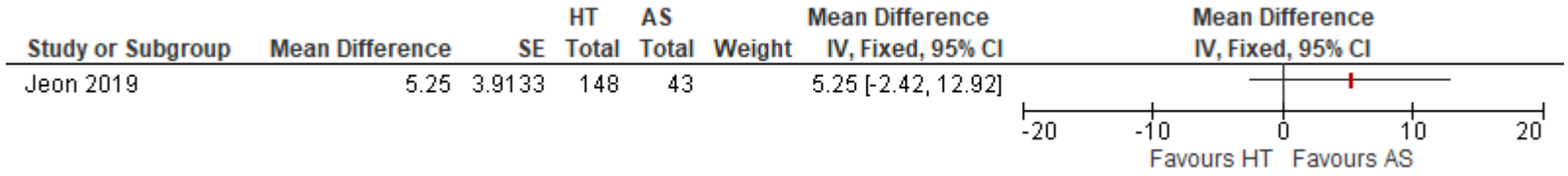
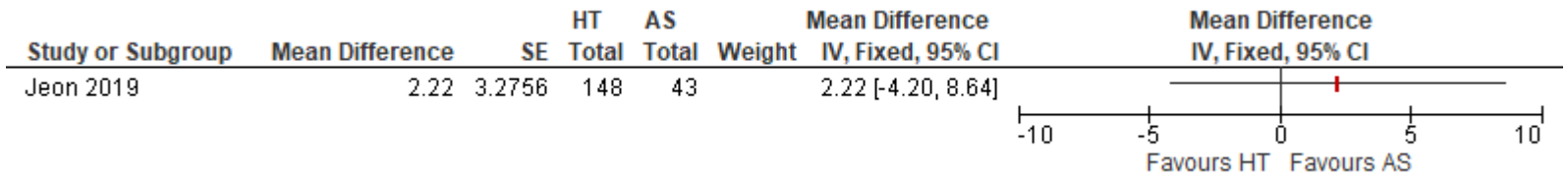
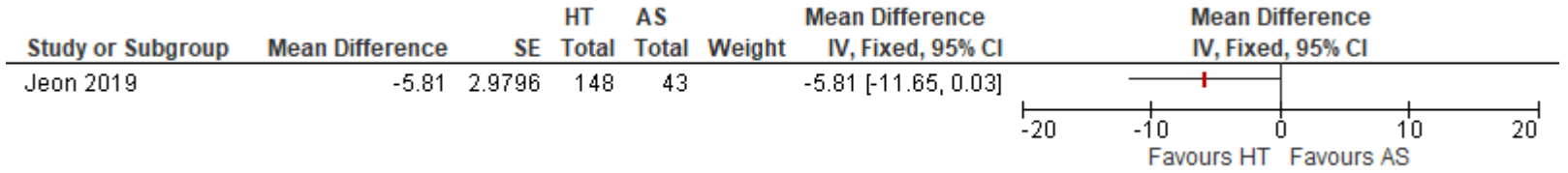


Figure 34: Quality of life: THYCA-QoL headache



1

Figure 35: Quality of life: THYCA-QoL less interest in sex



2

Appendix F – GRADE tables

Stage I disease

Table 14: Clinical evidence profile: thyroid surgery (total thyroidectomy alone or hemithyroidectomy alone) versus active surveillance

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Thyroid surgery	Active surveillance	Relative (95% CI)	Absolute		
Overall mortality (follow-up 5 years)												
1	observational studies	Very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	2308	15	HR 0.11 (0.09 to 0.13)		LOW	CRITICAL

1. Downgraded for serious risk of bias due to selection bias

Table 15: Clinical evidence profile: Total thyroidectomy versus active surveillance

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	TT	Active surveillance	Relative (95% CI)	Absolute		
Quality of life (follow-up 2 years)												
1	observational studies	Very serious[1]	NA	no serious indirectness	Serious imprecision[2]	none	45	94	-	MD: -0.354 (-0.529 to -0.179)	⊕OOO VERY LOW	CRITICAL

1. Downgraded for very serious risk of bias due to selection bias and incomplete outcome data
2. Downgraded for imprecision on the basis of optimal information size <350. It was not possible to assess on the basis of 0.5 x sd in the control group as such data were not provided in the paper

Table 16: Clinical evidence profile: hemithyroidectomy versus active surveillance

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Hemi	Active surveillance	Relative (95% CI)	Absolute		
Quality of life(follow-up 2 years)												
1	observational studies	Very serious[1]	no serious inconsistency	no serious indirectness	Serious imprecision[2]	none	45	94		MD: -0.141 (-0.248 to -0.141)	⊕OOO VERY LOW	CRITICAL

1. Downgraded for very serious risk of bias due to selection bias and incomplete outcome data
2. Downgraded for imprecision on the basis of optimal information size <350. It was not possible to assess on the basis of 0.5 x sd in the control group as such data were not provided in the paper

Table 17: Clinical evidence profile: total thyroidectomy with CND versus total thyroidectomy alone

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	TT + CND	TT only (low risk stratum)	Relative (95% CI)	Absolute		
cancer recurrence												
1	RCT	Serious[1]	NA	No serious indirectness	Very serious imprecision[2]	none	7/51 (13.7%)	8/50 (16%)	RR 0.86 (0.34 to 2.19)	22 fewer per 1000 (from 106 fewer to 190 more)	VERY LOW	CRITICAL
Recurrent laryngeal nerve palsy (binary)												
1	RCT	Serious[1]	NA	No serious indirectness	Very serious imprecision[2]	none	5/51 (9.8%)	3/50 (6%)	RR 1.63 (0.41 to 6.48)	38 more per 1000 (from 35 fewer to 329 more)	VERY LOW	CRITICAL
Hypoparathyroidism												
1	RCT	Serious[1]	NA	No serious indirectness	serious imprecision[2]	none	7/51 (13.7%)	13/50 (26%)	RR 0.53 (0.23 to 1.21)	122 fewer per 1000 (from 200 fewer to 55 more)	LOW	CRITICAL

Need for further treatment												
1	RCT	Serious[1]	NA	No serious indirectness	Very serious imprecision[2]	none	11/51 (21.6%)	11/50 (22%)	RR 0.98 (0.47 to 2.05)	4 fewer per 1000 (from 117 fewer to 231 more)	VERY LOW	CRITICAL
EAT-10 swallowing score (Better indicated by lower values)												
1	RCT	Serious[1]	NA	No serious indirectness	Serious imprecision[2]	none	30	30	-	MD 0.46 higher (1.36 lower to 2.29 higher)	LOW	CRITICAL
Calcium levels (Better indicated by lower values)												
1	RCT	Serious[1]	NA	No serious indirectness	Serious imprecision[2]	none	30	30	-	MD 0.1 higher (0.18 lower to 0.38 higher)	LOW	CRITICAL
PTH levels (Better indicated by lower values)												
1	RCT	Serious[1]	NA	No serious indirectness	Serious imprecision[2]	none	30	30	-	MD 1.5 higher (11.95 lower to 14.95 higher)	LOW	CRITICAL

1. Downgraded for serious risk of bias due to selection bias secondary to no reports of allocation concealment
2. Serious imprecision if the 95% CIs crossed one MID and very serious if they cross two MIDs. For binary outcomes the MIDs were defined as a RR/HR or OR of 0.8 and 1.2, and for continuous outcomes the MIDs were defined as ± half the standard deviation of the control group. For the EAT-10 swallowing score, the MID was ±2.08, based on the control group sd of 4.16. For Calcium levels, the MID was ±0.2738, based on the control group sd of 0.5477. For PTH, the MID was ±12.87, based on the control group sd of 25.87.

Table 18: Clinical evidence profile: hemithyroidectomy with CND versus hemithyroidectomy alone

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Hemi with pCND**	**Hemi alone**	**Relative (95% CI)**	**Absolute**		
cancer recurrence												
1	RCT	Very serious[1]	NA	No serious indirectness	Very serious imprecision[2]	none	3/82 (3.7%)	1/82 (1.2%)	RR 3 (0.32 to 28.25)	24 more per 1000 (from 8 fewer to 332 more)	VERY LOW	CRITICAL

Need for further treatment												
1	RCT	Very serious[1]	NA	No serious indirectness	Serious imprecision[2]	none	10/82 (12.2%)	0/82 (0%)	Peto OR 8.31 (2.32 to 29.73)	120 more per 1000) from 50 more to 200 more)-	VERY LOW	CRITICAL
Recurrent laryngeal nerve palsy												
1	RCT	Very serious[1]	NA	No serious indirectness	Very serious imprecision[2]	none	0/82 (0%)	1/82 (1.2%)	Peto OR 0.14 (0.00 to 6.82)	10 less per 1000 (from 50 fewer to 20 more)	VERY LOW	CRITICAL
Hypoparathyroidism												
1	RCT	Very serious[1]	NA	No serious indirectness	Very serious imprecision[2]	none	0/82 (0%)	0/82 (0%)	RD: 0.00 (-0.02 to 0.02	0 less per 1000 (from 20 fewer to 20 more)	VERY LOW	CRITICAL

1. Downgraded for serious risk of bias due to selection bias secondary to no reports of allocation concealment
2. Serious imprecision if the 95% CIs crossed one MID and very serious if they cross two MIDs. The MIDs were defined as a RR/HR or OR of 0.8 and 1.2. For the outcome with a risk difference, the imprecision was based on the calculation of total information size.

Disease stage unclear or mixed

Table 19Table 17: Clinical evidence profile: total thyroidectomy with CND versus total thyroidectomy alone 19: Clinical evidence profile: Total thyroidectomy alone (TT) versus hemithyroidectomy plus isthmusectomy alone (HT+I)

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	TT	HT+I	Relative (95% CI)	Absolute		
Cancer recurrence (follow-up 3 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/30 (0%)	6/30 (20%)	Peto OR 0.11 (0.02 to 0.6)	-	⊕⊕OO LOW	CRITICAL
Hoarseness (follow-up 3 months)												
1	randomised trials	very serious[2]	no serious inconsistency	serious[3]	very serious[4]	none	2/30 (6.7%)	2/30 (6.7%)	RR 1 (0.15 to 6.64)	0 fewer per 1000 (from 57 fewer to 376 more)	⊕OOO VERY LOW	CRITICAL

[1] No description of sequence generation or allocation concealment.
[2] No description of sequence generation or allocation concealment. No blinding. No description of how the outcome was measured.
[3] Outcome was hoarseness whereas the protocol outcome was recurrent laryngeal nerve palsy.
[4] Downgraded by 2 increments as the confidence interval crossed both default MIDs. The MIDs were defined as a RR/HR or OR of 0.8 and 1.25.

Table 20: Clinical evidence profile: Total thyroidectomy plus prophylactic central lymph node dissection (TT+PCND) versus total thyroidectomy alone (TT)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	TT+PCCND	TT	Relative (95% CI)	Absolute		
Disease persistence												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/93 (7.5%)	7/88 (8%)	RR 0.95 (0.35 to 2.59)	4 fewer per 1000 (from 52 fewer to 126 more)	⊕OOO VERY LOW	CRITICAL
Need for additional 131Iodine ablation (follow-up 5 years)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	3/89 (3.4%)	15/86 (17.4%)	RR 0.19 (0.06 to 0.64)	141 fewer per 1000 (from 63 fewer to 164 fewer)	⊕⊕OO LOW	CRITICAL

Hypoparathyroidism (follow-up 5 years)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	18/93 (19.4%)	7/88 (8%)	RR 2.43 (1.07 to 5.54)	114 more per 1000 (from 6 more to 361 more)	⊕OOO VERY LOW	CRITICAL
Recurrent laryngeal nerve palsy (follow-up 5 years)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	4/93 (4.3%)	7/88 (8%)	RR 0.54 (0.16 to 1.78)	37 fewer per 1000 (from 67 fewer to 62 more)	⊕OOO VERY LOW	CRITICAL

[1] No description of sequence generation or allocation concealment.
[2] Downgraded by 2 increments as the confidence interval crossed both default MIDs. The MIDs were defined as a RR/HR or OR of 0.8 and 1.25
[3] Downgraded by one increment as the confidence interval crossed one default MID. The MIDs were defined as a RR/HR or OR of 0.8 and 1.25

Table 21: Clinical evidence profile: Hemithyroidectomy alone (HT) versus active surveillance (AS)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**HT**	**AS**	**Relative (95% CI)**	**Absolute**		
QoL (neuromuscular) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 4.99 higher (0.63 to 9.35 higher)	⊕OOO VERY LOW	CRITICAL
QoL (physical component summary) (measured with: SF-12; range of scores: 0-100; Better indicated by higher values)												

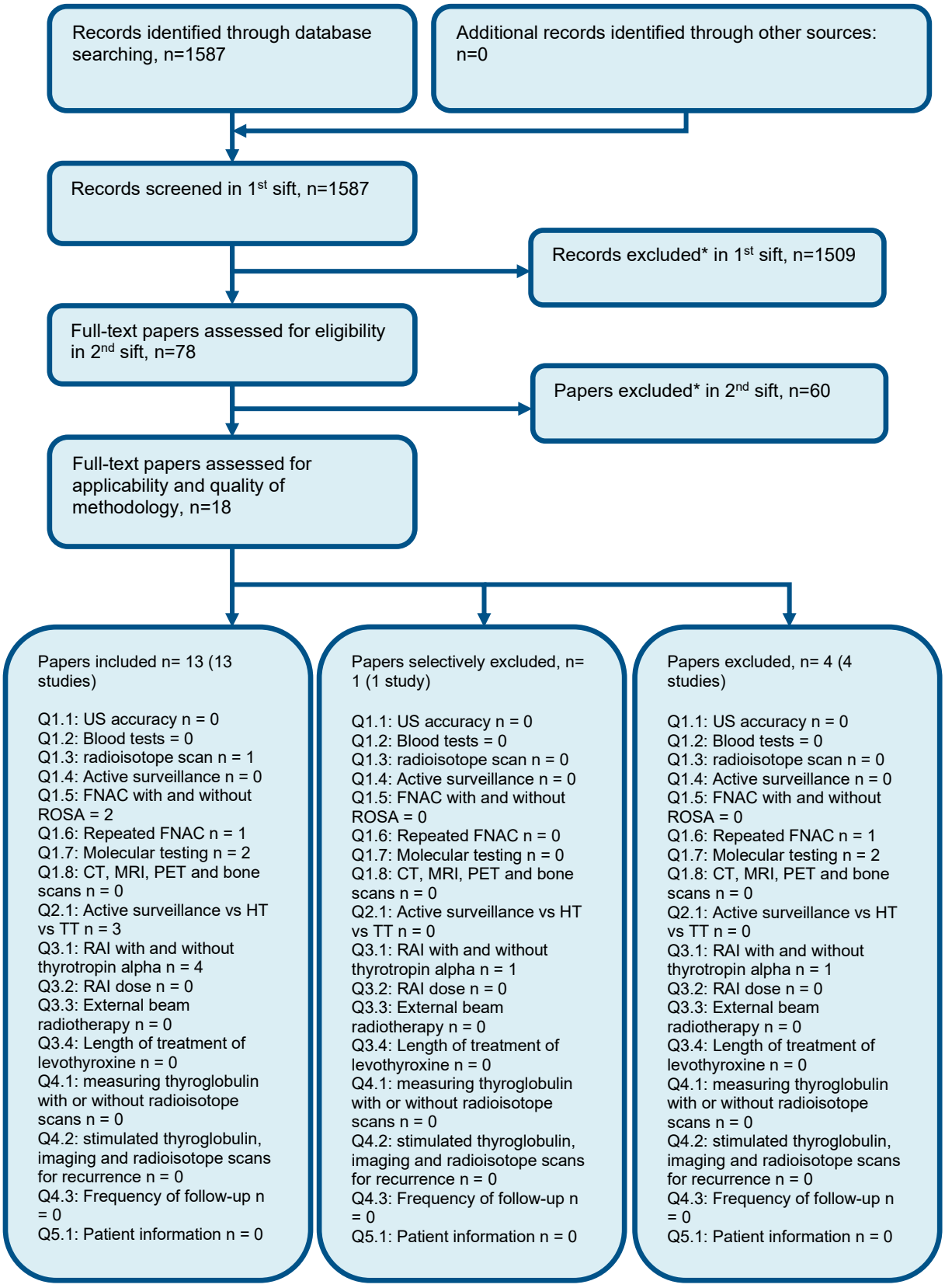
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 1.31 lower (3.66 lower to 1.04 higher)	⊕OOO VERY LOW	CRITICAL
QoL (mental component summary) (measured with: SF-12; range of scores: 0-100; Better indicated by higher values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 1.41 lower (4.16 lower to 1.34 higher)	⊕OOO VERY LOW	CRITICAL
QoL (voice) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 3.02 higher (1.95 lower to 7.99 higher)	⊕OOO VERY LOW	CRITICAL
QoL (concentration) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 5.25 higher (0.45 lower to 10.95 higher)	⊕OOO VERY LOW	CRITICAL
QoL (sympathetic symptoms) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 4.64 higher (1.73 lower to 11.01 higher)	⊕OOO VERY LOW	CRITICAL
QoL (throat/mouth) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 5.28 higher (0.18 to 10.38 higher)	⊕OOO VERY LOW	CRITICAL
QoL (psychological) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	148	43	-	MD 2.29 higher (3.29 lower to 7.87 higher)	⊕⊕OO LOW	CRITICAL
QoL (sensory) (measured with: THYCA-QoL; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	148	43	-	MD 0.4 higher (5.88 lower to 6.68 higher)	⊕⊕OO LOW	CRITICAL

QoL (scar) (measured with: THYCA-QoL; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	148	43	-	MD 9.34 higher (4.38 to 14.3 higher)	⊕⊕OO LOW	CRITICAL
QoL (felt chilly) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 4.61 higher (1.01 lower to 10.23 higher)	⊕OOO VERY LOW	CRITICAL
QoL (tingling hands/feet) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 3.27 higher (4.29 lower to 10.83 higher)	⊕OOO VERY LOW	CRITICAL
QoL (weight gain) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 5.25 higher (2.42 lower to 12.92 higher)	⊕OOO VERY LOW	CRITICAL
QoL (headache) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 2.22 higher (4.2 lower to 8.64 higher)	⊕OOO VERY LOW	CRITICAL
QoL (less interest in sex) (measured with: THYCA-QoL; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	148	43	-	MD 5.81 lower (11.65 lower to 0.03 higher)	⊕OOO VERY LOW	CRITICAL

[1] Downgraded for very serious risk of bias due to selection bias and blinding (Group allocation was determined by disease severity criteria. No blinding. Groups dissimilar for baseline characteristics).
[2] Downgraded by 1 increment as the confidence interval crossed one default MID. The MIDs were as follows. PCS: ±2.83, based on sd of control group sd of 5.66; MCS: ±4.09, based on sd of control group sd of 8.18; neuromuscular: ±7.03, based on sd of control group sd of 14.07; voice: ±6.41, based on sd of control group sd of 12.82; concentration: ±6.61, based on sd of control group sd of 13.22; sympathetic: ±8.55, based on sd of control group sd of 17.09; throat/mouth: ±5.79, based on sd of control group sd of 11.58; psychological: ±7.92, based on sd of control group sd of 15.84; sensory: ±6.89, based on sd of control group sd of 13.79; PCS: ±2.83, based on sd of control group sd of 5.66; scar: ±0, based on sd of control group sd of 0; Felt chilly: ±5.93, based on sd of control group sd of 11.85; tingling hands and feet: ±9.68, based on sd of control group sd of 19.35;

gained weight: ±8.20, based on sd of control group sd of 16.39; PCS: ±2.83, based on sd of control group sd of 5.66; headache: ±7.07, based on sd of control group sd of 14.14; less interest in sex: ±10.9, based on sd of control group sd of 21.79;

Appendix G – Economic evidence study selection



* Non-relevant population, intervention, comparison, design or setting; non-English language

1 # Appendix H – Economic evidence tables

Study	Kim 2019[62]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: CC (health outcome: None) **Study design:** Retrospective database analysis **Approach to analysis:** Descriptive statistical analyses (Kruskal-Wallis test for continuous variables, Pearson's chi-square and Fisher's exact test for categorical variables; all 2-tailed) **Perspective:** South Korean healthcare system **Time horizon:** 5 years **Treatment effect duration:** NA **Discounting:** Costs: NR Outcomes: NR	**Population:** Adults (≥18 years) with low-risk, advanced and recurrent differentiated thyroid cancer (papillary carcinoma) **Cohort settings:** Mean age: 48 Male: 24% N: 33 **Intervention 1:** Hemithyroidectomy **Intervention 2:** Total thyroidectomy **Intervention 3:** Total thyroidectomy with ipsilateral radical neck dissection **Intervention 4:** Total thyroidectomy with bilateral radical neck dissection and mediastinal dissection	**Total costs (mean per patient):** Intervention 1: £4,656 Intervention 2: £7,876 Intervention 3: £11,575 Intervention 4: £39,293 Incremental (2−1): (95% CI: NR; p=NR) Incremental (3−2): (95% CI: NR; p=NR) Incremental (4−2): (95% CI: NR; p=NR) **Currency & cost year:** 2015 South Korean won (presented here as 2015 UK pounds[(a)]) **Cost components incorporated:** Intervention (surgical admission), outpatient follow-up (outpatient visit, thyroid function test, thyroid ultrasound, neck	n/a	The cost of initial surgery, outpatient clinical treatment, and overall costs increase with increasing surgical extent and disease severity. **Analysis of uncertainty:** None

		CT, PET-CT), Radioiodine therapy (131-Iodine therapy, 131-Iodine full body scan)		
Data sources				
Health outcomes: NA **Quality-of-life weights:** NA **Cost sources:** NR.				
Comments				
Source of funding: National Research Foundation of Korea **Limitations:** Population included advanced and recurrent thyroid cancer patients who require radical neck dissection or mediastinal dissection and therefore the cost of high-dose RAI therapy should be considered in addition to the cost of surgery and outpatient follow-up. Patients were retrospectively categorized in each comparator group according to surgical extent which was also determined by disease severity: hence patient characteristics are different across the groups. Korean health system context. Clinical outcomes were not included. Discounting was not reported and QALYs were not included. Retrospective study that only included data from a single hospital of unclear representativeness or generalizability. Resource use and unit cost sources were not reported. Sensitivity analyses were not conducted and parameter uncertainty was not reported. **Other:** None				
Overall applicability:(b) Partially applicable **Overall quality:**(c) Potentially serious limitations				

Abbreviations: 95% CI= 95% confidence interval; CC= cost comparison; CT= computed tomography; ICER= incremental cost-effectiveness ratio; NA = not applicable; NR= not reported; PET = positron emission tomography; QALYs= quality-adjusted life years; RAI = radioactive iodine..
(a) Converted using 2015/2016 purchasing power parities[78]
(b) Directly applicable / Partially applicable / Not applicable
(c) Minor limitations / Potentially serious limitations / Very serious limitations

Study	**Oda 2017[76]**			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: CC (health outcome: None) **Study design:** Patient flow model	**Population:** Adults (≥18 years) with low-risk differentiated thyroid cancer (papillary microcarcinoma) **Cohort settings:**	**Total costs (mean per patient):** Intervention 1: £1,549 Intervention 2: £6,370 Incremental (2−1): £4,821 (95% CI: NR; p=NR)	n/a	Immediate surgery was 4.1 times more expensive than active surveillance **Analysis of uncertainty:** None

Approach to analysis: Cost comparison based on intention-to-treat basis based on Oda 2016[77] **Perspective:** Japanese health care insurance system **Time horizon:** 10 years **Treatment effect duration:**(a) NA **Discounting:** Costs: NR Outcomes: NR	Median age: 56 years Male: 12% N: 2,153 **Intervention 1:** Active surveillance (followed-up by ultrasound and blood tests at 6-months and 1-year) **Intervention 2:** Immediate surgery (total thyroidectomy with central node dissection or hemithyroidectomy with paratracheal dissection)	**Currency & cost year:** 2013 Japanese yen (presented here as 2013 UK pounds(b)) **Cost components incorporated:** Initial diagnosis (physician consultation, blood test, ultrasound, fine needle aspiration cytology), surgery (pre-operative examinations, surgery, anaesthesia, pathologic examination, and inpatient stay), follow-up care (physician consultation, blood test, ultrasound), medication (l-thyroxine, vitamin D supplements)		
Data sources				
Health outcomes: Not included **Quality-of-life weights:** Not included. **Cost sources:** Resource use and unit costs obtained from Kuma Hospital in Japan.				
Comments				
Source of funding: NR **Limitations:** Health and QoL outcomes were not included. Future costs were not discounted. Resource use and unit costs were obtained from one hospital with unclear representativeness or generalizability. Transient and permanent vocal cord paralysis (both potential outcomes of the surgery) were not included in the model. Patients chose whether undergo immediate surgery or be in active surveillance so baseline characteristics in the two arms are likely to be unbalanced; descriptive statistics were not reported. Sensitivity analyses were not conducted and parameter uncertainty was not reported. **Other:** None				
all applicability:(c) Partially applicable		**Overall quality:**(d) Potentially serious limitations		

Abbreviations: 95% CI= 95% confidence interval; CC= cost comparison; da= deterministic analysis; ICER= incremental cost-effectiveness ratio; NA = not applicable; NR= not reported; pa= probabilistic analysis; QALYs= quality-adjusted life years.
(a) Converted using 2013/14 purchasing power parities[78]
(b) Directly applicable / Partially applicable / Not applicable
(c) Minor limitations / Potentially serious limitations / Very serious limitations

Study	Lin 2020[67]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: CC (health outcome: None) **Study design:** Retrospective database analysis **Approach to analysis:** Statistical analysis (x^2 test, cox proportional hazard model, Kaplan-Meier curves) **Perspective:** Australian healthcare system perspective **Time horizon:** 3 years **Treatment effect duration:**[(a)] NA **Discounting:** Costs: NR Outcomes: NR	**Population:** Adults with low-risk differentiated thyroid cancer (papillary microcarcinoma) **Cohort settings:** Median age: 48 years Male: 19% N: 349 **Intervention 1:** Active surveillance (followed-up by ultrasound and blood tests at 6-months and 1-year) **Intervention 2:** Immediate surgery (total thyroidectomy with central node dissection or hemithyroidectomy with paratracheal dissection)	**Total costs (mean per patient):** Intervention 1: £1,524 • Initial cost: £450 • Annual cost: £358 Intervention 2: £5,177 • Initial cost: £4,843 • Annual cost: £111 Incremental (2−1): £3,653 **Currency & cost year:** 2017 Australian dollars (presented here as 2017 UK pounds[(b)]) **Cost components incorporated:** Initial diagnosis (endocrinologist consultation, blood test, ultrasound, fine needle aspiration cytology), surgery (pre-operative examinations, surgery, anaesthesia, pathologic examination, and inpatient	n/a	With a 3-year time horizon, the cost of active surveillance is lower than the cost of immediate treatment. Active surveillance surpasses the cost of immediate surgery after 16.2 years. **Analysis of uncertainty:** Decreasing the follow-up interval for active surveillance from twice to once a year halved the annual cost of active surveillance. Age has a big impact on the results of the analysis as younger patients have a higher risk of disease progression than older patients and therefore active surveillance is a much more costly strategy for people in their 20s, 30s and 40s than 50s, 60s and 70s).

		stay), follow-up care (endocrinologist consultation, blood test, ultrasound), medication (l-thyroxine, calcium, and vitamin D supplements)		
Data sources				
Health outcomes: Patient charts from the Endocrine Database at the University of Sydney Endocrine Surgery Unit were reviewed to verify patient characteristics, treatment details, complications, recurrences, recurrence-free survival and overall survival. Probability of operative intervention for patients receiving active surveillance was obtained from Oda 2017[76] **Quality-of-life weights:** Not included. **Cost sources:** The cost of surgery and hypothetic AS were derived from anonymized data provided by the clinical costing team from the Royal North Shore Hospital at the University of Sydney. Because active surveillance was not offered as a treatment option to patients at the Endocrine Surgery Unit at the University of Sydney, cost calculations for active surveillance were based on resource use for the program proposed by Oda 2017[76].				
Comments				
Source of funding: This research was supported without any funding. **Limitations:** Australian healthcare context. QALYs were not included. Discount rate was not reported. Outcomes were obtained from a single hospital of unclear representativeness or generalizability. Relative effectiveness was not included. Resource use estimates for active surveillance were based on resource use from Oda 2017[76] because active surveillance was not offered as a treatment option at the study site (therefore they are subject to the same limitations of Oda 2017). Descriptive statistics were not reported. Cost year was not reported and was assumed to be the final year of the database analysis. **Other:** None				
Overall applicability:[(c)] Partially applicable **Overall quality:**[(d)] Potentially serious limitations				

Abbreviations: 95% CI= 95% confidence interval; CC= cost comparison; ICER= incremental cost-effectiveness ratio; NA = not applicable; NR= not reported; pa= probabilistic analysis; QALYs= quality-adjusted life years.

(a) For studies where the time horizon is longer than the treatment duration, an assumption needs to be made about the continuation of the study effect. For example, does a difference in utility between groups during treatment continue beyond the end of treatment and if so for how long.

(b) Converted using 2017/2018 purchasing power parities[78]

(c) Directly applicable / Partially applicable / Not applicable

(d) Minor limitations / Potentially serious limitations / Very serious limitations

Appendix I – Excluded studies

I.1 Clinical studies

Table 19: Studies excluded from the clinical review

Study	Exclusion reason
Ahn 2020[2]	SR - references checked
Alabdrabalnabi 2020[3]	Literature review
Alhashemi 2016[4]	SR - references checked
Altedlawi albalawi 2020[6]	SR - references checked
Attene 2017[7]	SR - references checked
Bai 2018[8]	SR - references checked
Bojoga 2020[9]	SR - references checked
Bononi 2004[10]	Non RCT – case series
Burch 1995[11]	Literature review
Cabrera 2015[12]	Non RCT – case series
Caglia 2010[14]	Literature review
Caglia 2017[13]	SR - references checked
Carling 2007[16]	Literature review
Carling 2010[15]	Literature review
Caron 2006[17]	Literature review
Chan 2020[18]	SR - references checked
Chaukar 2010[19]	Literature review
Chen 2018[20]	SR - references checked
Chisholm 2009[21]	SR - references checked
Cho 2019[22]	SR - references checked
Cohen 2017[23]	Observational study with no adjustments for confounding; wrong comparators (diagnostic thyroidectomy)
Colombo, 2021[24]	Observational study with no adjustment for confounders

Cooper 2006[25]	Guidelines
De ceulaer 2012[26]	SR - references checked
Ding 2017[27]	Active surveillance paper so dropped down to observational: no adjustments for confounding
Dionigi 2014[28]	Literature review
Dong 2018[29]	SR - references checked
El-labban 2009[30]	Incorrect population (mixed diagnoses with no separate analysis for those with differentiated thyroid cancer).RCT but non-protocol comparison of two hemithyroidectomy techniques (conventional versus minimally-invasive video-assisted.
Fan 2012[31]	Literature review
Friedman 1990[33]	Literature review
Friedman 2002[32]	Literature review
Fu 2019[34]	Literature review
Gambale 2020[35]	SR - references checked
Gambardella 2016[36]	Literature review
Gartland 2018[37]	SR - references checked
Geramizadeh 2019[38]	Literature review
Gharib 2016[39]	Guidelines
Giuffrida 2012[40]	Literature review
Grani 2020[41]	Literature review
Grant 2014[42]	Literature review
Griffin 2017[43]	Non RCT: case series
Grossman 1997[44]	Literature review
Guo 2020[45]	SR - references checked
Hall 2017[46]	Did not include active surveillance as a comparator
Hassanain 2010[47]	Did not compare active surveillance to surgery
Hewitt 2006[48]	Non RCT - case series and commentary

Hu, 2021 [49]	Observational study with no adjustment for confounding
Huang 2013[50]	SR - references checked
Hughes 2011[51]	SR - references checked
Hughes 2018[52]	SR - references checked
Husson 2011[53]	SR - references checked
Ito 2003[58]	No comparison between intervention groups as initially assigned.
Ito 2010[55]	No multivariate analysis to compare active surveillance with surgery.
Ito 2011[57]	Observational study with no adjustment for confounding
Ito 2018[56]	SR - references checked
Ito 2020[54]	SR - references checked
Jackson 2014[59]	SR - references checked
Kong 2019[63]	Observational study with no appropriate multivariable analysis. Although a multivariable analysis was performed, this was in order to elucidate the factors influencing the choice of treatment, rather than to adjust the effects of treatments for plausible confounders
Kuo 2017[64]	Observational study compared untreated to treated, but the treated group were not specified and unclear if the treatments matched those in the protocol
Lee 2019[65]	Diagnostic accuracy study
Li, 2020 [66]	Total versus subtotal thyroidectomy not a protocol comparison
Lin 2020[67]	Case series - even though active surveillance is mentioned in the title this was only considered hypothetically and no patients received AS
Lin 2020[68]	Compared untreated to treated, but the treatments included non-protocol treatments like RAI and EBR.
Liu 2019[69]	No data presented for the multivariable associations between surgery/AS and outcome
Nagarkatti 2013[72]	Observational study no multivariable analysis performed
Nakamura 2020[73]	Observational study with no adjustments for confounding
Nakamura, 2020 [73]	Already excluded

Neagoe 2017[75]	SR - references checked
Oda 2016[77]	Observational study with no adjustments for confounding
Pan 2017[79]	SR - references checked
Parker 2017[80]	SR - references checked
Paschke 2015[81]	SR - references checked
Pisanu 2013[82]	SR - references checked
Qu 2015[83]	SR - references checked
Qu 2016[84]	SR - references checked
Raffaelli 2012[85]	Observational study with no adjustments for confounding
Ren 2017[86]	Non-RCT – case series
Rodriguez-martin 2018[87]	Cost effectiveness evaluation
Rosario 2019[88]	No multivariable analysis performed
Roti 2008[89]	SR - references checked
Ruggiero 2008[90]	Literature review
Sakai 2019[91]	Observational study with no adjustments for confounding
Sanabria 2021[92]	Literature review
Saravana-bawan 2020[93]	SR - references checked
Schmidbauer 2017[94]	SR - references checked
Sgourakis 2008[95]	SR - references checked
Shan 2012[96]	SR - references checked
Shan 2012[98]	Incorrect population. RCT but non-protocol comparison.
Shan 2019[97]	SR - references checked
Shen 2014[99]	SR - references checked
Sieda 2020[100]	Not review population. Not guideline condition
Singer 1996[101]	Guidelines
Sipos 2008[102]	Literature review
Smulever 2019[104]	Observational study with no adjustments for confounding

Son 2015[105]	SR - references checked
Sonkar 2010[106]	SR - references checked
Su 2018[107]	SR - references checked
Su 2019[108]	SR - references checked
Sugitani 2019[109]	Survey of treatment practices rather than effects of treatments on outcomes
Sun 2014[110]	SR - references checked
Sun 2015[111]	SR - references checked
Sun 2020[112]	SR - references checked
Sywak 2004[113]	Literature review
Sywak 2008[114]	Incorrect population (mixed diagnoses with no separate analysis for those with differentiated thyroid cancer).RCT but non-protocol comparison of two hemithyroidectomy techniques (conventional versus minimally-invasive).
Tan 2008[115]	Literature review
Tunca 2008[116]	RCT but non-protocol comparison of conventional versus radio-guided thyroidectomy.
Tuttle 2018[117]	Literature review
Udelsman 1996[118]	Discussion article
Van gerwen 2020[119]	SR - references checked
Vargas-pinto 2019[120]	SR - references checked
Vasileiadis 2018[121]	SR - references checked
Venkat 2013[122]	Literature review
Venkatesh 2017[123]	Health economics analysis paper
Vuong 2019[125]	SR - references checked
Walgama 2020[126]	SR - references checked
Wang 2013[128]	SR - references checked
Wang 2015[127]	Cost effectiveness analysis
Wang 2015[129]	SR - references checked

Wang 2015[131]	SR - references checked
Wang 2021[130]	SR - references checked
White 2007[132]	Literature review
White 2007[133]	SR - references checked
Witt 2013[134]	Literature review
Wojtczak 2010[135]	Literature review
Won 2018[136]	SR - references checked
Wong 2020[137]	Subjects did not have a diagnosis of differentiated thyroid cancer. Comparison was active surveillance versus non-protocol intervention (fine needle aspiration).
Wong 2020[138]	Cost effectiveness analysis
Yang 2015[139]	Incorrect population. RCT but non-protocol comparison.
Yi 2017[140]	SR - references checked
Yip 2016[141]	Literature review
Yuk-wah liu 2016[142]	Literature review
Zhan 2019[143]	SR - references checked
Zhang 2020[144]	SR - references checked
Zhang, 2021 [145]	Did not evaluate protocol comparisons; evaluated ambulatory thyroidectomy vs non-ambulatory thyroidectomy instead.
Zhao 2017[146]	SR - references checked
Zhao 2017[147]	SR - references checked
Zheng 2018[148]	SR - references checked
Zhu 2013[149]	SR - references checked

I.2 Health Economic studies

Published health economic studies that met the inclusion criteria (relevant population, comparators, economic study design, published 2005 or later and not from non-OECD country or USA) but that were excluded following appraisal of applicability and methodological quality are listed below. See the health economic protocol for more details.

None.

Appendix J – Research recommendations

J.1.1 Research recommendation

For people with stage 1 differentiated thyroid cancer, what is the clinical and cost effectiveness of active surveillance compared with surgery?

J.1.2 Why this is important

In patients with stage 1 differentiated thyroid cancer, there is genuine clinical uncertainty as to whether it is better to use active surveillance or to use surgery (total- or hemithyroidectomy). At present, no randomised controlled trials exist, although some observational trials suggest active surveillance may offer some benefits for patients. Until RCT evidence is available, it is not possible to make strong recommendations for people with stage I disease to have active surveillance, even though the clinical consensus is that surgery may be over-used and may cause more harm than good in some patients. There is thus a strong need for an RCT in this area. Because of the genuine uncertainty in this area, there should be few ethical issues around randomising patients to either surgery or active surveillance, given no strong current evidence that either is better.

J.1.3 Rationale for research recommendation

Importance to 'patients' or the population	At present many stage 1 patients undergo surgery because there is insufficiently robust evidence to suggest that active surveillance is a better option. An RCT in this area would be of great importance to patients by helping to resolve a very important question, helping to direct clinical practice on the basis of evidence. If active surveillance can be shown by robust experimental evidence to be a safe and effective option for stage 1 patients, then patients will avoid unnecessary surgery, feel less anxious when doing so, and gain better health outcomes overall.
Relevance to NICE guidance	The efficacy of active surveillance has been considered in this guideline, but we did not find any RCTs evaluating it. The development of such RCTs is therefore required.
Relevance to the NHS	If active surveillance can be shown to be cost effective for stage 1 patients this may reduce costs, morbidity and bed use.
National priorities	None known
Current evidence base	There is currently no RCT evidence.
Equality considerations	None known

J.1.4 Modified PICO table

Population	People with stage 1 differentiated thyroid cancer
Intervention	Active surveillance
Comparator	Surgery
Outcome	Quality of life, recurrence, progression, mortality
Study design	RCT
Timeframe	Long term
Additional information	None