National Institute for Health and Care Excellence

FINAL

Thyroid disease: assessment and management

[P] Management of non-malignant thyroid enlargement

NICE guideline NG145

Intervention evidence review underpinning recommendations 1.9.7 to 1.9.12 in the guideline. See also evidence review F

2019

FINAL

Developed by the National Guideline Centre, hosted by the Royal College of Physician

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, where appropriate, their carer or guardian.

Local commissioners and 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: 978-1-4731-3595-6

Contents

1

1 Management of thyroid enlargement

1.1 Review question: Which people with non-malignant thyroid enlargement should be referred for surgery?

Review question: What is the clinical and cost effectiveness of non-surgical treatments (for example radiofrequency ablation, high intensity focused ultrasound(HIFU)) for non-malignant thyroid enlargement?

1.2 Introduction

In many patients, thyroid enlargement requires no surgical intervention and may be managed according to the patients' thyroid function. Thyroidectomy may be performed in the management of non-malignant thyroid disease. Recognised indications for surgery include patient preference, failed medical management (in the context of concomitantly deranged thyroid function), mass effect and cosmetic embarrassment. Careful discussion is warranted to ensure patients understand the sequelae and risks of surgery, which will differ with the extent of the operation. Robot assisted surgery remains uncommon in the UK.

In recent years percutaneous ablation techniques have been used in the treatment of thyroid malignancies, with many of these deriving from ablation of other solid organ tumours. Techniques include radiofrequency ablation, microwave ablation, laser ablation, and high intensity focused ultrasound. Ethanol ablation has also been described but tends to be used in the ablation of thyroid cysts when aspiration has failed. The use of percutaneous ablation by any method in the management of benign disease is a new, but growing, development and is gaining traction in centres across the UK. This review, whilst recognising the novelty of these applications in non-malignant thyroid disease, seeks to assess the current evidence for them whilst also outlining the situations in which surgical (and by extension these non-surgical percutaneous alternatives) treatment is warranted.

1.3 PICO table

For full details see the review protocol in Appendix A:.

Table 1: PICO characteristics of review question

Population	People with non-malignant thyroid enlargement
Interventions	Surgery Radiofrequency ablation High intensity focused ultrasound Ethanol ablation Radioactive iodine ablation Levothyroxine therapy Microwave ablation Laser ablation Monitoring only Placebo/sham
Comparisons	Any of the above vs any other
Outcomes	**Critical** • Mortality (dichotomous, ≥1 year)

	• Quality of life (continuous) **Important** • Percent change in size of nodule/goitre (continuous) • Malignancy (dichotomous) • Hypothyroidism (dichotomous) • Hyperthyroidism (dichotomous) • Hypoparathyroidism (dichotomous) • Recurrent laryngeal nerve damage (dichotomous) • Bleeding (dichotomous) • Infection (dichotomous) • Arrhythmia (dichotomous) • Osteoporosis (dichotomous) • Pain (continuous) • Compressive symptoms (continuous) • Patient/family/carer experience (continuous)
Study design	RCTs preferred, if no RCTs available (on an intervention by intervention basis) to consider non-randomised cohort studies in which key confounders (age, sex, co-existing conditions, size of nodule, type of nodule) are addressed, either through restriction or appropriate matching/statistical adjustment

1.4 Clinical evidence

1.4.1 Included studies

Twenty studies were included in the review;[1, 5-8, 13, 21, 24, 34-36, 39-42, 45, 48, 50, 53, 54] these are summarised in Table 2 below. Evidence from these studies is summarised in the clinical evidence summary below (Table 3).

One Cochrane review in this area was identified[3], the studies included in this review were checked against the protocol and included as appropriate.

All studies were in adults (between 18 and 65).

Eleven RCTs were found comparing levothyroxine with either placebo or follow-up only, in people with non-cystic nodules. Two RCTs were found comparing radiofrequency ablation with follow-up only, in people with non-cystic nodules. One RCT was found with three arms comparing levothyroxine, laser ablation and follow-up only in people with non-cystic nodules. One RCT was found comparing ethanol ablation, with levothyroxine in people with non-cystic nodules. One RCT was found comparing microwave ablation with surgery. Two RCTs were found comparing radiofrequency ablation with ethanol ablation, in people with cystic nodules.

One non-randomised study was found comparing radiofrequency ablation with surgery, in people with non-specified types of nodules. One non-randomised study compared HIFU with surgery in people with non-cystic nodules.

No evidence was found comparing radioactive iodine ablation with any other intervention. No evidence was found comparing surgery or ethanol ablation with a non-active intervention.

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:.

1.4.2 Excluded studies

See the excluded studies list in Appendix J:.

1.4.3 Summary of clinical studies included in the evidence review

Table 2: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Baek 2015[1]	Radiofrequency ablation, n = 22, 1 dose, 40-80 W RF power, moving-shot technique Percutaneous ethanol injection, n = 24, 1 dose, 99% ethanol, volume ~50% of aspirated	Adults (mean age 51, SD 15) Euthyroid Nodule (predominantly cystic (all 50-90% cystic component), cosmetic/compressive symptoms) South Korea	Nodule volume Compressive symptoms 6 month follow-up	
Bennedbaek 1998[5]	Percutaneous ethanol injection, n = 25, 1 dose, 98% ethanol, volume 20-50% of pre-treatment nodule volume Levothyroxine, n = 25, 1.5 µg/kg, titrated to TSH 0.1-0.4 (mU/L) after 1 month	Adults (mean age 44, range 37-52) Euthyroid Nodule (benign on FNAB, cold, causing discomfort, solid) Denmark	Nodule volume Compressive symptoms Pain Hyperthyroidism 12 month follow-up	
Cesareo 2010[7]	Levothyroxine, n = 36, 2 µg/kg body weight Follow-up only, n = 35	Adults (mean age 36, SD 10) Euthyroid Multinodule (colloid, hypofunctioning) Italy	Nodule volume Thyroid volume 12 month follow-up	Area of iodine deficiency
Cesareo 2015[8]	Radiofrequency ablation, n = 42, 1 dose, 60 W RF power, moving-shot technique Follow-up only, n = 42	Adults (mean age 55, SD 13) Euthyroid Nodule (solid, cosmetic/compressive symptoms or volume >5ml/diameter >2cm + growing) Italy	Nodule volume Hypothyroidism Hyperthyroidism Compressive symptoms 6 month follow-up	
Deandrea	Radiofrequency	Adults (mean age	Nodule volume	

Study	Intervention and comparison	Population	Outcomes	Comments
2015[13]	ablation, n = 40, 1 dose, moving-shot technique Follow-up only, n = 40	52, SD 12) Euthyroid Nodule (solid, benign as confirmed by FNAB, volume 10-20ml, pressure or cosmetic symptoms) Italy, South Korea	Compressive symptoms 6 month follow-up	
Gharib 1987[21]	Levothyroxine, n = 28, daily dose of 3 µg/kg body weight Placebo, n = 25	Adults (mean age ~45, SD 16) Nodule (benign as confirmed by FNAB, 43% solid) USA	Nodule volume 6 month follow-up	
Grussendorf 2011[24]	Levothyroxine, n = 206, initial 75 µg/d, titrated to a TSH between 0.2 and 0.8 from 3 months Placebo, n = 199	Adults (mean age 47, range 44.6-48.5) Euthyroid Nodule (at least 1 cm in diameter, ≤20% cystic component, non-malignant according to guidelines) Germany	Nodule volume Arrhythmias 12 month follow-up	'Majority' of participants achieved TSH suppression, exact proportion not stated
Lang 2018[34]	High intensity focused ultrasound, n = 77 Surgery, open lobectomy, n = 77	Adults (mean age 50, SD 14) Euthyroid Nodule (index nodule causing symptoms, 10-50mm, predominantly solid) China	Nodule volume Hypothyroidism RLN damage Bleeding Infection 6 month follow-up	Non-randomised study, propensity score matched for age, sex, BMI, TSH, size of nodule, volume of nodule
Larijani 1999[36]	Levothyroxine, n = 32, daily dose of 1.5 to 2.0 µg/kg body weight	Adults (mean age 32.9, SD 10.8) Nodule (40% of nodules were	Nodule volume 12 month follow-up	

Study	Intervention and comparison	Population	Outcomes	Comments
	Placebo, n = 30	solid, benign as confirmed by FNAB, mean starting volume 13ml) Iran		
Larijani 2005[35]	Levothyroxine, n = 31, 1.5-2 µg/kg, aimed at complete TSH suppression (<0.1 µIU/ml) Placebo, n = 27	Adults (mean age 36, SD 10) Nodule (benign as confirmed by FNAB) Iran	Nodule volume 24 month follow-up	
Ozkaya 2010[6]	Levothyroxine, n = 35, 50-100mg/day Follow-up only, n = 27	Adults (mean age not stated) Euthyroid Nodule (benign by FNAB, <2cm, non-cystic) Turkey	Nodule volume 12 month follow-up	
Papini 1993[39]	Levothyroxine, n = 54, daily dose of 2.0 ug/kg initially, then titrated to TSH suppression Placebo, n = 56	Adults (mean age 43, SD 11) Euthyroid Nodule (colloid, volume >1ml, solid, hypofunctioning/normal) Italy	Nodule volume 12 month follow-up	TSH suppression achieved in 88% of levothyroxine group
Papini 1998[41]	Levothyroxine, n = 42, daily dose of 2.0 µg/kg Follow-up only, n = 41	Adults (mean age 42, SD 13) Euthyroid Nodule (colloid, greatest diameter 10-30mm, hypofunctioning) Italy	Nodule volume 5 year follow-up	TSH suppression achieved in 48% of levothyroxine group
Papini 2007[40]	Levothyroxine, n = 21, daily dose of 1.5 µg/kg, titrated at 1 month to TSH <0.3 µIU/ml	Adults (mean age 47, SD 7) Euthyroid	Nodule volume Compressive symptoms 12 month follow-	

Study	Intervention and comparison	Population	Outcomes	Comments
	Laser ablation, n = 21, 1 session for all participants, US guided, 1-4 illuminations, mean 1200J/ml of thyroid tissue Follow-up only, n = 20	Nodule (non-cystic, >5ml, hypofunctioning, benign per FNAB) Italy	up	
Reverter 1992[42]	Levothyroxine, n = 20, daily dose of 100 µg/d for 2 weeks, then 200 µg/d and titrated to TSH <0.1 mIU/ml Follow-up only, n = 20	Adults (mean age 40, SD 10) Nodule (majority solid or mixed, hypofunctoining, benign per FNAB) Spain	Nodule volume 12 month follow-up	TSH suppression achieved in 100% of levothyroxine group
Sung 2013[45]	Radiofrequency ablation, n = 25, 1 dose, 50-70 W RF power, moving-shot technique Percutaneous ethanol injection, n = 25, 1 dose, 99% ethanol, volume ~50% of aspirated	Adults (mean age 42, SD 10) Euthyroid Nodule (cystic (cystic portion >90%), cosmetic/compression symptoms, benign by FNAB) South Korea	Nodule volume Compressive symptoms Pain 6 month follow-up	
Wemeau 2002[48]	Levothyroxine, n = 64, daily dose of 2.5 to 2.0 µg/kg body weight, titrated to TSH <0.3µIU/ml after 4 weeks Placebo, n = 59	Adults (mean age ~39, SD 9) Euthyroid Nodules (benign as confirmed by FNAB, non-cystic cold nodules only) France	Nodule volume Hyperthyroidism 18 month follow-up	
Yue 2016[50]	Radiofrequency ablation, n = 108, moving shot technique Surgery, n = 108, hemithyroidectomy	Adults (mean age 51, SD 13) Euthyroid Nodules (symptoms or anxiety over malignancy)	Quality of life 6 month follow-up	Non-randomised study Surgical cohort included some people with uncertain malignancy status

Study	Intervention and comparison	Population	Outcomes	Comments
		China		
Zelmanovitz 1998[53]	Levothyroxine, n = 24, initial 2.5-3.0 µg/kg, adjusted after 4 weeks to TSH <0.3 µIU/ml Placebo, n = 27	Adults (mean age 43, SD 12) Nodules (hypofunctioning, non-cystic) Brazil	Nodule volume 12 month follow-up	Suppression maintained in 86% of participants
Zhi 2018[54]	Microwave ablation, n = 30, excluded those (n = 2) with treatment failure (VRR <50% at 6 months) Surgery, n = 30 hemithyroidectomy for those without a >1cm contralateral nodule (n = 24), those needing total excluded	Adults (mean age 53, SD 10) Nodules (mix of solid and cystic, >2cm in at least one dimension, causing symptoms or rapidly growing) China	Nodule volume Hoarseness Pain Symptoms 12 month follow-up	

See Appendix D: for full evidence tables.

1.4.4 Quality assessment of clinical studies included in the evidence review

Table 3: Clinical evidence summary: Levothyroxine vs placebo/follow-up only, non-cystic nodules

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo or follow-up only	Risk difference with T4 (95% CI)
Nodule volume ml	735 (11 studies) 6-60 months	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean nodule volume in the control groups was 2.5 ml	The mean nodule volume in the intervention groups was 0.79 lower (1.13 to 0.45 lower)
Reduction in nodule volume %	405 (1 study) 12 months	⊕⊕⊕⊕ HIGH		The mean reduction in nodule volume in the control groups was 5.2 %	The mean reduction in nodule volume in the intervention groups was 6.9% higher (0.91 lower to 14.71 higher)
Thyroid volume ml	71 (1 study) 12 months	⊕⊕⊕⊖ MODERATE2 due to imprecision		The mean thyroid volume in the control groups was 20 ml	The mean thyroid volume in the intervention groups was 9.4 lower (14.53 to 4.27 lower)
Hyperthyroidism 'Severe' (nil else specified)	123 (1 study) 18 months	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 0.92 (0.06 to 14.41)	17 per 1000	1 fewer per 1000 (from 16 fewer to 228 more)
Improvement in symptoms	29 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	Peto OR 7.43 (0.44 to 125.12)	0 per 1000	130 more per 1000 (from 70 fewer to 330 more)3
Arrhythmias Reported as AF	405 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	Peto OR 0.13 (0.01 to 2.09)	10 per 1000	9 fewer per 1000 (from 10 fewer to 11 more)

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo or follow-up only	Risk difference with T4 (95% CI)
2 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs 3 Zero events in control arm					

Table 4: Clinical evidence summary: Radiofrequency ablation vs follow-up only, non-cystic nodules

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with follow-up only	Risk difference with RFA (95% CI)
Nodule volume ml	164 (2 studies) 6 months	⨁⨁⨁⨁ HIGH		The mean nodule volume in the control groups was 21.5 ml	The mean nodule volume in the intervention groups was 10.8ml lower (12.14 to 9.45 lower)
Hypothyroidism	84 (1 study) 6 months	⨁⨀⨀⨀ VERY LOW2,3 due to risk of bias, imprecision	Not estimable	0 per 1000	not estimable1
Hyperthyroidism	84 (1 study) 6 months	⨁⨀⨀⨀ VERY LOW2,3 due to risk of bias, imprecision	Peto OR 7.39 (0.15 to 372.38)	0 per 1000	23 more per 1000 (from 40 fewer to 90 more)4
Compressive symptoms VAS (0-10, higher is worse)	164 (2 studies) 6 months	⨁⨁⨁⨀ MODERATE2 due to risk of bias		The mean compressive symptoms in the control groups was 3.1	The mean compressive symptoms in the intervention groups was 2.8 lower (3.3 to 2.31 lower)
1 Zero events in either arm 2 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias 3 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs					

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with follow-up only	Risk difference with RFA (95% CI)
4 Zero events in control arm					

Table 5: Clinical evidence summary: Laser ablation vs follow-up only, non-cystic nodules

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Follow-up only	Risk difference with LA (95% CI)
Change in nodule volume ml	41 (1 study) 12 months	⨁⨁⨁⨀ MODERATE1 due to imprecision		The mean change in nodule volume in the control groups was 0.7 ml	The mean change in nodule volume in the intervention groups was 5.9 lower (7.54 to 4.26 lower)
Improvement in symptoms	41 (1 study) 12 months	⨁⨀⨀⨀ VERY LOW1,3 due to risk of bias, imprecision	Peto OR 24.5 (5.9 to 101.6)		810 more per 1000 (from 60 more to 1000 more)2

1 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
2 Zero events in control arm
3 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Table 6: Clinical evidence summary: Laser ablation vs levothyroxine, non-cystic nodules

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4	Risk difference with LA (95% CI)
Change in nodule	42	⨁⨁⨁⨀		The mean change in nodule volume in	The mean change in nodule volume in the

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4	Risk difference with LA (95% CI)
volume ml	(1 study) 12 months	MODERATE1 due to imprecision		the control groups was -0.6 ml	intervention groups was 4.6ml lower (6.25 to 2.95 lower)
Improvement in symptoms	42 (1 study) 12 months	⊕⊕⊖⊖ LOW2 due to risk of bias	RR 6.09 (1.64 to 22.62)	133 per 1000	677 more per 1000 (from 85 more to 1000 more)

1 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
2 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Table 7: Clinical evidence summary: Percutaneous ethanol injection vs levothyroxine, non-cystic nodules

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4	Risk difference with PEI (95% CI)
Reduction in nodule volume % changes	50 (1 study) 12 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean reduction in nodule volume in the control groups was 9 %	The mean reduction in nodule volume in the intervention groups was 38 higher (19.08 to 56.92 higher)
Compressive symptoms Any improvement	44 (1 study) 12 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision	RR 1.55 (0.96 to 2.49)	500 per 1000	275 more per 1000 (from 20 fewer to 745 more)
Pain On procedure	50 (1 study)	⊕⊕⊕⊖ MODERATE1	Peto OR 43.35	0 per 1000	960 more per 1000 (from 860 more to 1000 more)3

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4	Risk difference with PEI (95% CI)
	12 months	due to risk of bias	(14.45 to 130.02)		
Hyperthyroidism Presence of symptoms	50 (1 study) 12 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision	Peto OR 0.11 (0.2 to 0.71)	200 per 1000	173 fewer per 1000 (from 49 fewer to 152 fewer)

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
3 Zero events in control arm

Table 8: Clinical evidence summary: HIFU vs surgery, non-cystic nodules, NRS

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Surgery	Risk difference with HIFU (95% CI)
Nodule volume (% reduction)	154 (1 study) 6 months	⊕⊕⊖⊖ LOW2		The mean nodule volume (% reduction) in the control groups was 100 %	The mean nodule volume (% reduction) in the intervention groups was 36 less 1
Hypothyroidism	154 (1 study) 6 months	⊕⊖⊖⊖ VERY LOW2,3 due to imprecision	Peto OR 0.13 (0.02 to 0.94)	52 per 1000	45 fewer per 1000 (from 3 fewer to 51 fewer)
RLN palsy (temporary)	154 (1 study) 6 months	⊕⊖⊖⊖ VERY LOW2,3 due to	RR 1 (0.21 to 4.8)	39 per 1000	0 fewer per 1000 (from 31 fewer to 148 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Surgery	Risk difference with HIFU (95% CI)
		imprecision			
Bleeding (requiring re-exploration)	154 (1 study) 6 months	⨁⊖⊖⊖ VERY LOW2,3 due to imprecision	Peto OR 0.14 (0 to 6.82)	13 per 1000	11 fewer per 1000 (from 13 fewer to 69 more)
Infection	154 (1 study) 6 months	⨁⊖⊖⊖ VERY LOW4 due to imprecision	Not estimable		not estimable 4
Improvement in symptoms (any increase (slight/moderate/significant)	154 (1 study) 6 months	⨁⊖⊖⊖ VERY LOW3,5 due to risk of bias, imprecision	RR 1.09 (0.98 to 1.21)	870 per 1000	78 more per 1000 (from 17 fewer to 183 more)

1 95% confidence intervals cannot be calculated as no SD available for surgery (all participants with whole nodule removed), therefore there is no forest plot displayed in the appendix.
2 No additional risk of bias beyond default selection bias for non-randomised studies
3 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
4 Zero events in either arm
5 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Table 9: Clinical evidence summary: Radiofrequency ablation vs percutaneous ethanol injection, cystic nodules

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with PEI	Risk difference with RFA (95% CI)
Nodule volume % reduction	88 (2 studies)	⨁⨁⨁⊖ MODERATE1		The mean nodule reduction in the control groups was	The mean nodule reduction in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with PEI	Risk difference with RFA (95% CI)
	6 months	due to risk of bias		90 %	3.15 lower (5.97 to 0.33 lower)
Compressive symptoms VAS, 0-10, higher is worse	89 (2 studies) 6 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean compressive symptoms in the control groups was 0.6	The mean compressive symptoms in the intervention groups was 0.2 lower (0.55 lower to 0.15 higher)
Pain Mild or greater, during procedure	50 (1 study) 6 months	⊕⊕⊕⊖ MODERATE1 due to risk of bias	Peto OR 29.35 (9.65 to 89.23)	0 per 1000	840 more per 1000 (from 690 more to 990 more)3

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
3 Zero events in control arm

Table 10: Clinical evidence summary: Radiofrequency ablation vs surgery, nodule type unspecified, NRS

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Surgery	Risk difference with RFA (95% CI)
Quality of life SF-36, general health, 0-100, higher is better	216 (1 study) 6 months	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life in the control groups was 66.7	The mean quality of life in the intervention groups was 1.8 higher (0.2 to 3.4 higher)
Quality of life SF-36, vitality, 0-100, higher is better	216 (1 study) 6 months	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life in the control groups was 67.5	The mean quality of life in the intervention groups was 3.8 higher (1.56 to 6.04 higher)
Quality of life	216	⊕⊖⊖⊖		The mean quality of life in the	The mean quality of life in the

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Surgery	Risk difference with RFA (95% CI)
SF-36, mental health, 0-100, higher is better	(1 study) 6 months	VERY LOW1,2 due to risk of bias, imprecision		control groups was 79.3	intervention groups was 1.6 higher (0.11 to 3.09 higher)

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

Table 11: Clinical evidence summary: Microwave ablation vs surgery, nodule type unspecified, RCT

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Surgery	Risk difference with Microwave ablation (95% CI)
Nodule volume	52 (1 study) 12 months	⨁⊖⊖⊖ VERY LOW1 due to risk of bias, imprecision		See footnote 3	The mean final nodule volume in the intervention groups was 0.69ml (SD 0.89) compared with baseline 17.1ml (SD 14.4)3
Hoarseness (temporary)	52 (1 study) 12 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 0.43 (0.04 to 4.44)	83 per 1000	47 fewer per 1000 (from 80 fewer to 286 more)
Post-operative pain (temporary)	52 (1 study) 12 months	⨁⨁⊖⊖ LOW1 due to risk of bias	RR 0.08 (0.02 to 0.30)	917 per 1000	844 fewer per 1000 (from 642 fewer to 899 fewer)
Resolution of compressive symptoms	21 (1 study) 12 months	⨁⨁⊖⊖ LOW1 due to risk of bias	RR 1 (0.84 to 1.19)	1000 per 1000	0 fewer per 1000 (from 160 fewer to 190 more)

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Surgery	Risk difference with Microwave ablation (95% CI)
at very high risk of bias 2 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs 3 95% confidence intervals cannot be calculated as no SD available for surgery (all participants with whole nodule removed)					

See Appendix F: for full GRADE tables.

1.5 Economic evidence

1.5.1 Included studies

No relevant health economic studies were identified.

1.5.2 Excluded studies

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

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

1.5.3 Health economic model

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

1.5.4 Resource costs

Relevant unit costs are provided below to aid consideration of cost effectiveness.

Table 12: UK costs of thyroid enlargement drugs

Drug	Daily dose	Cost per Month	Cost per Year
Levothyroxine (T_4)	100µg once daily (a)	£1.03	£13.43

Source: BNF, August 2019.
(a) Maintenance dose 100-200mcg once daily

Table 13: UK costs of hospital procedures

Intervention	Unit costs
Radiofrequency Ablation (a)	£733.53
Radioactive iodine (b)	£286.32
Surgery (Thyroid Procedures with CC Score 0-4+)(c)	£3,689
Ethanol ablation (d)	£91.91 (e)

Source: NHS reference costs 2016-17, total HRG schedule [14].
(a) Cost code AB15Z, Radiofrequency Ablation or Cryoablation, for Pain Management
(b) Cost code RN51Z, Oral Delivery of Radiotherapy for Thyroid Ablation
(c) Weighted average of all 3 combined thyroid procedures with CC scores 0-1, 2-3, 4+(KA09C, KA09D, KA09E) including excess bed days and the average length of stay is 1.6 days
(d) Cost code RD42Z, Ultrasound Scan with duration of 20 minutes and over, without Contrast plus cost of ethanol BP, Martindale pharmaceuticals
*(e) Ultrasound £59.91 + ethanol ampule 20ml (5ml *4) £32*

1.6 Evidence statements

1.6.1 Clinical evidence statements

1.6.1.1.1 Levothyroxine vs placebo/follow-up

No clinically important difference was identified for nodule volume (ml, 11 studies, moderate quality), reduction in nodule volume (%, 1 study, high quality), hyperthyroidism (1 study, very low quality), arrhythmias (1 study, very low quality).

There was a clinically important benefit of levothyroxine for thyroid volume (1 study, moderate quality) and improvement in symptoms (1 study, very low quality).

No evidence identified for other outcomes.

1.6.1.1.2 Radiofrequency ablation vs follow-up only

No clinically important difference was identified for hypothyroidism or hyperthyroidism (1 study, very low quality).

There was a clinically important benefit of radiofrequency ablation for nodule volume (2 studies, high quality) and compressive symptoms (2 studies, moderate quality).

No evidence identified for other outcomes.

1.6.1.1.3 Laser ablation vs follow-up only

There was a clinically important benefit of laser ablation for nodule volume (1 study, moderate quality) and improvement in symptoms (1 study, very low quality).

No evidence identified for other outcomes.

1.6.1.1.4 Laser ablation vs levothyroxine

No clinically important difference was identified for nodule volume (1 study, moderate quality).

There was a clinically important benefit of laser ablation for improvement in symptoms (1 study, low quality).

No evidence identified for other outcomes.

1.6.1.1.5 Ethanol injection vs levothyroxine

No clinically important difference was identified for reduction in nodule volume (1 study, low quality).

There was a clinically important benefit of ethanol injection for compressive symptoms (1 study, low quality) and hyperthyroidism (1 study, low quality).

There was a clinically important harm of ethanol injection for pain on procedure (1 study, moderate quality).

No evidence identified for other outcomes.

1.6.1.1.6 HIFU vs surgery

No clinically important difference was identified for reduction in nodule volume (1 study, low quality), hypothyroidism, RLN palsy, bleeding, infection, improvement in symptoms (1 study, very low quality).

No evidence identified for other outcomes.

1.6.1.1.7 Radiofrequency ablation vs ethanol injection

No clinically important difference was identified for reduction in nodule volume (2 studies, moderate quality), compressive symptoms (2 studies, low quality).

There was a clinically important harm of radiofrequency ablation for pain (1 study, moderate quality).

No evidence identified for other outcomes.

1.6.1.1.8 Radiofrequency ablation vs surgery

No clinically important difference was identified for quality of life – general health, mental health (1 study, very low quality).

There was a clinically important benefit of radiofrequency ablation for quality of life – vitality (1 study, very low quality).

No evidence identified for other outcomes.

1.6.1.1.9 Microwave ablation vs surgery

No clinically important difference identified for reduction in nodule volume, hoarseness (1 study, very low quality), resolution of compressive symptoms (1 study, low quality).

There was a clinically important benefit of microwave ablation for post-operative pain (1 study, low quality).

No evidence identified for other outcomes.

1.6.2 Health economic evidence statements

- No relevant economic evaluations were identified.

1.7 The committee's discussion of the evidence

1.7.1 Interpreting the evidence

1.7.1.1 The outcomes that matter most

The committee agreed that the critical outcomes for this review were mortality and quality of life. Important outcomes included change in size of nodule/goitre, malignancy, hypothyroidism, hyperthyroidism, hypoparathyroidism, recurrent laryngeal nerve damage, bleeding, infection, arrhythmia, osteoporosis, pain, compressive symptoms and experience of care.

1.7.1.2 The quality of the evidence

The majority of studies included in the review only reported on size of nodule. A few studies reported compressive symptoms and pain but there was little evidence on any other outcome. Thyroid dysfunction was reported occasionally but it was not always clear how this was defined or assayed.

There was no evidence on management of goitre.

The quality of the evidence in the review varied from very low to high quality. The levothyroxine vs placebo/follow-up comparison generally had the largest number of

participants and the highest quality evidence, while the majority of other comparisons involved smaller numbers of participants.

The committee noted that the majority of studies did not specify the presence of symptoms as an inclusion criteria. However, because participants were typically recruited from specialty centres, referral to these centres was likely to be prompted by symptoms or cosmetic concerns.

1.7.1.3 Benefits and harms

Levothyroxine

The committee noted that there was no clinically important difference between levothyroxine and placebo/follow-up for the most commonly reported outcome of nodule volume. There was some evidence of benefit for compressive symptoms and thyroid volume as a whole, although these outcomes each came from single, smaller studies. Direct comparisons with other interventions (ethanol and laser ablation) showed a benefit for the other interventions on symptoms, although volume changes did not reach the level of clinical importance. The committee noted, based on their experience, that the long term use of levothyroxine is likely to be accompanied by adverse effects (for example TSH suppression). Overall the committee agreed that while there was insufficient evidence to strongly recommend against the use of levothyroxine, there was no evidence to support its use.

Ethanol ablation

There was a clinically important benefit of ethanol ablation compared to levothyroxine in terms of compressive symptoms but no difference in reduction in nodule volume. This comparison was from a small study and the committee noted that the ethanol ablation did result in a greater nodule volume reduction, even if it did not reach the minimal important difference. Ethanol ablation resulted in people experiencing mild pain during the procedure but had a benefit compared with thyroxine in terms of people experiencing hyperthyroidism symptoms during the follow-up.

Percutaneous thermal ablation

There was a clinically important benefit of radiofrequency ablation vs follow-up only in terms of nodule volume and compressive symptoms. There was no clinically important difference between RFA and ethanol ablation for volume or compressive symptoms and a clinically important harm of RFA vs ethanol ablation for mild pain during the procedure, although this was self-limiting. There was a clinically important benefit of radiofrequency ablation vs surgery in a non-randomised study for the quality of life vitality subscale, but no difference for any other quality of life subscale.

There was no clinically important difference between HIFU and surgery in terms of nodule volume, hypothyroidism, RLN palsy, bleeding, infection or improvement in symptoms.

There was no clinically important difference between microwave ablation and surgery in terms of nodule volume, hoarseness and resolution of compressive symptoms. Microwave ablation did have a clinically important benefit in terms of post-operative pain, however this was only temporary pain that resolved spontaneously with time.

There was a clinically important benefit of laser ablation for nodule volume and improvement in symptoms but only available from 1 study with very low quality evidence.

Overall the committee agreed that there may be a benefit from percutaneous thermal ablation techniques but that this is based on a small and generally low quality evidence base. Furthermore these techniques are not widely available in the UK. They agreed that they

could be considered in some people with thyroid enlargement but that there was insufficient evidence to distinguish between the various options and that more research was required before a strong recommendation could be made or specific subgroups who may benefit most could be firmly identified.

1.7.2 Cost effectiveness and resource use

There was no health economic evidence identified for either of the two-review question, therefore recommendations were based on consensus around the likely cost effectiveness of the interventions. Unit costs were presented for all the interventions considered to inform the qualitative considerations about cost effectiveness.

The recommendation made by the committee not to offer treatment to patients unless they have symptoms or there is clinical concern, is thought to be cost effective as only patients who need any type of treatment would then be considered. This can reduce the number of people receiving unnecessary treatment saving NHS money.

The committee made a consensus recommendation that treatment of children and young people should be discussed with an appropriate multidisciplinary team. The committee agreed this was important to ensure that appropriate management is agreed on as early as possible, preventing any deterioration in health or quality of life and, hence, ensuring the delivery of cost-effective care. However, the committee noted that, in general, thyroid enlargement is very rare in children and hence, small numbers will require such referral and this was seen to be current practice. Thus, the cost impact of this recommendation is likely to be minimal.

The committee recommended aspiration with possible ethanol ablation, as first line treatment for people with symptomatic cystic nodules, based on the clinical evidence and costs. The average cost of ethanol ablation was estimated to be £92, compared to radiofrequency ablation (RFA), which costs £734 (NHS reference cost code AB15Z).The committee did not recommend the use of RFA in this group of people as it was unlikely to be cost effective, same benefits as ethanol ablation and higher costs. However, the committee noted that percutaneous thermal ablation using either RFA or laser ablation should be considered in non-cystic nodule rather than ethanol ablation where there is a risk of ethanol leakage, which can cause complications. The committed considered the benefits from reducing the symptom burden and the likely complications and improving quality of life to justify the cost of providing thermal ablation.

The committee noted that for non-cystic nodule, multinodular goitre, or diffuse thyroid goitre causing compressive symptoms, radioactive iodine ablation, surgery or percutaneous thermal ablation (see above) should be considered. Radioactive iodine, costing £286 according to the NHS reference costs, should be considered where there is demonstrable radionuclide uptake, as first line. The use of radioactive iodine in this population reflects current practice; hence, unlikely to result in substantial cost impact. It was noted however, that there is uncertainty around the most appropriate dose of radioactive iodine to use in this population; hence, a research recommendation was also made.

The committee considered that for people with thyroid enlargement, who are contra-indicated or cannot tolerate the relevant interventions recommended above, referral for assessment for surgery should be considered. The committee discussed that despite the fact that surgery is more costly (approximately £4,265), it is likely to increase the patient's quality of life and avoid downstream costs due to complications. Furthermore, repeat procedures are unlikely to be required. In addition, not all patients referred for assessments for surgery will have surgery the number is excepted to be small and most patients with no symptoms will not require any interventions. Given the small number requiring surgery, no significant resource impact is expected.

1.7.3 Other factors the committee took into account

GPs may refer people to endocrinologists or directly to surgeons depending on local pathways available.

The committee noted that, based on their experience and the wider body of evidence not included in this review, larger nodules may require longer and more resource intensive interventions.

The committee noted that the cost of RFA includes the cost of the kits only as the machines are provided on a hiring agreement. There are capacity constraints on providing this intervention however, as there is only one centre in the UK that manages approximately 40 cases. Hence, the committee felt that including the option of laser ablation as another modality of thermal ablation can partly address the issue of access.

References

1. Baek JH, Ha EJ, Choi YJ, Sung JY, Kim JK, Shong YK. Radiofrequency versus Ethanol Ablation for Treating Predominantly Cystic Thyroid Nodules: A Randomized Clinical Trial. Korean Journal of Radiology. 2015; 16(6):1332-40

2. Baek JH, Kim YS, Lee D, Huh JY, Lee JH. Benign predominantly solid thyroid nodules: Prospective study of efficacy of sonographically guided radiofrequency ablation versus control condition. American Journal of Roentgenology. 2010; 194(4):1137-1142

3. Bandeira-Echtler E, Bergerhoff K, Richter B. Levothyroxine or minimally invasive therapies for benign thyroid nodules. Cochrane Database of Systematic Reviews 2014, Issue 6. Art. No.: CD004098. DOI: https://dx.doi.org/10.1002/14651858.CD004098.pub2.

4. Bayani M, Amani M, Moazezi Z. Efficacy of levothyroxine on benign thyroid nodule. Caspian Journal of Internal Medicine. 2012; 3(1):359-62

5. Bennedbaek FN, Nielsen LK, Hegedus L. Effect of percutaneous ethanol injection therapy versus suppressive doses of L-thyroxine on benign solitary solid cold thyroid nodules: a randomized trial. Journal of Clinical Endocrinology and Metabolism. 1998; 83(3):830-5

6. Cakir Ozkaya E, Aydin Y, Ozkan B, Karaahmetoglu Ozkan S, Eskioglu E, Guler S. The effect of thyroxine-suppressive therapy in patients with euthyroid nodular disease: A randomized controlled study. Endocrinologist. 2010; 20(4):182-184

7. Cesareo R, Iozzino M, Isgro MA, Annunziata F, Di Stasio E. Short term effects of levothyroxine treatment in thyroid multinodular disease. Endocrine Journal. 2010; 57(9):803-9

8. Cesareo R, Pasqualini V, Simeoni C, Sacchi M, Saralli E, Campagna G et al. Prospective study of effectiveness of ultrasound-guided radiofrequency ablation versus control group in patients affected by benign thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 2015; 100(2):460-6

9. Chen F, Tian G, Kong D, Zhong L, Jiang T. Radiofrequency ablation for treatment of benign thyroid nodules: A PRISMA-compliant systematic review and meta-analysis of outcomes. Medicine. 2016; 95(34):e4659

10. Cheng Z, Che Y, Yu S, Wang S, Teng D, Xu H et al. US-guided percutaneous radiofrequency versus microwave ablation for benign thyroid nodules: A prospective multicenter study. Scientific Reports. 2017; 7:9554

11. Cheung PS, Lee JM, Boey JH. Thyroxine suppressive therapy of benign solitary thyroid nodules: a prospective randomized study. World Journal of Surgery. 1989; 13(6):818-21; discussion 822

12. Chung SR, Suh CH, Baek JH, Park HS, Choi YJ, Lee JH. Safety of radiofrequency ablation of benign thyroid nodules and recurrent thyroid cancers: a systematic review and meta-analysis. International Journal of Hyperthermia. 2017; 33(8):920-930

13. Deandrea M, Sung JY, Limone P, Mormile A, Garino F, Ragazzoni F et al. Efficacy and safety of radiofrequency ablation versus observation for nonfunctioning benign

thyroid nodules: A randomized controlled international collaborative trial. Thyroid. 2015; 25(8):890-6

14. Department of Health. NHS reference costs 2016-17. 2017. Available from: https://www.gov.uk/government/collections/nhs-reference-costs Last accessed: 15/02/2019

15. Dossing H, Bennedbaek FN, Bonnema SJ, Grupe P, Hegedus L. Randomized prospective study comparing a single radioiodine dose and a single laser therapy session in autonomously functioning thyroid nodules. European Journal of Endocrinology. 2007; 157(1):95-100

16. Dossing H, Bennedbaek FN, Hegedus L. Interstitial laser photocoagulation (ILP) of benign cystic thyroid nodules--a prospective randomized trial. Journal of Clinical Endocrinology and Metabolism. 2013; 98(7):E1213-7

17. Faggiano A, Ramundo V, Assanti AP, Fonderico F, Macchia PE, Misso C et al. Thyroid nodules treated with percutaneous radiofrequency thermal ablation: a comparative study. Journal of Clinical Endocrinology and Metabolism. 2012; 97(12):4439-45

18. Ferreira MC, Piaia C, Cadore AC. Percutaneous ethanol injection versus conservative treatment for benign cystic and mixed thyroid nodules. Archives of Endocrinology & Metabolism. 2016; 60(3):211-6

19. Gambelunghe G, Fatone C, Ranchelli A, Fanelli C, Lucidi P, Cavaliere A et al. A randomized controlled trial to evaluate the efficacy of ultrasound-guided laser photocoagulation for treatment of benign thyroid nodules. Journal of Endocrinological Investigation. 2006; 29(9):RC23-6

20. Gangappa RB, Kenchannavar MB, Chowdary PB, Patanki AM, Ishwar M. Total Thyroidectomy for Benign Thyroid Diseases: What is the Price to be Paid? Journal of Clinical and Diagnostic Research JCDR. 2016; 10(6):PC04-7

21. Gharib H, James EM, Charboneau JW. Suppressive therapy with levothyroxine for solitary thyroid nodules. A double-blind controlled clinical study. New England Journal of Medicine. 1987; 317(2):70-75

22. Gharib H, Mazzaferri EL. Thyroxine suppressive therapy in patients with nodular thyroid disease. Annals of Internal Medicine. 1998; 128(5):386-94

23. Giovanella L, Piccardo A, Pezzoli C, Bini F, Ricci R, Ruberto T et al. Comparison of high intensity focused ultrasound and radioiodine for treating toxic thyroid nodules. Clinical Endocrinology. 2018; 89(2):219-225

24. Grussendorf M, Reiners C, Paschke R, Wegscheider K, Investigators L. Reduction of thyroid nodule volume by levothyroxine and iodine alone and in combination: a randomized, placebo-controlled trial. Journal of Clinical Endocrinology and Metabolism. 2011; 96(9):2786-95

25. Ha EJ, Baek JH, Kim KW, Pyo J, Lee JH, Baek SH et al. Comparative efficacy of radiofrequency and laser ablation for the treatment of benign thyroid nodules: systematic review including traditional pooling and bayesian network meta-analysis. Journal of Clinical Endocrinology and Metabolism. 2015; 100(5):1903-11

26. Jin H, Fan J, Liao K, He Z, Li W, Cui M. A propensity score matching study between ultrasound-guided percutaneous microwave ablation and conventional thyroidectomy for benign thyroid nodules treatment. International Journal of Hyperthermia. 2018:1-7

27. Joint Formulary Committee. British National Formulary (BNF) December 2017 update. 2017. Available from: http://www.bnf.org.uk Last accessed: 01/03/2018

28. Kim SH, Choo JH, Oh KC, Cho HC, Cho JH, Lim JK. Suppressive Therapy with Levothyroxine in Benignn Solitary Thyroid Nodule. Journal of korean society of endocrinology. 1997; 12(1):45-52

29. Korkusuz Y, Groner D, Raczynski N, Relin O, Kingeter Y, Grunwald F et al. Thermal ablation of thyroid nodules: are radiofrequency ablation, microwave ablation and high intensity focused ultrasound equally safe and effective methods? European Radiology. 2018; 28(3):929-935

30. La Rosa GL, Lupo L, Giuffrida D, Gullo D, Vigneri R, Belfiore A. Levothyroxine and potassium iodide are both effective in treating benign solitary solid cold nodules of the thyroid. Annals of Internal Medicine. 1995; 122(1):1-8

31. Lang BH, Woo YC, Wong CKH. High-intensity focused ultrasound for treatment of symptomatic benign thyroid nodules: A prospective study. Radiology. 2017; 284(3):897-906

32. Lang BH, Wu ALH. High intensity focused ultrasound (HIFU) ablation of benign thyroid nodules - a systematic review. Journal of Therapeutic Ultrasound. 2017; 5:11

33. Lang BHH, Wong CKH, Ma EPM. Single-session high intensity focussed ablation (HIFU) versus open cervical hemithyroidectomy for benign thyroid nodule: analysis on early efficacy, safety and voice quality. International Journal of Hyperthermia. 2017; 33(8):868-874

34. Lang BHH, Wong CKH, Ma EPM, Woo YC, Chiu KWH. A propensity-matched analysis of clinical outcomes between open thyroid lobectomy and high-intensity focused ultrasound (HIFU) ablation of benign thyroid nodules. Surgery. 2019; 165(1):85-91

35. Larijani B, Pajouhi M, Bastanhagh MH, Sadjadi A, Aghakhani S, Zare F et al. Role of levothyroxine suppressive therapy for benign cold nodules of thyroid: A randomized, double-blind, placebo-controlled clinical trial. Therapy. 2005; 2(6):883-888

36. Larijani B, Pajouhi M, Bastanhagh MH, Sadjadi A, Sedighi N, Eshraghian MR. Evaluation of suppressive therapy for cold thyroid nodules with levothyroxine: double-blind placebo-controlled clinical trial. Endocrine Practice. 1999; 5(5):251-6

37. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2018]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

38. Pacella CM, Mauri G, Cesareo R, Paqualini V, Cianni R, De Feo P et al. A comparison of laser with radiofrequency ablation for the treatment of benign thyroid nodules: a propensity score matching analysis. International Journal of Hyperthermia. 2017; 33(8):911-919

39. Papini E, Bacci V, Panunzi C, Pacella CM, Fabbrini R, Bizzarri G et al. A prospective randomized trial of levothyroxine suppressive therapy for solitary thyroid nodules. Clinical Endocrinology. 1993; 38(5):507-13

40. Papini E, Guglielmi R, Bizzarri G, Graziano F, Bianchini A, Brufani C et al. Treatment of benign cold thyroid nodules: a randomized clinical trial of percutaneous laser ablation versus levothyroxine therapy or follow-up. Thyroid. 2007; 17(3):229-35

41. Papini E, Petrucci L, Guglielmi R, Panunzi C, Rinaldi R, Bacci V et al. Long-term changes in nodular goiter: a 5-year prospective randomized trial of levothyroxine suppressive therapy for benign cold thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 1998; 83(3):780-3

42. Reverter JL, Lucas A, Salinas I, Audi L, Foz M, Sanmarti A. Suppressive therapy with levothyroxine for solitary thyroid nodules. Clinical Endocrinology. 1992; 36(1):25-8

43. Sdano MT, Falciglia M, Welge JA, Steward DL. Efficacy of thyroid hormone suppression for benign thyroid nodules: Meta-analysis of randomized trials. Otolaryngology - Head and Neck Surgery. 2005; 133(3):391-396

44. Sewefy AM, Tohamy TA, Esmael TM, Atyia AM. Intra-capsular total thyroid enucleation versus total thyroidectomy in treatment of benign multinodular goiter. A prospective randomized controlled clinical trial. International Journal of Surgery. 2017; 45:29-34

45. Sung JY, Baek JH, Kim KS, Lee D, Yoo H, Kim JK et al. Single-session treatment of benign cystic thyroid nodules with ethanol versus radiofrequency ablation: a prospective randomized study. Radiology. 2013; 269(1):293-300

46. Verde G, Papini E, Pacella CM, Gallotti C, Delpiano S, Strada S et al. Ultrasound guided percutaneous ethanol injection in the treatment of cystic thyroid nodules. Clinical Endocrinology. 1994; 41(6):719-24

47. Wang JF, Wu T, Hu KP, Xu W, Zheng BW, Tong G et al. Complications following radiofrequency ablation of benign thyroid nodules: A systematic review. Chinese Medical Journal. 2017; 130(11):1361-1370

48. Wemeau JL, Caron P, Schvartz C, Schlienger JL, Orgiazzi J, Cousty C et al. Effects of thyroid-stimulating hormone suppression with levothyroxine in reducing the volume of solitary thyroid nodules and improving extranodular nonpalpable changes: a randomized, double-blind, placebo-controlled trial by the French Thyroid Research Group. Journal of Clinical Endocrinology and Metabolism. 2002; 87(11):4928-34

49. Yan J, Qiu T, Lu J, Wu Y, Yang Y. Microwave ablation induces a lower systemic stress response in patients than open surgery for treatment of benign thyroid nodules. International Journal of Hyperthermia. 2018; 34(5):606-610

50. Yue WW, Li XL, Xu HX, Lu F, Sun LP, Guo LH et al. Quality of life and cost-effectiveness of radiofrequency ablation versus open surgery for benign thyroid nodules: A retrospective cohort study. Scientific Reports. 2016; 6:37838

51. Yue WW, Wang SR, Lu F, Li XL, Xu HX, Sun LP et al. Corrigendum: Quality of life and cost-effectiveness of radiofrequency ablation versus open surgery for benign thyroid nodules: A retrospective cohort study. Scientific Reports. 2017; 7:41342

52. Yue WW, Wang SR, Lu F, Sun LP, Guo LH, Zhang YL et al. Radiofrequency ablation vs. microwave ablation for patients with benign thyroid nodules: a propensity score matching study. Endocrine. 2017; 55(2):485-495

53. Zelmanovitz F, Genro S, Gross JL. Suppressive therapy with levothyroxine for solitary thyroid nodules: a double-blind controlled clinical study and cumulative meta-analyses. Journal of Clinical Endocrinology and Metabolism. 1998; 83(11):3881-5

54. Zhi X, Zhao N, Liu Y, Liu JB, Teng C, Qian L. Microwave ablation compared to thyroidectomy to treat benign thyroid nodules. International Journal of Hyperthermia. 2018; 34(5):644-652

55. Zingrillo M, Modoni S, Conte M, Frusciante V, Trischitta V. Percutaneous ethanol injection plus radioiodine versus radioiodine alone in the treatment of large toxic thyroid nodules. Journal of Nuclear Medicine. 2003; 44(2):207-10

56. Zingrillo M, Torlontano M, Ghiggi MR, Frusciante V, Varraso A, Liuzzi A et al. Radioiodine and percutaneous ethanol injection in the treatment of large toxic thyroid nodule: a long-term study. Thyroid. 2000; 10(11):985-9

Appendices

Appendix A: Review protocols

Table 14:

ID	Field	Content
I	Review questions	Which people with non-malignant thyroid enlargement should be referred for surgery? What is the clinical and cost effectiveness of non-surgical treatments (e.g. radiofrequency ablation, high intensity focused ultrasound) for non-malignant thyroid enlargement?
II	Type of review question	Intervention A review of health economic evidence related to the same review question was conducted in parallel with this review. For details see the health economic review protocol for this NICE guideline.
III	Objective of the review	Determine the most clinically and cost effective way to manage non-malignant thyroid enlargement
IV	Eligibility criteria – population / disease / condition / issue / domain	People with non-malignant thyroid enlargement
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	Surgery Radiofrequency ablation High intensity focused ultrasound Ethanol ablation Radioiodine ablation Levothyroxine therapy Microwave ablation Laser ablation Monitoring only Placebo/sham
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	Any of above vs any other
VII	Outcomes and prioritisation	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Percent change in volume of nodule/goitre (continuous) • Malignancy (dichotomous) • Hypothyroidism (dichotomous) • Hyperthyroidism (dichotomous) • Hypoparathyroidism (dichotomous)

		• Recurrent laryngeal nerve damage (dichotomous) • Bleeding (dichotomous) • Infection (dichotomous) • Arrhythmia (dichotomous) • Osteoporosis (dichotomous) • Pain (continuous) • Compressive symptoms (continuous) • Patient/family/carer experience (continuous) Minimum duration as for the minimum duration for inclusion of studies unless specified.
VIII	Eligibility criteria – study design	• RCTs preferred, if no RCTs available (on an intervention by intervention basis) to consider non-randomised cohort studies in which key confounders (age, sex, co-existing conditions, size of nodule, type of nodule) are addressed, either through restriction (see stratifications below) or appropriate matching/statistical adjustment • Minimum duration of 3 months • Crossover studies excluded
IX	Other inclusion exclusion criteria	-
X	Proposed sensitivity / subgroup analysis, or meta-regression	**Stratifications** • Nodule vs goitre • Age - children (4-18), adults (>18-65), older adults (>65) • Type of nodule (cystic vs non-cystic) **Subgroup analyses** • Age subdivisions (18-50, 50-65, 65-85, >85) • Surgical indications (compressive symptoms, cosmesis, uncertainty in diagnosis)
XI	Selection process – duplicate screening / selection / analysis	• A sample of at least 10% of the abstract lists were double-sifted by a senior research fellow and discrepancies rectified, with committee input where consensus could not be reached, for more information please see the separate Methods report for this guideline.
XII	Data management (software)	Pairwise meta-analyses were performed using Cochrane Review Manager (RevMan5). GRADEpro was used to assess the quality of evidence for each outcome. Endnote was used for bibliography, citations, sifting and reference management
XIII	Information sources – databases and dates	• Medline, Embase and The Cochrane Library
XIV	Identify if an update	Not an update
XV	Author contacts	https://www.nice.org.uk/guidance/indevelopment/gid-ng10074
XVI	Highlight if amendment to previous protocol	Not an amendment
XVI	Search	For details please see Appendix B:

I	strategy – for one database	
XVII	Data collection process – forms / duplicate	A standardised evidence table format will be used and published as Appendix D: of the evidence report.
XIX	Data items – define all variables to be collected	For details please see evidence tables in Appendix D: (clinical evidence tables) or Appendix H: (health economic evidence tables).
XX	Methods for assessing bias at outcome / study level	Standard study checklists were used to critically appraise individual studies. For details please see section 6.2 of Developing NICE guidelines: the manual The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/
XXI	Criteria for quantitative synthesis	For details please see section 6.4 of Developing NICE guidelines: the manual.
XXII	Methods for quantitative analysis – combining studies and exploring (in)consistency	For details please see the separate Methods report for this guideline.
XXIII	Meta-bias assessment – publication bias, selective reporting bias	For details please see section 6.2 of Developing NICE guidelines: the manual.
XXIV	Confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual.
XXV	Rationale / context – what is known	For details please see the introduction to the evidence review.
XXVI	Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the National Guideline Centre (NGC) and chaired by Sarah Fishburn in line with section 3 of Developing NICE guidelines: the manual. Staff from NGC undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
XXVII	Sources of funding / support	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXVIII	Name of sponsor	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXIX	Roles of sponsor	NICE funds NGC to develop guidelines for those working in the NHS, public health and social care in England.
XXX	PROSPERO registration	Not registered

	number	

Table 15: Health economic review protocol

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 2003, 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).[37] **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 2003 or later but that depend on unit costs and resource data entirely or predominantly from before 2003 will be rated as 'Not applicable'. • Studies published before 2003 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 this review are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual 2014, updated 2018 https://www.nice.org.uk/guidance/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869

For more detailed information, please see the Methodology Review.

B.1 Clinical search literature search strategy

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.

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 07 January 2019	Exclusions Randomised controlled trials Systematic review studies Observational studies
Embase (OVID)	1974 – 07 January 2019	Exclusions Randomised controlled trials Systematic review studies Observational studies
The Cochrane Library (Wiley)	Cochrane Reviews to 2019 Issue 1 or 12 CENTRAL to 2019 Issue 1 or 12 DARE, and NHSEED to 2015 Issue 2 of 4 HTA to 2016 Issue 2 of 4	None

Medline (Ovid) search terms

1.	(((thyroid adj4 (swell* or enlarg* or nodule* or node*)) or (goitre* or goiter*)) adj5 (non-malignan* or nonmalignan* or benign)).ti,ab.
2.	Thyroid Nodule/
3.	1 or 2
4.	letter/
5.	editorial/
6.	news/
7.	exp historical article/
8.	Anecdotes as Topic/
9.	comment/
10.	case report/
11.	(letter or comment*).ti.
12.	or/4-11
13.	randomized controlled trial/ or random*.ti,ab.

14.	12 not 13
15.	animals/ not humans/
16.	exp Animals, Laboratory/
17.	exp Animal Experimentation/
18.	exp Models, Animal/
19.	exp Rodentia/
20.	(rat or rats or mouse or mice).ti.
21.	or/14-20
22.	3 not 21
23.	Thyroidectomy/
24.	(thyroidectomy or total thryoidectomy or total-thyroidectomy or lobectomy or hemithyroidectomy or hemi-thyroidectomy or near total thryoidectomy or near-total thyroidectomy or isthmusectomy or surg* or interstitial laser photocoagulation or ILP).ti,ab.
25.	((radioiodine or microwave or radiofrequency or ethanol) adj3 ablation).ti,ab.
26.	(radioactive iodine or RAI).ti,ab.
27.	High-Intensity Focused Ultrasound Ablation/
28.	(high intensity focused ultrasound or HIFU).ti,ab.
29.	(Magnetic resonance guided focused ultrasound or MRgFUS).ti,ab.
30.	(percutaneous ethanol injection or PEI).ti,ab.
31.	levothyroxine.ti,ab.
32.	or/23-31
33.	22 and 32
34.	randomized controlled trial.pt.
35.	controlled clinical trial.pt.
36.	randomi#ed.ti,ab.
37.	placebo.ab.
38.	randomly.ti,ab.
39.	Clinical Trials as topic.sh.
40.	trial.ti.
41.	or/34-40
42.	Meta-Analysis/
43.	exp 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 or cross sectional) 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.	or/53-62
64.	exp case control study/
65.	case control*.ti,ab.
66.	or/64-65
67.	63 or 66
68.	Cross-sectional studies/
69.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
70.	or/68-69
71.	63 or 70
72.	63 or 66 or 70
73.	33 and (41 or 52 or 72)
74.	limit 73 to English language

Embase (Ovid) search terms

1.	(((thyroid adj4 (swell* or enlarg* or nodule* or node*)) or (goitre* or goiter*)) adj5 (non-malignan* or nonmalignan* or benign)).ti,ab.
2.	Thyroid Nodule/
3.	1 or 2
4.	letter.pt. or letter/
5.	note.pt.
6.	editorial.pt.
7.	case report/ or case study/
8.	(letter or comment*).ti.
9.	or/4-8
10.	randomized controlled trial/ or random*.ti,ab.
11.	9 not 10
12.	animal/ not human/
13.	nonhuman/
14.	exp Animal Experiment/
15.	exp Experimental Animal/
16.	animal model/
17.	exp Rodent/
18.	(rat or rats or mouse or mice).ti.
19.	or/11-18
20.	3 not 19
21.	Thyroidectomy/

22.	(thyroidectomy or total thryoidectomy or total-thyroidectomy or lobectomy or hemithyroidectomy or hemi-thyroidectomy or near total thryoidectomy or near-total thyroidectomy or isthmusectomy or surg* or interstitial laser photocoagulation or ILP).ti,ab.
23.	((radioiodine or microwave or radiofrequency or ethanol) adj3 ablation).ti,ab.
24.	radioactive iodine/
25.	(radioactive iodine or RAI).ti,ab.
26.	High-Intensity Focused Ultrasound Ablation/
27.	(high intensity focused ultrasound or HIFU).ti,ab.
28.	(Magnetic resonance guided focused ultrasound or MRgFUS).ti,ab.
29.	(percutaneous ethanol injection or PEI).ti,ab.
30.	levothyroxine/
31.	or/21-30
32.	20 and 31
33.	random*.ti,ab.
34.	factorial*.ti,ab.
35.	(crossover* or cross over*).ti,ab.
36.	((doubl* or singl*) adj blind*).ti,ab.
37.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
38.	crossover procedure/
39.	single blind procedure/
40.	randomized controlled trial/
41.	double blind procedure/
42.	or/33-41
43.	systematic review/
44.	meta-analysis/
45.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
46.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
47.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
48.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
49.	(search* adj4 literature).ab.
50.	(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.
51.	cochrane.jw.
52.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
53.	or/43-52
54.	Clinical study/
55.	Observational study/
56.	family study/
57.	longitudinal study/
58.	retrospective study/
59.	prospective study/
60.	cohort analysis/
61.	follow-up/
62.	cohort*.ti,ab.

63.	61 and 62
64.	(cohort adj (study or studies or analys* or data)).ti,ab.
65.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
66.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
67.	(before adj2 after adj2 (study or studies or data)).ti,ab.
68.	or/54-60,63-67
69.	exp case control study/
70.	case control*.ti,ab.
71.	or/69-70
72.	68 or 71
73.	cross-sectional study/
74.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
75.	or/73-74
76.	68 or 75
77.	68 or 71 or 75
78.	32 and (42 or 53 or 77)
79.	limit 78 to English language

Cochrane Library (Wiley) search terms

#1.	(((thyroid near/4 (swell* or enlarg* or nodule* or node*)) or (goitre* or goiter*)) near/5 (non-malignan* or nonmalignan* or benign)):ti,ab
#2.	MeSH descriptor: [Thyroid Nodule] explode all trees
#3.	#1 or #2
#4.	MeSH descriptor: [Thyroidectomy] explode all trees
#5.	("total thryoidectomy" or "total-thyroidectomy" or "lobectomy" or "hemithyroidectomy" or "hemi-thyroidectomy" or "near total thryoidectomy" or "near-total thyroidectomy" or "isthmusectomy" or "surg*" or "interstitial laser photocoagulation" or "ILP"):ti,ab
#6.	(radioactive iodine or RAI):ti,ab
#7.	(high intensity focused ultrasound or HIFU):ti,ab
#8.	(Magnetic resonance guided focused ultrasound or MRgFUS):ti,ab
#9.	(percutaneous ethanol injection or PEI):ti,ab
#10.	levothyroxine:ti,ab
#11.	(or #4-#10)
#12.	#3 and #11

B.2 Health Economics literature search strategy

Health economic evidence was identified by conducting a broad search relating to a thyroid disease population in NHS Economic Evaluation Database (NHS EED – this ceased to be updated after March 2015) and the Health Technology Assessment database (HTA) with no date restrictions. NHS EED and HTA databases are hosted by the Centre for Research and Dissemination (CRD). Additional searches were run on Medline and Embase for health economics, economic modelling and quality of life studies.

Table 16: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline	2014 – 07 January 2019	Exclusions

Database	Dates searched	Search filter used
		Health economics studies Health economics modelling studies Quality of life studies
Embase	2014 – 07 January 2019	Exclusions Health economics studies Health economics modelling studies Quality of life studies
Centre for Research and Dissemination (CRD)	HTA - Inception – 07 January 2019 NHSEED - Inception to March 2015	None

Medline (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice).ti.
24.	or/17-23
25.	6 not 24
26.	limit 25 to English language
27.	Economics/
28.	Value of life/
29.	exp "Costs and Cost Analysis"/
30.	exp Economics, Hospital/

31.	exp Economics, Medical/
32.	Economics, Nursing/
33.	Economics, Pharmaceutical/
34.	exp "Fees and Charges"/
35.	exp Budgets/
36.	budget*.ti,ab.
37.	cost*.ti.
38.	(economic* or pharmaco?economic*).ti.
39.	(price* or pricing*).ti,ab.
40.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
41.	(financ* or fee or fees).ti,ab.
42.	(value adj2 (money or monetary)).ti,ab.
43.	or/27-42
44.	exp models, economic/
45.	*Models, Theoretical/
46.	*Models, Organizational/
47.	markov chains/
48.	monte carlo method/
49.	exp Decision Theory/
50.	(markov* or monte carlo).ti,ab.
51.	econom* model*.ti,ab.
52.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
53.	or/44-52
54.	quality-adjusted life years/
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
62.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
63.	(hui or hui1 or hui2 or hui3).ti,ab.
64.	(health* year* equivalent* or hye or hyes).ti,ab.
65.	discrete choice*.ti,ab.
66.	rosser.ti,ab.
67.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
68.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
69.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
70.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
71.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
72.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
73.	or/54-72
74.	26 and (43 or 53 or 73)

Embase (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis*.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11
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).ti.
22.	or/14-21
23.	6 not 22
24.	limit 23 to English language
25.	health economics/
26.	exp economic evaluation/
27.	exp health care cost/
28.	exp fee/
29.	budget/
30.	funding/
31.	budget*.ti,ab.
32.	cost*.ti.
33.	(economic* or pharmaco?economic*).ti.
34.	(price* or pricing*).ti,ab.
35.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
36.	(financ* or fee or fees).ti,ab.
37.	(value adj2 (money or monetary)).ti,ab.
38.	or/25-37

39.	statistical model/
40.	exp economic aspect/
41.	39 and 40
42.	*theoretical model/
43.	*nonbiological model/
44.	stochastic model/
45.	decision theory/
46.	decision tree/
47.	monte carlo method/
48.	(markov* or monte carlo).ti,ab.
49.	econom* model*.ti,ab.
50.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
51.	or/41-50
52.	quality adjusted life year/
53.	"quality of life index"/
54.	short form 12/ or short form 20/ or short form 36/ or short form 8/
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
62.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
63.	(hui or hui1 or hui2 or hui3).ti,ab.
64.	(health* year* equivalent* or hye or hyes).ti,ab.
65.	discrete choice*.ti,ab.
66.	rosser.ti,ab.
67.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
68.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
69.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
70.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
71.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
72.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
73.	or/52-72
74.	24 and (38 or 51 or 73)

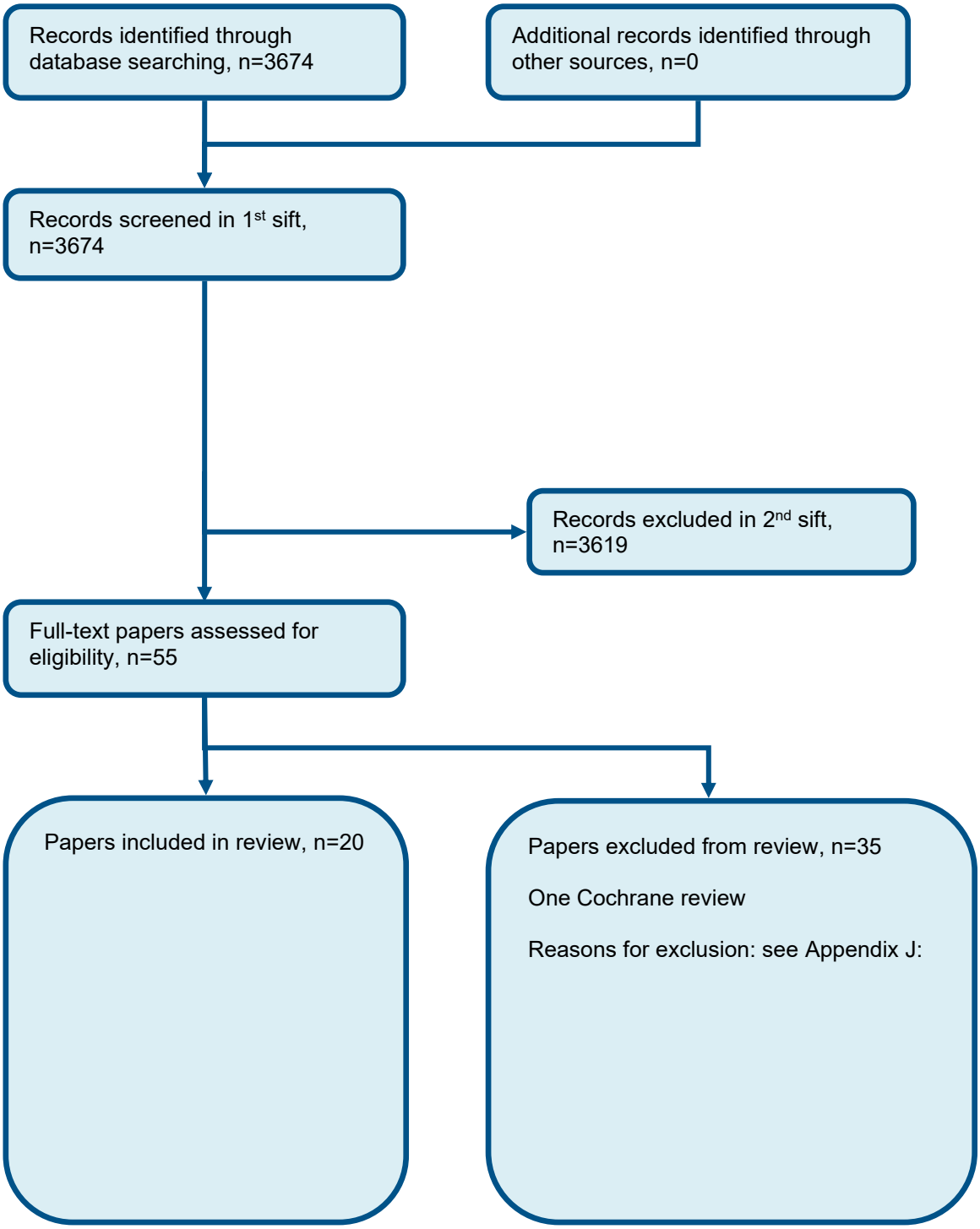
NHS EED and HTA (CRD) search terms

#1.	MeSH DESCRIPTOR Thyroid Diseases EXPLODE ALL TREES
#2.	hyperthyroid*

#3.	hypothyroid*
#4.	thyrotoxicosis*
#5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*))
#6.	#1 OR #2 OR #3 OR #4 or #5

Appendix C: Clinical evidence selection

Figure 1: Flow chart of clinical study selection for the review of benign thyroid enlargement



Appendix D: Clinical evidence tables

Study	Baek 2015[1]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=46)
Countries and setting	Conducted in South Korea; Setting: Not stated
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	Define
Exclusion criteria	Define
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (SD): 51 (15). Gender (M:F): 20:80. Ethnicity: Not stated
Further population details	1. Age groups: 50-65 2. Surgical indication: Compressive symptoms
Extra comments	Inclusion - predominantly cystic nodules (nodules 50-90% cystic), pressure or cosmetic symptoms, benign by 2 FNABs, normal thyroid hormones, older than 20, not pregnant
Indirectness of population	No indirectness
Interventions	(n=22) Intervention 1: Radiofrequency ablation. Moving shot technique, began with 40W, increased up to 80W to form hyperechoic zone. Duration 6 months. Concurrent medication/care: Usual care (n=24) Intervention 2: Ethanol ablation. 99% ethanol injection, 50% volume aspirated and replaced. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: RADIOFREQUENCY ABLATION versus ETHANOL ABLATION Protocol outcome 1: Percent change in size of nodule/goitre at 3 months - Actual outcome for Cystic nodule : % reduction in volume of nodule at end of study at 6 months; Group 1: mean 87.5 (SD 11.5); n=22, Group 2: mean 82.4 (SD 28.6); n=24 Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: RFA group mean volume at baseline 8.6, PEI group 14.7; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 2: Compressive symptoms at 3 months - Actual outcome for Cystic nodule : Symptom score at end of follow-up (VAS, 0-10, higher is worse) at 6 months; Group 1: mean 0.2 (SD 0.4); n=22, Group 2: mean 0.7 (SD 1.3); n=24 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 ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Patient/family/carer experience of care at 3 months

Study	Bennedbaek 1998[5]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=50)
Countries and setting	Conducted in Denmark; Setting: Denmark, area with borderline deficient iodine supply
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	20-70, benign solitary cold nodule causing discomfort, solid, FNAB colloid, euthyroid, normal calcium + calcitonin, no major comorbidity, no medication affecting thyroid function, no history of head or neck irradiation, normal laryngoscopy
Exclusion criteria	Nil else
Recruitment/selection of patients	Nil else
Age, gender and ethnicity	Age - Mean (range): 44 (37-52). Gender (M:F): Define. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=25) Intervention 1: Ethanol ablation. 1 dose, 98% ethanol, volume 20-50% of pre-treatment nodule volume. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=25) Intervention 2: Levothyroxine therapy. 1.5 µg/kg, titrated to TSH 0.1-0.4 (mU/L) after 1 month. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Academic or government funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: ETHANOL ABLATION versus LEVOTHYROXINE THERAPY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Median reduction in nodule volume at end of follow-up at 12 months; Group 1: mean 47 (SD 31); n=25, Group 2: mean 9 (SD 37); n=25: Comments: SDs calculated from confidence intervals

Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Median volume in T4 group 7.1ml, median volume in PEI group 9.2ml; Group 1 Number missing: 2, Reason: Had surgery as no improvement in symptoms, values LOCF from 6 months; Group 2 Number missing: 0 Protocol outcome 2: Hyperthyroidism at 3 months - Actual outcome for Non-cystic nodule : Number of people with symptoms of hyperthyroidism (resolved with dose titration) at 12 months; Group 1: 0/25, Group 2: 5/25 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Median volume in T4 group 7.1ml, median volume in PEI group 9.2ml; Group 1 Number missing: 2, Reason: Had surgery as no improvement in symptoms, values LOCF from 6 months; Group 2 Number missing: 0 Protocol outcome 3: Pain at 3 months - Actual outcome for Non-cystic nodule : Number of people reporting pain during procedure (resolved within days) at 12 months; Group 1: 24/25, Group 2: 0/25 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Median volume in T4 group 7.1ml, median volume in PEI group 9.2ml; Group 1 Number missing: 2, Reason: Had surgery as no improvement in symptoms, values LOCF from 6 months; Group 2 Number missing: 0 Protocol outcome 4: Compressive symptoms at 3 months - Actual outcome for Non-cystic nodule : Number of people with symptoms (mild/mod/sev) reporting an improvement by at least one category by end of follow-up at 12 months; Group 1: 17/22, Group 2: 11/22 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Median volume in T4 group 7.1ml, median volume in PEI group 9.2ml; Group 1 Number missing: 2, Reason: Had surgery as no improvement in symptoms, values LOCF from 6 months; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Patient/family/carer experience of care at 3 months

Study	**Cakir ozkaya 2010[6]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=62)
Countries and setting	Conducted in Turkey; Setting: Outpatient clinic
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Benign by FNAB, <2cm, non-cystic, euthyroid, no CV/liver/renal disease, no pregnancy, no previous thyroid treatment
Exclusion criteria	Nil else
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (SD): Not stated. Gender (M:F): Not stated. Ethnicity: Not stated
Further population details	1. Age groups: Not stated / Unclear 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=35) Intervention 1: Levothyroxine therapy. 50 to 100mg/day. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=27) Intervention 2: Monitoring only. Nil else stated. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus MONITORING ONLY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome: Mean volume of nodule (cm3, used as approximation of ml) at end of follow-up at 12 months; Group 1: mean 0.748 (SD 8); n=33, Group 2: mean 0.346 (SD 4); n=25
Risk of bias: All domain - High. Selection - High. Blinding - Low. Incomplete outcome data - Low. Outcome reporting - Low. Measurement - Low. Crossover - Low:

Indirectness of outcome: No indirectness ; Baseline details: LT4 0.76, FU 0.35; Group 1 Number missing: 2, Reason: Not stated; Group 2 Number missing: 2, Reason: Not stated	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Cesareo 2010[7]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=71)
Countries and setting	Conducted in Italy; Setting: Italy
Line of therapy	1st line
Duration of study	Intervention + follow up: 24 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Non-cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	Pre-menopausal women, 2-5 nodules, colloid, hypofunctioning, normal T3/T4/TSH, no thyroid autoantibodies, no previous treatment for thyroid disease, no smoking, no pregnancy in last 12 months, BMI 18.5 to 30, no history of neck irradiation or surgery
Exclusion criteria	Nil else
Recruitment/selection of patients	Nil specified
Age, gender and ethnicity	Age - Mean (SD): 36 (10). Gender (M:F): 0:100. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Extra comments	Area of iodine deficiency
Indirectness of population	No indirectness
Interventions	(n=36) Intervention 1: Levothyroxine therapy. 2 micrograms/kg body weight. Duration 24 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=35) Intervention 2: Monitoring only. Nil else. Duration 24 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Academic or government funding
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus FOLLOW-UP ONLY Protocol outcome 1: Percent change in size of nodule/goitre at 3 months	

- Actual outcome for Non-cystic nodule : Dominant nodule final volume (ml) at 12 months; Group 1: mean 0.8 (SD 0.8); n=36, Group 2: mean 2.5 (SD 2.4); n=35 Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - High; Indirectness of outcome: No indirectness ; Baseline details: T4 group mean baseline dominant nodule 1.8ml, control 1.2ml; Group 1 Number missing: ; Group 2 Number missing: - Actual outcome for Non-cystic nodule : Total thyroid final volume (ml) at 12 months; Group 1: mean 10.6 (SD 4.9); n=36, Group 2: mean 20 (SD 14.7); n=35 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - High; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Cesareo 2015[8]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=84)
Countries and setting	Conducted in Italy; Setting: Italy, hospital
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Non-cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	-
Exclusion criteria	-
Recruitment/selection of patients	Consecutive patients
Age, gender and ethnicity	Age - Mean (SD): 54 (13). Gender (M:F): 39:61. Ethnicity: Not stated
Further population details	1. Age groups: 50-65 2. Surgical indication: Compressive symptoms
Extra comments	Included - >18, compressive symptoms or cosmetic concerns + >5ml or max diameter >2cm and steadily growing, benign by 2 FNA, solid >70%, cold nodule, normal thyroid hormones, no previous thyroid treatment. Excluded - pregnancy, confluent nodules in compressive mass.
Indirectness of population	No indirectness
Interventions	(n=42) Intervention 1: Radiofrequency ablation. 1 dose, 60W, moving shot technique. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=42) Intervention 2: Monitoring only. Nil else specified. Duration 6 months. Concurrent medication/care: Nil else specified. Indirectness: No indirectness
Funding	No funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: RADIOFREQUENCY ABLATION versus MONITORING ONLY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months

- Actual outcome for Non-cystic nodule : Size of nodule at end of follow-up (ml) at 6 months; Group 1: mean 8.6 (SD 9.5); n=42, Group 2: mean 27.8 (SD 22.1); n=42 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 2: Hypothyroidism at 3 months - Actual outcome for Non-cystic nodule : Hypothyroidism at 6 months; Group 1: 0/42, Group 2: 0/42 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 ; Group 1 Number missing: ; Group 2 Number missing: - Actual outcome for Non-cystic nodule : Hyperthyroidism at 6 months; Group 1: 1/42, Group 2: 0/42 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 ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 3: Compressive symptoms at 3 months - Actual outcome for Non-cystic nodule : Compressive symptoms at end of follow-up (0-10 VAS, higher worse) at 6 months; Group 1: mean 0.4 (SD 0.8); n=42, Group 2: mean 2.9 (SD 3.2); n=42 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 ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Patient/family/carer experience of care at 3 months

Study	Deandrea 2015[13]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=80)
Countries and setting	Conducted in Multiple countries; Setting: Italian and Korean specialty centres
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable:
Inclusion criteria	Older than 18, solid nodule (>70%), volume 10-20ml, pressure symptoms/cosmetic problems, benign by 2 biopsies, normal thyroid hormones and calcitonin
Exclusion criteria	Nodules with US features of malignancy, treatment for nodule in 6 months prior to study
Recruitment/selection of patients	Not specified
Age, gender and ethnicity	Age - Mean (SD): 52 (12). Gender (M:F): 10:90. Ethnicity: Not stated
Further population details	1. Age groups: 50-65 2. Surgical indication: Not applicable
Indirectness of population	No indirectness
Interventions	(n=40) Intervention 1: Radiofrequency ablation. RFA, skin punctured using transisthmic approach method, 'moving-shot technique'. Duration 6-month follow-up . Concurrent medication/care: Not specified. Indirectness: No indirectness (n=40) Intervention 2: Monitoring only. Nil specified. Duration 6 month follow-up. Concurrent medication/care: Nil specified. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: RADIOFREQUENCY ABLATION versus MONITORING ONLY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome: Mean volume of nodule at end of follow-up at 6 months; Group 1: mean 4.7 (SD 2.7); n=40, Group 2: mean 15.2 (SD 3.5); n=40
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Compressive symptoms at 3 months - Actual outcome: Compressive symptom score (0-10, VAS, higher is worse) at 6 months; Group 1: mean 0.4 (SD 0.7); n=40, Group 2: mean 3.3 (SD 1.7); n=40 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 ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Patient/family/carer experience of care at 3 months

Study	Gharib 1987[21]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=53)
Countries and setting	Conducted in USA; Setting: Patients seen in thyroid outpatients clinic
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Single palpable nodule, benign on FNAB
Exclusion criteria	Multiple nodules, uncertain biopsy findings, pregnancy, cardiovascular disease
Recruitment/selection of patients	Consecutive patients seen in thyroid clinic
Age, gender and ethnicity	Age - Mean (SD): 45 (16). Gender (M:F): 9:91. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Extra comments	All patients had colloid nodules
Indirectness of population	No indirectness
Interventions	(n=28) Intervention 1: Levothyroxine therapy. 3ug/kg/d, no titration specified. Duration 6 months. Concurrent medication/care: Nil else stated (n=25) Intervention 2: Placebo/sham. Identical placebo tablet . Duration 6 months. Concurrent medication/care: Nil else specified . Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus PLACEBO/SHAM

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Change in mls nodule volume at end of follow-up at 6 months; Group 1: mean -0.5 (SD 1.2); n=28, Group 2: mean -0.2 (SD 1.2); n=25

Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Grussendorf 2011[24]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=405)
Countries and setting	Conducted in Germany; Setting: 60 German centres
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Non-cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	Caucasian, 18-65, euthyroid, at least one nodule >1cm in diameter, normal or enlarged thyroid, nodule with </= 20% cystic component, non-malignant c.f. guidelines,
Exclusion criteria	Thyroid therapy in last 3 years, autonomous thyroid structure, cysts, CI to iodine, thyroid antibodies, co-existing autoimmune thyroid disease/CHD/endocrine disease/acute illness/pregnancy
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (range): 47 (44.6 - 48.5). Gender (M:F): 27:73. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=206) Intervention 1: Levothyroxine therapy. Initial 75micrograms/day, titrated to a TSH between 0.2 and 0.8 from 3 months initial 75 µg/d, titrated to a TSH between 0.2 and 0.8 from 3 months . Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness

	(n=199) Intervention 2: Placebo/sham. Placebo. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Study funded by industry
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus PLACEBO/SHAM Protocol outcome 1: Percent change in size of nodule/goitre at 3 months - Actual outcome for Non-cystic nodule : % Change from baseline in thyroid nodule volume at 12 months; Group 1: mean 12.1 (SD 38.5); n=206, Group 2: mean 5.2 (SD 41.5); n=199; Comments: SDs calculated from CIs Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 2: Arrhythmia at 3 months - Actual outcome for Non-cystic nodule : Atrial fibrillation at 12 months; Group 1: 0/206, Group 2: 2/199 Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	**Lang 2018[34]**
Study type	Non-randomised comparative study
Number of studies (number of participants)	1 (n=154)
Countries and setting	Conducted in China; Setting: Nil else stated
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months

Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Benign, causing pressure +/- cosmetic symptoms, >10mm but <50mm, solid (<30% cystic), within treatable ablation depth (5-30mm between skin and nodule centre), normal thyroid hormones
Exclusion criteria	Under 18, pregnant/lactating, macrocalcifications, history of H&N irradiation, FMH of thyroid cancer, pre-existing vocal cord palsy, medical conditions precluding sedation
Recruitment/selection of patients	Nil else stated
Age, gender and ethnicity	Age - Mean (SD): 50 (14). Gender (M:F): 20:80. Ethnicity: Not stated
Further population details	1. Age groups: 2. Surgical indication:
Indirectness of population	No indirectness
Interventions	(n=77) Intervention 1: High intensity focused ultrasound. Used EchoPulse, diazepam and pethedine before treatment, each subunit of nodule received 8 second pulse of HIFU energy followed by 30 seconds of cooling, nearby structures marked and left unaffected, oral diet resumed immediately, discharged 2 hours after treatment. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=77) Intervention 2: Surgery. Open lobectomy via cervical approach, RLN and external branch of superior laryngeal nerve sought, mapped and monitored using IONM, strap muscles not transected routinely, no drain after procedure, oral diet resumed immediately afterwards, allowed to go home after one overnight hospital stay. Duration 6 months. Concurrent medication/care: Usual care . Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: HIGH INTENSITY FOCUSED ULTRASOUND versus SURGERY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule <x cm: % nodule shrinkage at 6 months at 6 months; Group 1: mean 64 (SD 26); n=77, Group 2: mean 100 (SD 0); n=77
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 2: Hypothyroidism at 3 months

- Actual outcome for Non-cystic nodule <x cm: Hypothyroidism at 6 months at 6 months; Group 1: 0/77, Group 2: 4/77 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 3: Recurrent laryngeal nerve damage at 3 months - Actual outcome for Non-cystic nodule <x cm: RLN damage (all recovered within 3 months) at 6 months; Group 1: 3/77, Group 2: 3/77 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 4: Bleeding at 3 months - Actual outcome for Non-cystic nodule <x cm: Bleeding (requiring re-exploration) at 6 months; Group 1: 0/77, Group 2: 1/77 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 5: Infection at 3 months - Actual outcome for Non-cystic nodule <x cm: Infection at 6 months; Group 1: 0/77, Group 2: 0/77 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 6: Compressive symptoms at 3 months - Actual outcome for Non-cystic nodule <x cm: Any improvement in compressive symptoms (slight/moderate/significant) at 6 months; Group 1: 73/77, Group 2: 67/77 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 ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Patient/family/carer experience of care at 3 months

Study	Larijani 1999[36]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=62)
Countries and setting	Conducted in Iran; Setting: Referred to Endocrine outpatient department
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	One benign nodule by FNAB
Exclusion criteria	Multiple nodules, suspicious/malignant findings on FNAB, pregnancy, history of CV disease, younger than 16 or older than 60
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (SD): 32.9 (10.8). Gender (M:F): 23:77. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=32) Intervention 1: Levothyroxine therapy. Daily dose of 1.5 to 2.0 micrograms/kg body weight, . Duration 12 months. Concurrent medication/care: Not stated. Indirectness: No indirectness (n=30) Intervention 2: Placebo/sham. Placebo. Duration 12 months. Concurrent medication/care: Not stated. Indirectness: No indirectness
Funding	Academic or government funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus PLACEBO/SHAM

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Nodule volume at end of follow-up at 12 months; Group 1: mean 12.4 (SD 16.7); n=32, Group 2: mean 11.7 (SD 13.6); n=30
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Larijani 2005[35]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=58)
Countries and setting	Conducted in Iran; Setting: Iran Endocrinology centre
Line of therapy	1st line
Duration of study	Intervention + follow up: 2 years
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Single palpable nodule, benign on FNAB
Exclusion criteria	History of T4 use, abnormal thyroid hormones, multiple nodules on exam, pregnant, history of CVD, younger than 15 or older than 60
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (SD): 36 (10). Gender (M:F): Define. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=31) Intervention 1: Levothyroxine therapy. 1.5-2 microg/kg/d, titrated to TSH <0.1mIU/ml. Duration 2 years. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=27) Intervention 2: Monitoring only. Nil else stated. Duration 2 years. Concurrent medication/care: Usual care . Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus MONITORING ONLY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome: Nodule volume ml at end at 24 months; Group 1: mean 10.8 (SD 9.7); n=31, Group 2: mean 11.6 (SD 8.5); n=27
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Papini 1993[39]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=110)
Countries and setting	Conducted in Italy; Setting: Italy
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Non-cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	Single nodule, colloid on FNA, US showing solid (i.e. <1ml fluid volume), decreased/normal uptake on scan, normal thyroid antibodies, normal thyroid hormones, diagnosis no more than 2 years before enrollment, no previous treatment for nodules, 18-60 years old, no clinically relevant CV/hepatic/pulmonary/renal disease
Exclusion criteria	-
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (SD): 43 (11). Gender (M:F): Define. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=54) Intervention 1: Levothyroxine therapy. 2 micrograms/kg day, adjusted to induce TSH suppression. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=56) Intervention 2: Placebo/sham. Nil else specified. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus PLACEBO/SHAM

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Nodule volume ml at end of follow-up at 12 months; Group 1: mean 5.86 (SD 8.9); n=51, Group 2: mean 6.48 (SD 7.4); n=50
Risk of bias: All domain - Low. Selection - Low. Blinding - Low. Incomplete outcome data - Low. Outcome reporting - Low. Measurement - Low. Crossover - Low:

Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 6	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Papini 1998[41]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=83)
Countries and setting	Conducted in Italy; Setting: Outpatient clinic in Rome
Line of therapy	1st line
Duration of study	Intervention + follow up: 5 years
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Single palpable nodule, greatest diameter 10-30mm, colloid by FNAB, solid on US, hypofunctioning, normal thyroid hormones, no previous treatment
Exclusion criteria	Nil else
Recruitment/selection of patients	100 consecutive patients at the centre
Age, gender and ethnicity	Age - Mean (SD): 42 (13). Gender (M:F): 16:84. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=42) Intervention 1: Levothyroxine therapy. Daily dose of 2.0 micrograms/kg . Duration 5 years. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=41) Intervention 2: Monitoring only. Nil else specified. Duration 5 years. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus MONITORING ONLY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule: Final nodule volume (ml) at end of follow-up at 5 years; Group 1: mean 1.45 (SD 1.17); n=42, Group 2: mean 2.12 (SD 1.46); n=41
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Group 1 Number missing: 8, Reason: Not specified ; Group 2 Number missing: 6, Reason: Not specified	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	**Papini 2007[40]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=41)
Countries and setting	Conducted in Italy; Setting: Outpatient clinic in Rome
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Non-cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	Single/dominant nodule, <20% fluid, volume >5ml, at least one diameter >30mm, hypoactive on scintiscan, benign on 2 consecutive FNABs, normal TSH/thyroid hormones/thyroid antibodies, 18 to 60 years old, refused/ineligible for surgery, untreated thyroid disease
Exclusion criteria	Nil else
Recruitment/selection of patients	2500 patients screened, 86 met criteria
Age, gender and ethnicity	Age - Mean (SD): 47 (7). Gender (M:F): 12:88. Ethnicity: Not stated

Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=21) Intervention 1: Levothyroxine therapy. 1.5 micrograms/kg/day modification of dose at day 35 increased if TSH >0.3 microIU/ml, decreased if undetectable/nervousness/tremors/tachycardia present. Duration 12 months . Concurrent medication/care: Nil else. Indirectness: No indirectness (n=20) Intervention 2: Monitoring only. Follow-up only. Duration 12 months . Concurrent medication/care: Nil else specified. Indirectness: No indirectness (n=21) Intervention 3: Monitoring only. US guided, pair of spinal needles inserted into lesions under US, optical fibre guided in through needles, all treatments done in one session but due to variability in size of nodules 6 patients had 1 illumination, 14 had 2 illuminations through two needles and one patient had 2 illuminations through four needles. 3W power, illuminated for 10 minutes, 1800J per fibre per treatment. IM steroids given before procedure. Duration 12 months. Concurrent medication/care: Usual care . Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus MONITORING ONLY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Change in nodule volume (ml) at end of follow up at 12 months; Group 1: mean -0.6 (SD 2.3); n=21, Group 2: mean 0.7 (SD 2.2); n=20
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Compressive symptoms at 3 months
- Actual outcome for Non-cystic nodule : Improvement in symptoms at 12 months; Group 1: 2/15, Group 2: 0/14
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: PERCUTANEOUS LASER ABLATION versus LEVOTHYROXINE THERAPY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months - Actual outcome for Non-cystic nodule : Change in nodule volume (ml) at end of follow up at 12 months; Group 1: mean -5.2 (SD 3.1); n=21, Group 2: mean -0.6 (SD 2.3); n=21 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 2: Compressive symptoms at 3 months - Actual outcome for Non-cystic nodule : Improvement in symptoms at 12 months; Group 1: 13/16, Group 2: 2/15 Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0 RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: PERCUTANEOUS LASER ABLATION versus MONITORING ONLY Protocol outcome 1: Percent change in size of nodule/goitre at 3 months - Actual outcome for Non-cystic nodule : Change in nodule volume (ml) at end of follow up at 12 months; Group 1: mean -5.2 (SD 3.1); n=21, Group 2: mean 0.7 (SD 2.2); n=20 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 2: Compressive symptoms at 3 months - Actual outcome for Non-cystic nodule : Improvement in symptoms at 12 months; Group 1: 13/16, Group 2: 0/14 Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Patient/family/carer experience of care at 3 months

Study	Reverter 1992[42]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=40)
Countries and setting	Conducted in Spain; Setting: Not stated
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Non-cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	Female, single nodule on palpation, cold on scan, benign by FNAB,
Exclusion criteria	Hashimoto's, pregnancy, CIs to T4 treatment
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (SD): 40 (10). Gender (M:F): 0:100. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Levothyroxine therapy. 100ug/d T4 for 2 weeks and 200ug/day thereafter, titrated to TSH <0.1mIU/l (100% suppressed). Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=20) Intervention 2: Monitoring only. Nil else specified. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus MONITORING ONLY

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Final volume of nodules at 12 months; Group 1: mean 10.1 (SD 12.2); n=20, Group 2: mean 9.2 (SD 9.5); n=20
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Group 1 Number missing: 6, Reason: Chose surgery for cosmetic reasons; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Sung 2013[45]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=50)
Countries and setting	Conducted in South Korea; Setting: Not stated
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Cystic nodules (>90%), pressure symptoms/cosmetic problems, benign by 2 FNABs, normal thyroid hormones, no previous treatment for nodules in last 6 months
Exclusion criteria	Nil else
Recruitment/selection of patients	Not stated
Age, gender and ethnicity	Age - Mean (SD): 42 (10). Gender (M:F): 10:90. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Compressive symptoms
Indirectness of population	No indirectness
Interventions	(n=25) Intervention 1: Radiofrequency ablation. 1 dose, 50-70W RF power, moving-shot technique. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=25) Intervention 2: Ethanol ablation. 1 dose 99% ethanol, aimed for 50% of volume aspirated and replaced. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Other (Some funding to the centre as a whole provided by pharma)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: RADIOFREQUENCY ABLATION versus ETHANOL ABLATION

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Cystic nodule : % volume reduction at end of follow-up at 6 months; Group 1: mean 93.3 (SD 5.4); n=21, Group 2: mean 96.9 (SD 4.1); n=21
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness : Baseline details: Mean volume in RFA 12.2ml, mean in PEI 9.3ml : Group 1 Number missing: 4, Reason: Not stated; Group 2

Number missing: 4, Reason: Not stated Protocol outcome 2: Pain at 3 months - Actual outcome for Cystic nodule : Grade 1 pain or more during and after procedure at 6 months; Group 1: 21/25, Group 2: 0/25 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 3: Compressive symptoms at 3 months - Actual outcome for Cystic nodule : Symptom score (VAS, 0-10, higher is worse) at 6 months; Group 1: mean 0.5 (SD 0.8); n=21, Group 2: mean 0.5 (SD 0.7); n=21 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 ; Group 1 Number missing: 4, Reason: Not stated; Group 2 Number missing: 4, Reason: Not stated	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Patient/family/carer experience of care at 3 months

Study	Wemeau 2002[48]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=123)
Countries and setting	Conducted in France; Setting: France, multicentre, nil else specified
Line of therapy	1st line
Duration of study	Intervention + follow up: 18 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Non-cystic nodule
Subgroup analysis within study	Not applicable
Inclusion criteria	Single nodule, benign by FNAB, 18-55
Exclusion criteria	More than one nodule, history of CVD/osteoporosis/thyroid surgery/neck irradiation/thyroiditis, abnormal thyroid hormone/TSH levels, autoantibodies, nodules with cystic >20%, hot nodules
Recruitment/selection of patients	Nil specified
Age, gender and ethnicity	Age - Mean (SD): 39 (9). Gender (M:F): 10:90. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=64) Intervention 1: Levothyroxine therapy. Initial dose of 2.5ug/kg/d, titrated after 4 weeks to maintain TSH <0.3mIU/L. Duration 18 months. Concurrent medication/care: Seen after 3, 6, 12, 18 months of treatment, beta blocker prescription allowed if tachycardia. Indirectness: No indirectness (n=59) Intervention 2: Placebo/sham. Nil else. Duration 18 months. Concurrent medication/care: Seen after 3, 6, 12, 18 months of treatment, beta blocker prescription allowed if tachycardia. Indirectness: No indirectness
Funding	Equipment / drugs provided by industry

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus PLACEBO/SHAM

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Change in volume of nodule (ml) by end of follow-up at 18 months; Group 1: mean -0.36 (SD 1.71); n=64, Group 2: mean 0.62 (SD 3.67): n=59

Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6, Reason: Not specified; Group 2 Number missing: 11, Reason: Not specified Protocol outcome 2: Hyperthyroidism at 3 months - Actual outcome for Non-cystic nodule : Severe hyperthyroidism (nil else specified) at 18 months; Group 1: 1/64, Group 2: 1/59 Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6, Reason: Not specified; Group 2 Number missing: 11, Reason: Not specified	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Yue 2016[50]
Study type	Non-randomised comparative study
Number of studies (number of participants)	1 (n=216)
Countries and setting	Conducted in China; Setting: China, RFA carried out by group with 5 years' experience, surgery by group with 6 years' experience
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Propensity matched cohort, euthyroid, treated with either RFA or surgery (indication = nodule + compressive symptoms/cosmetic problems/anxiety about malignancy), for RFA also benign on FNAC and no changes on US for at least 12 months
Exclusion criteria	Major comorbidities
Recruitment/selection of patients	NRS, retrospective cohort study
Age, gender and ethnicity	Age - Mean (SD): 51 (13). Gender (M:F): 34:66. Ethnicity: 97% Han Chinese
Further population details	1. Age groups: 50-65 2. Surgical indication: Compressive symptoms
Extra comments	PS matching on sex, age, nodule volume, ethnicity, SF-36 scores, smoking, drinking, chronic disease, marital status, monthly income, education, current job
Indirectness of population	Serious indirectness: Some in surgery group with ?malignancy
Interventions	(n=108) Intervention 1: Radiofrequency ablation. Moving shot technique, hydrodissection, only the largest nodules was treated. Duration 6 months. Concurrent medication/care: Usual care . Indirectness: No indirectness (n=108) Intervention 2: Surgery. Open hemithyroidectomy under GA. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Academic or government funding
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: RADIOFREQUENCY ABLATION versus SURGERY	

Protocol outcome 1: Quality of life at 3 months - Actual outcome: SF-36 (general health) at end of follow-up at 6 months; MD; 1.8, Comments: SD (estimated) 6.0; Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: - Actual outcome: SF-36 (vitality) at end of follow-up at 6 months; MD; 3.8, Comments: SD (estimated) 8.4; Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: - Actual outcome: SF-36 (mental health) at end of follow-up at 6 months; MD; 1.6, Comments: SD (estimated from p value) 5.6; Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Mortality at 1 year; Percent change in size of nodule/goitre at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	Zelmanovitz 1998[53]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=45)
Countries and setting	Conducted in Brazil; Setting: Secondary care in Brazil
Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Single nodule, hypofunctioning, benign
Exclusion criteria	>20% cystic, history of HT/CVD
Recruitment/selection of patients	Consecutive patients screened
Age, gender and ethnicity	Age - Mean (SD): 43 (12). Gender (M:F): 7:93. Ethnicity: Not stated
Further population details	1. Age groups: 18-50 2. Surgical indication: Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=24) Intervention 1: Levothyroxine therapy. Initial dose of 2.5-3.0microg/kg/d, after 1.5 months titrated to TSH <0.3microIU/ml. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=27) Intervention 2: Placebo/sham. Placebo. Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness
Funding	Equipment / drugs provided by industry

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: LEVOTHYROXINE THERAPY versus PLACEBO/SHAM

Protocol outcome 1: Percent change in size of nodule/goitre at 3 months
- Actual outcome for Non-cystic nodule : Volume (ml) at end of follow-up at 12 months; Group 1: mean 17.7 (SD 31.6); n=21, Group 2: mean 11.6 (SD 9.7); n=24
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low: Indirectness of outcome: No indirectness : Baseline details: T4 mean volume 16.4ml, PBO mean volume 13.6ml; Group 1 Number missing: 3, Reason: Lost to follow-

up; Group 2 Number missing: 3, Reason: Lost to follow-up	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Recurrent laryngeal nerve damage at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Pain at 3 months; Compressive symptoms at 3 months; Patient/family/carer experience of care at 3 months

Study	**Zhi 2018[54]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=60)
Countries and setting	Conducted in China; Setting: Consecutive patients referred to specialist centre in China

Line of therapy	1st line
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Benign by 2x FNAs, max diameter greater than 2 cm, continuing to grow, structural symptoms or clinical concern, increasing rapidly in volume or patients anxious about malignant transformation
Exclusion criteria	Current pregnancy/lactation, head or neck irradiation history, head or neck surgery history, medical conditions precluding intravenous sedation, vocal cord palsy
Recruitment/selection of patients	Nil else stated
Age, gender and ethnicity	Age - Mean (SD): 53 (10). Gender (M:F): 34/18. Ethnicity: Not stated
Further population details	1. Age groups: 2. Surgical indication:
Extra comments	Mix of nodule type (15 solid, 37 cystic)
Indirectness of population	No indirectness
Interventions	(n=30) Intervention 1: Microwave ablation. Outpatient setting, KY-2000, LA with lidocaine, cystic components aspirated with ethanol, 1-2mm skin incision, 30-50W power and frequency of 2450MHz, observation for "a few hours" after procedure . Duration 12 months. Concurrent medication/care: Usual care. Indirectness: No indirectness (n=30) Intervention 2: Surgery. GA, nerve monitoring, 2-4cm incision, linea alba cervicalis incised, cervical anterior muscles retracted, thyroid medium vein ligated or closed by clipping, vagus nerve identified, thyroid isthmus removed to expose trachea, unilateral thyroid gland completely resected. Duration 12 months. Concurrent medication/care: Usual care . Indirectness: No indirectness

Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MICROWAVE ABLATION versus SURGERY Protocol outcome 1: Recurrent laryngeal nerve damage at 3 months - Actual outcome: Hoarseness (temporary) at 12 months; Group 1: 1/28, Group 2: 2/24 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - High, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: More solid nodules in surgery group (42% vs 18%); Group 1 Number missing: 2, Reason: 2 missing because volume reduction was low so excluded as needed more treatment; Group 2 Number missing: 6, Reason: 6 missing as had contralateral nodule so needed total thyroidectomy Protocol outcome 2: Pain at 3 months - Actual outcome: Post-operative pain (temporary) at 12 months; Group 1: 2/28, Group 2: 22/24 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - High, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: More solid nodules in surgery group (42% vs 18%); Group 1 Number missing: 2, Reason: 2 missing because volume reduction was low so excluded as needed more treatment; Group 2 Number missing: 6, Reason: 6 missing as had contralateral nodule so needed total thyroidectomy Protocol outcome 3: Compressive symptoms at 3 months - Actual outcome: Compressive symptoms resolution at 12 months; Group 1: 10/10, Group 2: 11/11 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - High, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: More solid nodules in surgery group (42% vs 18%); Group 1 Number missing: 2, Reason: 2 missing because volume reduction was low so excluded as needed more treatment; Group 2 Number missing: 6, Reason: 6 missing as had contralateral nodule so needed total thyroidectomy	
Protocol outcomes not reported by the study	Mortality at 1 year; Quality of life at 3 months; Percent change in size of nodule/goitre at 3 months; Malignancy at 3 months; Hypothyroidism at 3 months; Hyperthyroidism at 3 months; Hypoparathyroidism at 3 months; Bleeding at 3 months; Infection at 3 months; Arrhythmia at 3 months; Osteoporosis at 3 months; Patient/family/carer experience of care at 3 months

Appendix E: Forest plots

E.1 Non-cystic nodules – T4 vs placebo/follow-up only

Figure 2: Nodule volume (ml, 6-60 months)

Study or Subgroup	T4 Mean	T4 SD	T4 Total	PBO Mean	PBO SD	PBO Total	Weight	Mean Difference IV, Fixed, 95% CI
Cesareo 2010	0.8	0.8	36	2.5	2.4	35	16.5%	-1.70 [-2.54, -0.86]
Gharib 1987	-0.5	1.2	28	-0.2	1.2	25	27.6%	-0.30 [-0.95, 0.35]
Larijani 1999	12.4	16.7	32	11.7	13.6	30	0.2%	0.70 [-6.86, 8.26]
Larijani 2005	10.8	9.7	31	11.6	8.5	27	0.5%	-0.80 [-5.48, 3.88]
Ozkaya 2010	0.75	8	33	0.35	4	25	1.2%	0.40 [-2.75, 3.55]
Papini 1993	5.86	8.9	51	6.48	7.36	50	1.1%	-0.62 [-3.80, 2.56]
Papini 1998	1.45	1.17	42	2.12	1.46	41	35.6%	-0.67 [-1.24, -0.10]
Papini 2007	-0.6	2.3	21	0.7	2.2	20	6.1%	-1.30 [-2.68, 0.08]
Wemeau 2002	-0.36	1.71	64	0.62	3.67	59	11.0%	-0.98 [-2.01, 0.05]
Zelmanovitz 1998	17.7	31.6	21	11.6	9.7	24	0.1%	6.10 [-7.96, 20.16]
Total (95% CI)			**359**			**336**	**100.0%**	**-0.79 [-1.13, -0.45]**

Heterogeneity: Chi² = 9.20, df = 9 (P = 0.42); I² = 2%
Test for overall effect: Z = 4.56 (P < 0.00001)

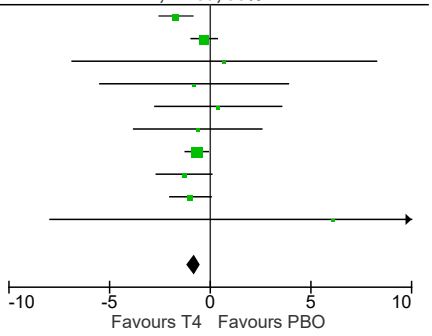


Figure 3: Nodule volume (% reduction, 12 months)

Study or Subgroup	T4 Mean	T4 SD	T4 Total	PBO Mean	PBO SD	PBO Total	Weight	Mean Difference IV, Fixed, 95% CI
Grussendorf 2011	12.1	38.5825	206	5.2	41.49	199	100.0%	6.90 [-0.91, 14.71]
Total (95% CI)			**206**			**199**	**100.0%**	**6.90 [-0.91, 14.71]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.73 (P = 0.08)

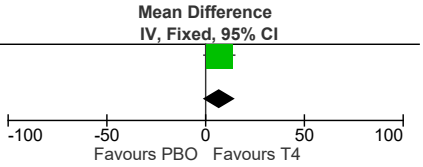


Figure 4: Thyroid volume (ml, 12 months)

Study or Subgroup	T4 Mean	T4 SD	T4 Total	FU Mean	FU SD	FU Total	Weight	Mean Difference IV, Fixed, 95% CI
Cesareo 2010	10.6	4.9	36	20	14.7	35	100.0%	-9.40 [-14.53, -4.27]
Total (95% CI)			**36**			**35**	**100.0%**	**-9.40 [-14.53, -4.27]**

Heterogeneity: Not applicable
Test for overall effect: Z = 3.59 (P = 0.0003)

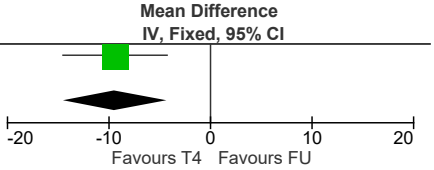


Figure 5: Hyperthyroidism (18 months)

Study or Subgroup	T4 Events	T4 Total	PBO Events	PBO Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Wemeau 2002	1	64	1	59	100.0%	0.92 [0.06, 14.41]
Total (95% CI)		**64**		**59**	**100.0%**	**0.92 [0.06, 14.41]**
Total events	1		1			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.06 (P = 0.95)

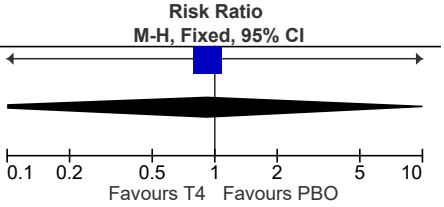


Figure 6: Improvement in symptoms (12 months)

Study or Subgroup	T4 Events	T4 Total	PBO Events	PBO Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Papini 2007	2	15	0	14	100.0%	7.43 [0.44, 125.12]
Total (95% CI)		**15**		**14**	**100.0%**	**7.43 [0.44, 125.12]**
Total events	2		0			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.39 (P = 0.16)

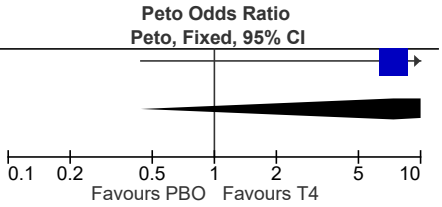
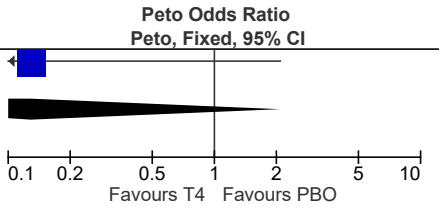


Figure 7: Arrhythmias (AF, 12 months)

Study or Subgroup	T4 Events	T4 Total	PBO Events	PBO Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Grussendorf 2011	0	206	2	199	100.0%	0.13 [0.01, 2.09]
Total (95% CI)		**206**		**199**	**100.0%**	**0.13 [0.01, 2.09]**
Total events	0		2			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.44 (P = 0.15)



E.2 Non-cystic nodules – RFA vs follow-up

Figure 8: Nodule volume (ml, 6 months)

Study or Subgroup	RFA Mean	RFA SD	RFA Total	FU Mean	FU SD	FU Total	Weight	Mean Difference IV, Fixed, 95% CI
Cesareo 2015	8.6	9.5	42	27.8	22.1	42	3.4%	-19.20 [-26.48, -11.92]
Deandrea 2015	4.7	2.7	40	15.2	3.5	40	96.6%	-10.50 [-11.87, -9.13]
Total (95% CI)			**82**			**82**	**100.0%**	**-10.80 [-12.14, -9.45]**

Heterogeneity: Chi² = 5.31, df = 1 (P = 0.02); I² = 81%
Test for overall effect: Z = 15.72 (P < 0.00001)

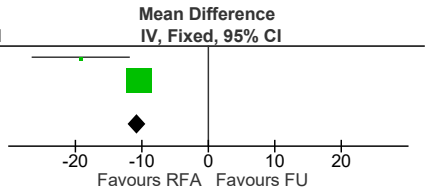


Figure 9: Hypothyroidism (6 months)

Study or Subgroup	RFA Events	RFA Total	FU Events	FU Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Cesareo 2015	0	42	0	42		Not estimable
Total (95% CI)		**42**		**42**		**Not estimable**
Total events	0		0			

Heterogeneity: Not applicable
Test for overall effect: Not applicable

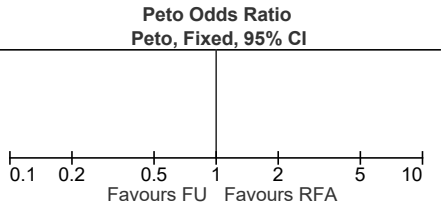


Figure 10: Hyperthyroidism (6 months)

Study or Subgroup	RFA Events	RFA Total	FU Events	FU Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Cesareo 2015	1	42	0	42	100.0%	7.39 [0.15, 372.38]
Total (95% CI)		**42**		**42**	**100.0%**	**7.39 [0.15, 372.38]**
Total events	1		0			

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

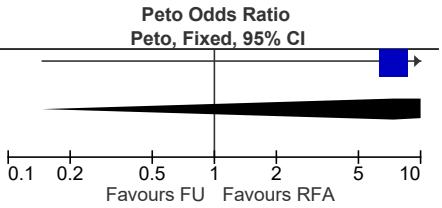
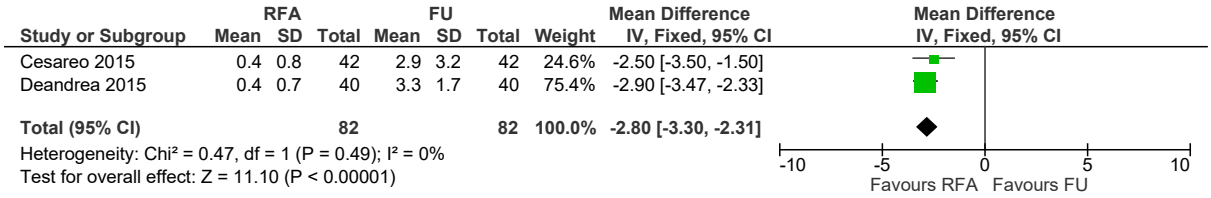


Figure 11: Compressive symptoms (6 months, 0-10, VAS, higher is worse)



E.3 Non-cystic nodules – LA vs FU

Figure 12: Change in nodule volume (ml, 12 months)

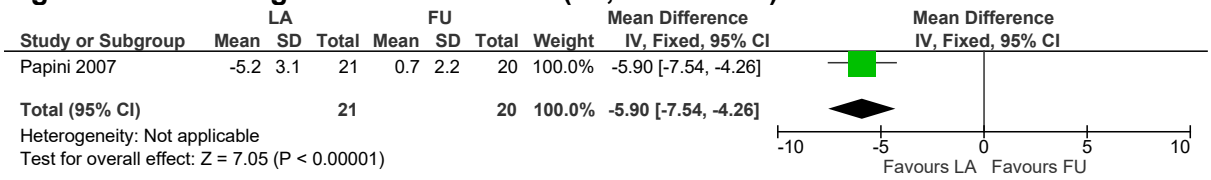
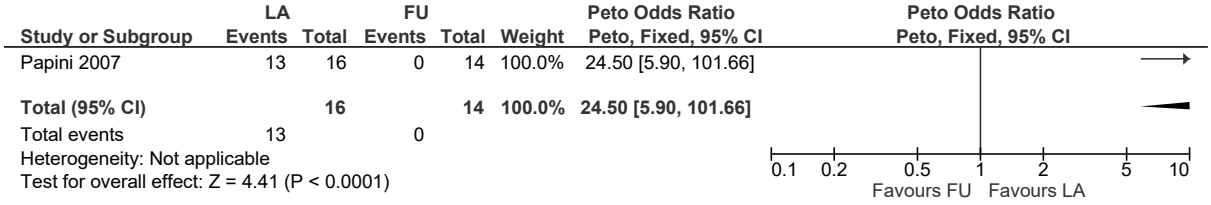
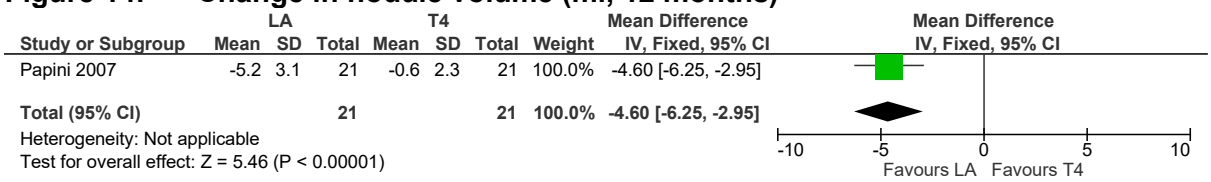


Figure 13: Improvement in symptoms (12 months)



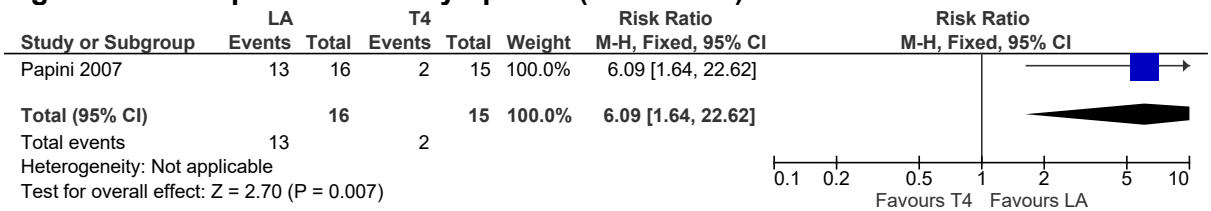
E.4 Non-cystic nodules – LA vs T4

Figure 14: Change in nodule volume (ml, 12 months)



<Insert Note here>

Figure 15: Improvement in symptoms (12 months)



E.5 Non-cystic nodules – PEI vs T4

Figure 16: Reduction in nodule volume (%, 12 months)

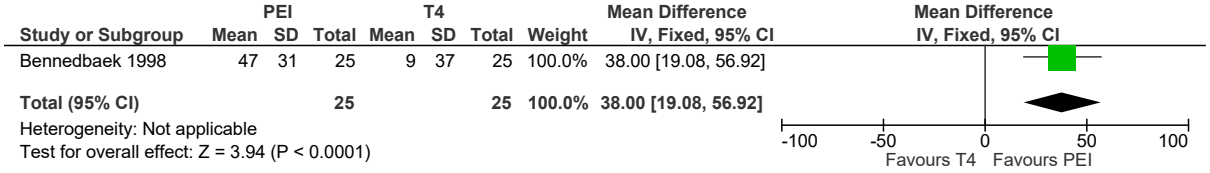


Figure 17: Compressive symptoms (any improvement, 12 months)

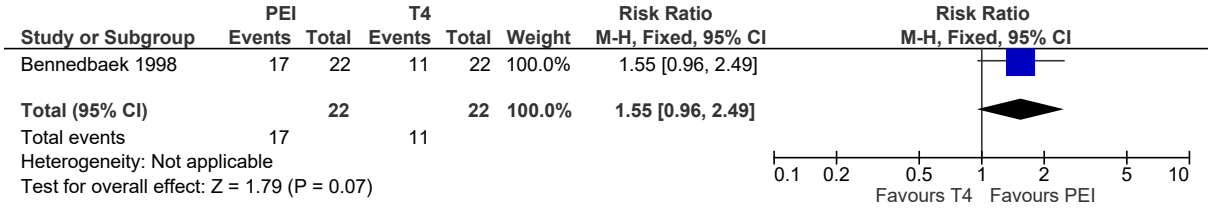


Figure 18: Pain (on procedure, 12 months)

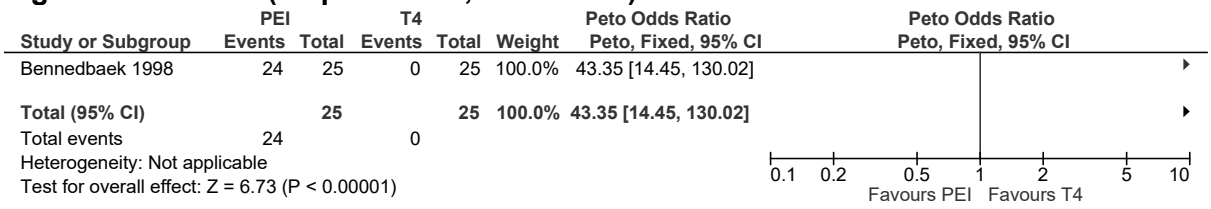
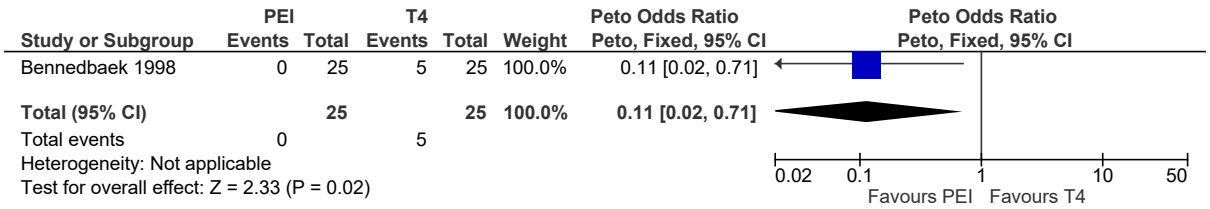


Figure 19: Hyperthyroidism (symptoms, 12 months)



E.6 Non-cystic nodules – HIFU vs surgery

Figure 20: Hypothyroidism (6 months)

Study or Subgroup	HIFU Events	HIFU Total	Surgery Events	Surgery Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Lang 2018	0	77	4	77	100.0%	0.13 [0.02, 0.94]
Total (95% CI)		**77**		**77**	**100.0%**	**0.13 [0.02, 0.94]**
Total events	0		4			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.02 (P = 0.04)

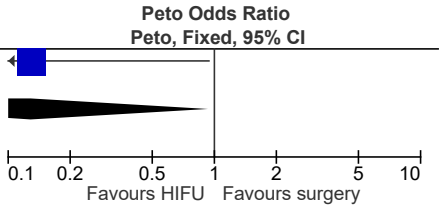


Figure 21: RLN damage (temporary, 6 months)

Study or Subgroup	HIFU Events	HIFU Total	Surgery Events	Surgery Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Lang 2018	3	77	3	77	100.0%	1.00 [0.21, 4.80]
Total (95% CI)		**77**		**77**	**100.0%**	**1.00 [0.21, 4.80]**
Total events	3		3			

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

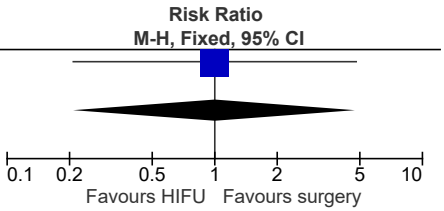


Figure 22: Bleeding (6 months)

Study or Subgroup	HIFU Events	HIFU Total	Surgery Events	Surgery Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Lang 2018	0	77	1	77	100.0%	0.14 [0.00, 6.82]
Total (95% CI)		**77**		**77**	**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)

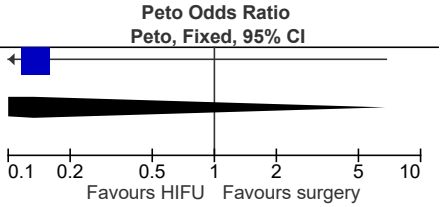


Figure 23: Infection (6 months)

Study or Subgroup	HIFU Events	HIFU Total	Surgery Events	Surgery Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Lang 2018	0	77	0	77		Not estimable
Total (95% CI)		**77**		**77**		**Not estimable**
Total events	0		0			

Heterogeneity: Not applicable
Test for overall effect: Not applicable

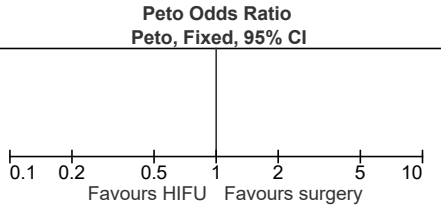
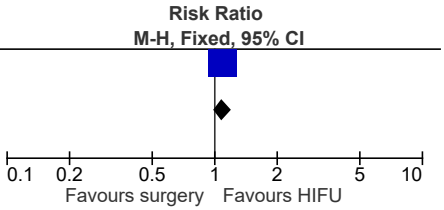


Figure 24: Improvement in symptoms (any (slight/moderate/severe), 6 months)

Study or Subgroup	HIFU Events	HIFU Total	Surgery Events	Surgery Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Lang 2018	73	77	67	77	100.0%	1.09 [0.98, 1.21]
Total (95% CI)		**77**		**77**	**100.0%**	**1.09 [0.98, 1.21]**
Total events	73		67			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.67 (P = 0.10)



E.7 Cystic nodules – RFA vs PEI

Figure 25: Nodule volume (% reduction, 6 months)

Study or Subgroup	RFA Mean	RFA SD	RFA Total	PEI Mean	PEI SD	PEI Total	Weight	Mean Difference IV, Fixed, 95% CI
Baek 2015	87.5	11.5	22	82.4	28.6	24	5.2%	5.10 [-7.31, 17.51]
Sung 2013	93.3	5.4	21	96.9	4.1	21	94.8%	-3.60 [-6.50, -0.70]
Total (95% CI)			**43**			**45**	**100.0%**	**-3.15 [-5.97, -0.33]**

Heterogeneity: Chi² = 1.79, df = 1 (P = 0.18); I² = 44%
Test for overall effect: Z = 2.19 (P = 0.03)

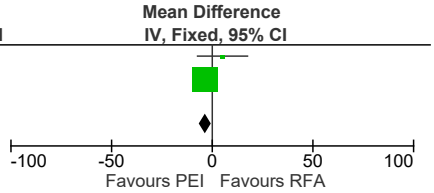


Figure 26: Compressive symptoms (0-10, VAS, higher is worse, 6 months)

Study or Subgroup	RFA Mean	RFA SD	RFA Total	PEI Mean	PEI SD	PEI Total	Weight	Mean Difference IV, Fixed, 95% CI
Baek 2015	0.2	0.4	22	0.7	1.3	24	40.3%	-0.50 [-1.05, 0.05]
Sung 2013	0.5	0.8	22	0.5	0.7	21	59.7%	0.00 [-0.45, 0.45]
Total (95% CI)			**44**			**45**	**100.0%**	**-0.20 [-0.55, 0.15]**

Heterogeneity: Chi² = 1.92, df = 1 (P = 0.17); I² = 48%
Test for overall effect: Z = 1.14 (P = 0.25)

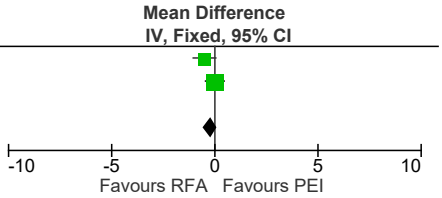
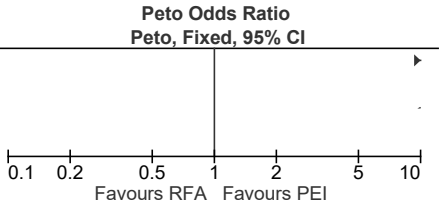


Figure 27: Pain (mild or greater, during procedure)

Study or Subgroup	RFA Events	RFA Total	PEI Events	PEI Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Sung 2013	21	25	0	25	100.0%	29.35 [9.65, 89.23]
Total (95% CI)		**25**		**25**	**100.0%**	**29.35 [9.65, 89.23]**
Total events	21		0			

Heterogeneity: Not applicable
Test for overall effect: Z = 5.96 (P < 0.00001)



E.8 Non-specific nodules – RFA vs surgery

Figure 28: Quality of life (SF-36, general health, 0-100, higher is better)

Study or Subgroup	Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Yue 2016	1.8	0.8165	108	108	100.0%	1.80 [0.20, 3.40]
Total (95% CI)			**108**	**108**	**100.0%**	**1.80 [0.20, 3.40]**

Heterogeneity: Not applicable
Test for overall effect: Z = 2.20 (P = 0.03)

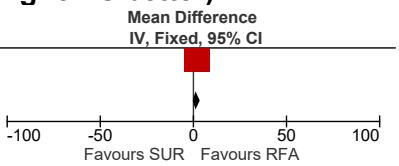


Figure 29: Quality of life (SF-36, vitality, 0-100, higher is better)

Study or Subgroup	Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Yue 2016	3.8	1.1431	108	108	100.0%	3.80 [1.56, 6.04]
Total (95% CI)			**108**	**108**	**100.0%**	**3.80 [1.56, 6.04]**

Heterogeneity: Not applicable
Test for overall effect: Z = 3.32 (P = 0.0009)

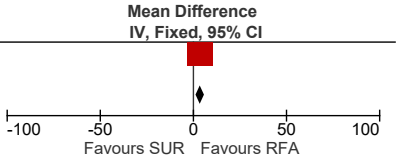
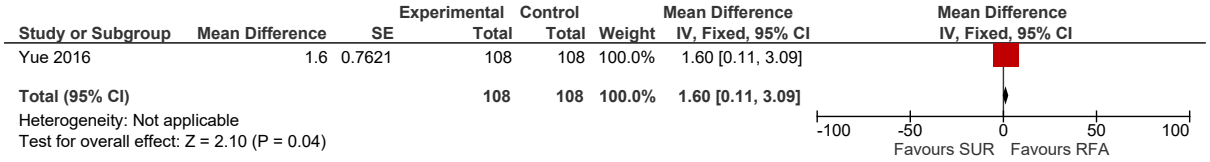


Figure 30: Quality of life (SF-36, mental health, 0-100, higher is better)



E.9 Non-specific nodules – Microwave ablation vs surgery

Figure 31: Hoarseness (temporary, 12 months)

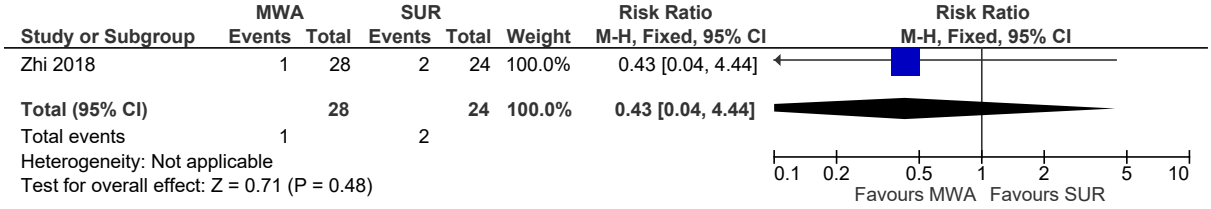


Figure 32: Post-operative pain (temporary, 12 months)

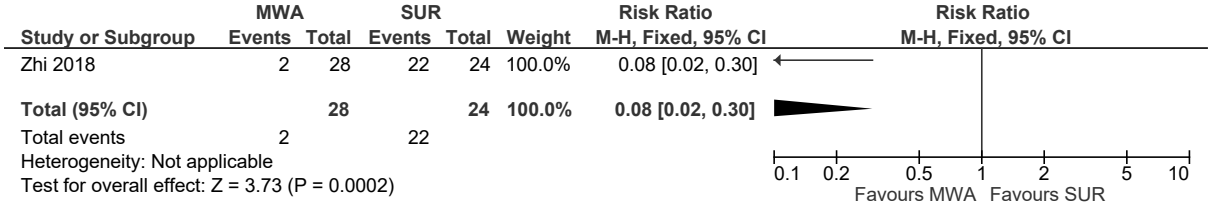
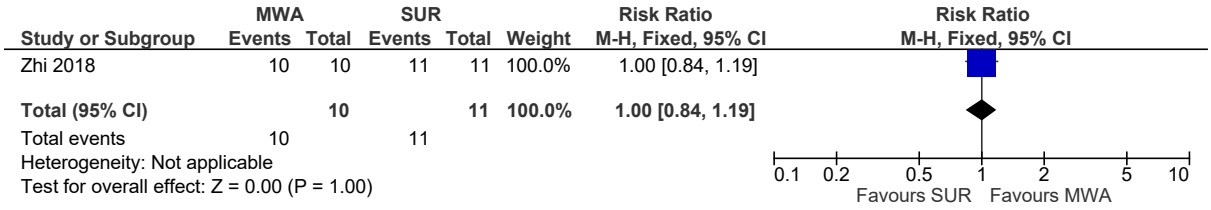


Figure 33: Resolution of compressive symptoms (12 months)



Appendix F: GRADE tables

Table 17: Clinical evidence profile: T4 vs PBO/FU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	T4	PBO/FU	Relative (95% CI)	Absolute		
Nodule volume (follow-up 6-60 months; measured with: ml)												
11	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	379	356	-	MD 0.79 lower (1.13 to 0.45 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Reduction in nodule volume (follow-up 12 months; measured with: %)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	206	199	-	MD 6.9 higher (0.91 lower to 14.71 higher)	⊕⊕⊕⊕ HIGH	IMPORTANT
Thyroid volume (follow-up 12 months; measured with: ml)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	36	35	-	MD 9.4 lower (14.53 to 4.27 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Hyperthyroidism (follow-up 18 months; assessed with: 'Severe' (nil else specified))												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/64 (1.6%)	1.7%	RR 0.92 (0.06 to 14.41)	1 fewer per 1000 (from 16 fewer to 228 more)	⊕OOO VERY LOW	IMPORTANT
Improvement in symptoms (follow-up 12 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	2/15 (13.3%)	0%	Peto OR 7.43 (0.44 to 125.12)	130 more per 1000 (from 70 fewer to 330 more)[3]	⊕OOO VERY LOW	IMPORTANT
Arrhythmias (follow-up 12 months; assessed with: Reported as AF)												
1	randomised	serious[1]	no serious	no serious	very serious[2]	none	0/206	1%	Peto OR 0.13	9 fewer per 1000 (from	⊕OOO	IMPORTANT

	trials		inconsistency	indirectness			(0%)		(0.01 to 2.09)	10 fewer to 11 more)	VERY LOW	

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[3] Zero events in control arm

Table 18: Clinical evidence profile: RFA vs FU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**RFA**	**Follow-up only**	**Relative (95% CI)**	**Absolute**		
Nodule volume (follow-up 6 months; measured with: ml)												
2	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	82	82	-	MD 10.8 lower (12.14 to 9.45 lower)	⊕⊕⊕⊕ HIGH	IMPORTANT
Hypothyroidism (follow-up 6 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	0/42 (0%)	0/42 (0%)	-[3]	not estimable	⊕OOO VERY LOW	IMPORTANT
Hyperthyroidism (follow-up 6 months)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/42 (2.4%)	0%	Peto OR 7.39 (0.15 to 372.38)	23 more per 1000 (from 40 fewer to 90 more)[4]	⊕OOO VERY LOW	IMPORTANT
Compressive symptoms (follow-up 6 months; measured with: VAS (0-10, higher is worse))												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	82	82	-	MD 2.8 lower (3.3 to 2.31 lower)	⊕⊕⊕O MODERATE	IMPORTANT

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[3] Zero events in either arm
[4] Zero events in control arm

Table 19: Clinical evidence profile: LA vs FU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	LA	Follow-up only	Relative (95% CI)	Absolute		
Change in nodule volume (follow-up 12 months; measured with: ml; Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	21	20	-	MD 5.9 lower (7.54 to 4.26 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Improvement in symptoms (follow-up 12 months)												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[1]	none	-	0%	Peto OR 24.5 (5.9 to 101.6)	810 more per 1000 (from 60 more to 1000 more)[3]	⊕OOO VERY LOW	IMPORTANT

[1] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[2] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[3] Zero events in control arm

Table 20: Clinical evidence profile: LA vs T4

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	LA	T4	Relative (95% CI)	Absolute		
Change in nodule volume (follow-up 12 months; measured with: ml; Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	21	21	-	MD 4.6 lower (6.25 to 2.95 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Improvement in symptoms (follow-up 12 months)												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-	13.3%	RR 6.09 (1.64 to 22.62)	677 more per 1000 (from 85 more to 1000 more)	⊕⊕OO LOW	IMPORTANT

[1] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[2] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Table 21: Clinical evidence profile: PEI vs T4

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PEI	T4	Relative (95% CI)	Absolute		
Reduction in nodule volume (follow-up 12 months; measured with: % changes)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	25	25	-	MD 38 higher (19.08 to 56.92 higher)	⊕⊕OO LOW	IMPORTANT
Compressive symptoms (follow-up 12 months; assessed with: Any improvement)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	17/22 (77.3%)	50%	RR 1.55 (0.96 to 2.49)	275 more per 1000 (from 20 fewer to 745 more)	⊕⊕OO LOW	IMPORTANT
Pain (follow-up 12 months; assessed with: On procedure)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	24/25 (96%)	0%	Peto OR 43.35 (14.45 to 130.02)	960 more per 1000 (from 860 more to 1000 more)[3]	⊕⊕⊕O MODERATE	IMPORTANT
Hyperthyroidism (follow-up 12 months; assessed with: Presence of symptoms and TSH changes)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	0/25 (0%)	20%	Peto OR 0.11 (0.2 to 0.71)	173 fewer per 1000 (from 49 fewer to 152 fewer)	⊕⊕OO LOW	IMPORTANT

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[3] Zero events in control arm

Table 22: Clinical evidence profile: HIFU vs surgery

Quality assessment							No of patients		Effect		Quality	Importance
No of	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	HIFU	Surgery	Relative	Absolute		

studies						considerations			(95% CI)			
Nodule volume (% reduction) (follow-up 6 months; Better indicated by lower values)												
1	observational studies	no serious risk of bias[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	77	77	-	MD 36 higher[2]	⊕⊕OO LOW	IMPORTANT
Hypothyroidism (follow-up 6 months)												
1	observational studies	no serious risk of bias[1]	no serious inconsistency	no serious indirectness	serious[3]	none	0/77 (0%)	5.2%	Peto OR 0.13 (0.02 to 0.94)	45 fewer per 1000 (from 3 fewer to 51 fewer)	⊕OOO VERY LOW	IMPORTANT
RLN palsy (temporary) (follow-up 6 months)												
1	observational studies	no serious risk of bias[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	3/77 (3.9%)	3.9%	RR 1 (0.21 to 4.8)	0 fewer per 1000 (from 31 fewer to 148 more)	⊕OOO VERY LOW	IMPORTANT
Bleeding (requiring re-exploration) (follow-up 6 months)												
1	observational studies	no serious risk of bias[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	0/77 (0%)	1.30%	Peto OR 0.14 (0 to 6.82)	11 fewer per 1000 (from 13 fewer to 69 more)	⊕OOO VERY LOW	IMPORTANT
Infection (follow-up 6 months)												
1	observational studies	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[4]	none	0/77 (0%)	0%	-	not estimable	⊕OOO VERY LOW	IMPORTANT
Improvement in symptoms (follow-up 6 months; assessed with: (any increase (slight/moderate/significant))												
1	observational studies	serious[5]	no serious inconsistency	no serious indirectness	serious[3]	none	73/77 (94.8%)	87%	RR 1.09 (0.98 to 1.21)	78 more per 1000 (from 17 fewer to 183 more)	⊕OOO VERY LOW	IMPORTANT

[1] No additional risk of bias beyond default selection bias for non-randomised studies
[2] 95% confidence intervals cannot be calculated as no SD available for surgery (all participants with whole nodule removed), therefore there is no forest plot displayed in the appendix.
[3] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[4] Zero events in either arm
[5] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Table 23: Clinical evidence profile: RFA vs PEI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	RFA	PEI	Relative (95% CI)	Absolute		
Nodule volume (follow-up 6 months; measured with: % reduction)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	45	-	MD 3.15 lower (5.97 to 0.33 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Compressive symptoms (follow-up 6 months; measured with: VAS, 0-10, higher is worse)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	44	45	-	MD 0.2 lower (0.55 lower to 0.15 higher)	⊕⊕OO LOW	IMPORTANT
Pain (follow-up 6 months; assessed with: Mild or greater, during procedure)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	21/25 (84%)	0%	Pain OR 29.35 (9.65 to 89.23)	840 more per 1000 (from 690 more to 990 more)[3]	⊕⊕⊕O MODERATE	IMPORTANT

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[3] Zero events in control arm

Table 24: Clinical evidence profile: RFA vs Surgery

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	RFA	Surgery	Relative (95% CI)	Absolute		
Quality of life (follow-up 6 months; measured with: SF-36, general health, 0-100, higher is better)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	108	108	-	MD 1.8 higher (0.2 to 3.4 higher)	⊕OOO VERY LOW	IMPORTANT

Quality of life (follow-up 6 months; measured with: SF-36, vitality, 0-100, higher is better)											
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	108	108	-	MD 3.8 higher (1.56 to 6.04 higher)	⊕OOO VERY LOW	IMPORTANT
Quality of life (follow-up 6 months; measured with: SF-36, mental health, 0-100, higher is better)											
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	108	108	-	MD 1.6 higher (0.11 to 3.09 higher)	⊕OOO VERY LOW	IMPORTANT

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

Table 25: Clinical evidence profile: Microwave ablation vs Surgery

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Microwave ablation**	**Surgery**	**Relative (95% CI)**	**Absolute**		
Nodule volume (follow-up 12 months; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious	none	28	24	-[3]	Mean final nodule volume in intervention group 0.69ml[3]	⊕OOO VERY LOW	IMPORTANT
Hoarseness (temporary) (follow-up 12 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	1/28 (3.6%)	8.3%	RR 0.43 (0.04 to 4.44)	47 fewer per 1000 (from 80 fewer to 286 more)	⊕OOO VERY LOW	IMPORTANT
Post-operative pain (temporary) (follow-up 12 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	2/28 (7.1%)	91.7%	RR 0.08 (0.02 to 0.30)	844 fewer per 1000 (from 642 fewer to 899 fewer)	⊕⊕OO LOW	IMPORTANT
Resolution of compressive symptoms (follow-up 12 months)												

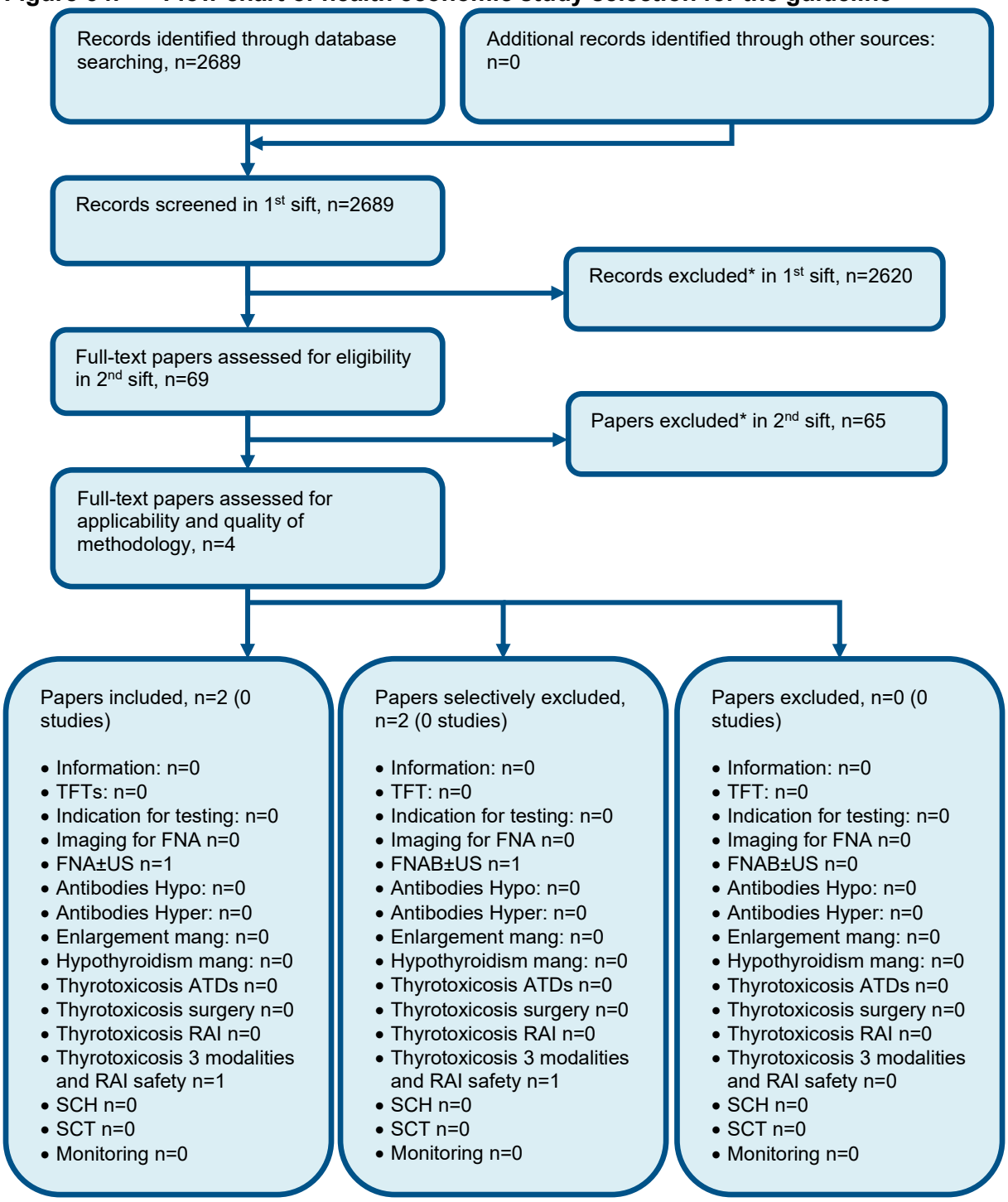
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	10/10 (100%)	100%	RR 1 (0.84 to 1.19)	0 fewer per 1000 (from 160 fewer to 190 more)	⊕⊕OO LOW	IMPORTANT

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

[3] 95% confidence intervals cannot be calculated as no SD available for surgery (all participants with whole nodule removed)

Appendix G: Health economic evidence selection

Figure 34: Flow chart of health economic study selection for the guideline



* Non-relevant population, intervention, comparison, design or setting; non-English language
TFT; thyroid function test, FNA; fine-needle aspiration, US; ultrasound, RAI; radioactive iodine, ATDs; antithyroid drugs, Mang; management, SCH; Subclinical hypothyroidism, SCT; Subclinical thyrotoxicosis.

Appendix H: Health economic evidence tables

None

Appendix I: Health economic analysis

None

Appendix J: Excluded studies

J.1 Excluded clinical studies

Table 26: Studies excluded from the clinical review

Study	Exclusion reason
Baek 2010[2]	NRS without adequate adjustment
Bayani 2012[4]	No usable outcomes
Chen 2016[9]	SR (checked for references)
Cheng 2017[10]	Incorrect interventions
Cheung 1989[11]	No usable outcomes
Chung 2017[12]	Systematic review is not relevant to review question or unclear PICO
Dossing 2007[15]	Inappropriate comparison
Dossing 2013[16]	Incorrect interventions
Faggiano 2012[17]	Not review population
Ferreira 2016[18]	NRS without adequate adjustment
Gambelunghe 2006[19]	Incorrect interventions
Gangappa 2016[20]	Incorrect interventions
Gharib 1998[22]	SR (checked for references)
Giovanella 2018[23]	NRS without adequate adjustment
Ha 2015[25]	SR (checked for references)
Jin 2018[26]	NRS without adequate adjustment
Kim 1997[28]	Not in English
Korkusuz 2018[29]	NRS without adequate adjustment
La rosa 1995[30]	No usable outcomes
Lang 2017[32]	SR (checked for references)
Lang 2017[31]	NRS without adequate adjustment
Lang 2017[33]	NRS without adequate adjustment
Pacella 2017[38]	NRS without adequate adjustment
Sdano 2005[43]	SR (checked for references)
Sewefy 2017[44]	Incorrect interventions
Verde 1994[46]	Incorrect interventions
Wang 2017[47]	SR (checked for references)
Yan 2018[49]	Not minimum duration
Yue 2017[52]	Incorrect interventions
Yue 2017[51]	Correction only
Zingrillo 2000[56]	NRS without adequate adjustment
Zingrillo 2003[55]	Inappropriate comparison

J.2 Excluded health economic studies

None

Appendix K: Research recommendations

K1 Research question: What is the clinical and cost effectiveness of percutaneous thermal ablation for benign thyroid nodules?

Why this is important:

Percutaneous interventional ablation is effective in the management of malignant solid organ disease where the cancer can be destroyed through the application of thermal destructive techniques but there is growing evidence in the literature in application of these to benign thyroid disease.

Benign thyroid nodules and enlargement is common with reported prevalence rates increasing concomitant with increased imaging – many cases being identified incidentally. The majority of cases do not require intervention; however, a subset of patients develop symptoms, usually relating to mass effect or cosmetic embarrassment, requiring volume reduction of the affected gland or portion thereof. This usually requires surgical intervention. Cysts may be aspirated or ablated with ethanol but solid (or predominantly solid) nodules require more invasive treatment – either surgery or percutaneous ablation.

Percutaneous thermal ablation treatments and particularly cryotherapy, microwave ablation and high intensity focused ultrasound are not widely available in the UK and there is currently very little evidence regarding their effectiveness in the management of benign thyroid nodules. Further research on the clinical and cost-effectiveness of those treatments could provide a rationale for the expansion of their availability across the UK, providing less invasive management options than currently available with potential benefits for patients with thyroid enlargement.

If percutaneous thermal ablation offers comparative clinical outcomes to traditional open surgical techniques, there would be potential benefits to a less invasive management strategy to debulk disease and possible opportunities for cost effectiveness, both warranting further investigation.

Criteria for selecting high-priority research recommendations:

PICO question	Population: People with symptomatic benign thyroid disease. Intervention(s): Percutaneous thermal ablation (such as Cryotherapy, Microwave Ablation, Laser Ablation and High Intensity Focused Ultrasound). Comparison: Open surgery Outcome(s): Symptom reduction, incidence and severity of side effects, cost, quality of life.
Importance to patients or the population	The use of less invasive treatments for thyroid nodules causing symptoms is important for patients who do not want to have a scar in the neck or be exposed to potential complications of thyroid surgery such as hypoparathyroidism and recurrent laryngeal nerve damage. In addition, these treatments may have important roles for patients with co-morbidities in whom surgery may be relatively contra-indicated. Furthermore if research supports the cost-effectiveness of these techniques, important savings could be made for the NHS and society as a whole.
Relevance to NICE guidance	Research would address the lack of available evidence for the clinical and cost effectiveness of less invasive therapies for people with benign thyroid nodules This would allow the development of evidence-based guidelines

	to clearly advise for or against the use of percutaneous thermal ablation across the UK either in place of, or alongside currently established surgical techniques.
Relevance to the NHS	Evidence based recommendations would enable clinicians to guide patients to the most clinically and cost-effective, management strategy for benign thyroid nodules, resulting in improved patient outcomes at a low cost/ without increased resource impact.
National priorities	A trial would support a national evidence based approach to the management of benign thyroid nodules with the potential of expanding the use of novel, less invasive treatment strategies not widely available in current practice.
Current evidence base	There is currently very little evidence available regarding the use of non-invasive treatments such as cryotherapy, microwave ablation and high intensity focused ultrasound for people with benign thyroid nodules to support their wide use in the UK's clinical setting.
Equality	The recommendation is not thought to affect equality. The trial would be specific to people with benign thyroid nodules as the examined treatment strategies are considered to be of potential benefit to people with this particular type pf thyroid enlargement.
Study design	Randomised controlled trial with health-economic analyses.
Feasibility	Benign thyroid nodules are common and surgical intervention not infrequent in this patient population. Percutaneous thermal ablation requires specific equipment and training but this is available in the UK. Trials are therefore considered feasible.
Other comments	The trial may attract commercial funders in the thermally ablative techniques as several are available both within each technique and also between the different thermal modalities themselves.
Importance	High: the research is essential to inform future recommendations for the management of patients with benign thyroid nodules.