National Institute for Health and Care Excellence

Draft for Consultation

Thyroid disease: assessment and management

[E] Management of hypothyroidism

NICE guideline

Intervention evidence review underpinning recommendations 1.3.1 to 1.3.7 in the guideline

June 2019

Draft for Consultation

This evidence review was developed by the National Guideline Centre

1

Disclaimer

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

Contents

1

1 Management of hypothyroidism

1.1 Review question: What is the clinical and cost effectiveness of using levothyroxine [L-T4], liothyronine [L-T3], combination of L-T4 and L-T3, thyroid extracts, and iodine and selenium supplementation to treat primary hypothyroidism?

1.2 Introduction

Hypothyroidism occurs when there are insufficient circulating levels of thyroid hormones. It can be subdivided into primary (where the abnormality is with the thyroid gland) or secondary (where the abnormality is in the pituitary gland or hypothalamus). This NICE review will focus on primary hypothyroidism, management and monitoring.

Primary hypothyroidism is common (occurring in about 1-2% of the population, with a much higher incidence in women than men and in the elderly). Symptoms can be non-specific, insidious and often take a while to resolve despite apparent biochemical correction.

Current practice is to diagnose hypothyroidism based on thyroid function tests (usually T4 and TSH) and treat with oral levothyroxine (LT4) in the first instance with the aim of achieving T4 and TSH in the normal range. Once this has been achieved then monitoring with TSH alone is usually appropriate if the patient remains well and on a stable dose. There are currently no national standards for monitoring and normal biochemical ranges vary depending on laboratory assays (as with many other biochemical investigations).

1.3 PICO table

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

Table 1: PICO characteristics of review question

Population	People with primary hypothyroidism
Interventions	T3 T4-initiation at high dose T4-initiation via gradual titration Combination of T3 & T4 Natural thyroid extract (mammalian only) Iodine supplementation Selenium supplementation Placebo
Comparisons	Any above vs any other, in isolation or combination
Outcomes	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Cardiovascular morbidity-ischemic heart disease, heart failure (dichotomous) • Arrhythmias (dichotomous) • Osteoporosis (dichotomous)

	• Impaired cognitive function (dichotomous) • Depression (dichotomous) • Patient/family/carer experience of care (continuous) • Healthcare contacts (rates/dichotomous) • Symptom scores (continuous) • Growth (continuous) • TSH suppression (dichotomous)
Study design	• RCTs only • Blinded comparisons prioritised, non-blinded comparisons only considered if blinded unavailable on an intervention by intervention basis • Minimum treatment duration of 3 months • Crossover studies included

1.4 Clinical evidence

1.4.1 Included studies

Nine RCTs were included in the review; [4, 9, 15, 33, 41, 46, 47, 50, 56] these are summarised in Table 2 below. Evidence from these studies is summarised in the clinical evidence summary below (Table 3).

Seven RCTs compared combined T4 and T3 with T4 alone.[4, 9, 33, 46, 47, 50, 56] One RCT compared natural thyroid extract with T4 alone. [15] One RCT compared a high T4 dose with a titrated T4 dose. [41]

No relevant clinical trials comparing iodine or selenium supplementation with any other intervention or placebo were identified.

All included studies were in the adult (18-65) age stratum. The RCT looking at T4 dose initiation strategies was in a treatment naïve population. All other RCTs were in people previously treated with T4. The primary cause of hypothyroidism varied across studies with autoimmune thyroiditis being the primary cause in six studies.[4, 9, 15, 33, 41, 47] Hypothyroidism was due to radioactive iodine or surgery for Grave's disease in one study [50] and the cause was not specified in the remaining two studies. [46, 47]

The follow-up period of the included studies ranged from 3 to 12 months.

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
Appelhof 2005 [4]	Combined T4+T3, n=93, T4: usual dose minus 25 µg/d; T3:dose required to achieve a 10:1 or a 5:1 T4 to T3 ratio (two separate study arms) T4 only, n=48	Adults (mean 48.38, SD 9.61) Previously receiving stable T4 resulting in TSH (0.11-4 µU/ml) Netherlands	Quality of life Depression TSH suppression (<0.11µU/ml) 15 week treatment	100 % Autoimmune Hypothyroidism Parallel study design
Clyde 2003 [9]	Combined T4+T3, n=23, T4: usual dose minus 50µg/d ;T3: 15 µg/d T4 only, n=23	Adults (mean 45.2, SD 9.7) Previously receiving stable T4 (131 ± 41 µg/d) >3 months, symptom state not reported USA	Quality of life Depression TSH suppression (< 0.20 mIU/L) 4 month treatment	70 % Autoimmune Thyroiditis Parallel study design
Hoang 2013 [15]	Natural thyroid extract, n=78, titrated, initial dose based on conversion of usual T4 (1mg DTE=1.667 µg L-T4) T4 only, n=78	Adults (mean 50.66,SD 23-65) Previously receiving T4 (112.4 ± 36.3 µg/d), symptom state not reported USA	Depression Symptom scores TSH suppression (<0.5 µIU/mL) 4 month treatment	50% Autoimmune hypothyroidism Cross-over study design
Nygaard 2009 [33]	Combined T4 + T3, n=68, T4: usual dose minus 50µg; T3: 20 µg	Adults (intervention: mean 46.5, SD 13.1, control: mean 47.6, SD 12.3)	Quality of Life Depression	85 % Autoimmune hypothyroidism

Study	Intervention and comparison	Population	Outcomes	Comments
	T4 only, n=68	Previously receiving stable T4 (129 ± 29 µg/d) for > 6 months, euthyroid for median 12 (8-34.5) months Denmark	3 month treatment	Cross-over study design
Roos 2005 [41]	High T4 dose, n=25, 1.6µg/kg Titrated T4 dose, n=25, started at 25 µg titrated by 25 µg every 4 weeks until 24 weeks and according to F T4 and TSH levels every 12 weeks onwards.	Adults (mean 47, range 25-86) First diagnosed, previously untreated Netherlands	Quality of life Cardiac events (at 6 months) 12 month treatment	100% untreated primary autoimmune hypothyroidism Parallel study design
Saravanan, 2005 [46]	Combined T4 + T3, n=344, T4: usual dose minus 50 µg/d; T3: 10 µg/d T4 only, n=353	Adults (intervention: mean 57.08, SD 11.31, control: mean 57.60, SD 10.8) Previously receiving stable T4 (127.3 ± 37.4 µg/d) > 3 months and TSH last known within 15 months within reference range United Kingdom	Depression, Symptom scores 3 month treatment	70% Primary hypothyroidism Parallel study design
Sawka, 2003 [47]	Combined T4 + T3, n=20, T4: 50% usual; T3: 25 µg/d (adjusted for normal TSH 0.52-5.0 mU/L) T4 only, n=20	Adults (intervention: mean 45, SD 10.1, control: mean 49.5, SD 11.8) Previously receiving stable T4 (T4 group:120± 38 µg/d; T4+T3 group: 132 ± 46 µg/d) for 6 months. Treated. Canada	Depression Quality of life 15 week treatment	100% Thyroiditis Parallel study design

Study	Intervention and comparison	Population	Outcomes	Comments
Siegmud 2004 [50]	Combined T4 + T3, n=26, T4: usual dose minus 5%; T3: dose required to achieve a 14:1 T4 to T3 ratio T4 only, n=26	Adults (age range 23-69) Previously receiving stable unspecified long-term T4 (100-175 µg/d), symptom state not reported (assume still symptomatic) Germany	Depression TSH suppression (<0.02 mU/l) 3 month treatment	92% surgery or radioactive iodine therapy Cross-over study design
Valizadeh 2014 [56]	Combined T4 + T3, n=36, T4: usual dose minus 50 µg/d; T3: 12.5 µg/d T4 only, n=35	Adults (intervention: mean 39.2, SD 11.2, control: mean 38.8, SD 11.7) Previously receiving T4 for > 6 months resulting in normal TSH (0.3-5.0 mIU/mL) Iran	Depression 4 months treatment	76.6% Autoimmune thyroiditis Parallel study design

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: Combination T4 + T3 versus T4 alone

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4 alone	Risk difference with Combined T4 and T3 (95% CI)
QoL-Disease specific hypo-specific HR-QoL, high is poor outcome. Scale from: 29 to 145.	41 (1 study) 4 months	⨁⊖⊖⊖ VERY LOW1,2 due to indirectness,		The mean QoL-disease specific in the control groups was	The mean QoL-disease specific 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 T4 alone	Risk difference with Combined T4 and T3 (95% CI)
		imprecision		19	4 lower (17.63 lower to 9.63 higher)
QoL-General health SF-36; high is good outcome. Scale from: 0 to 100.	97 (2 studies) 12-15 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to inconsistency, indirectness, imprecision		The mean QoL-general health in the control groups was 67.3	The mean QoL-general health in the intervention groups was 1.36 lower (16.62 lower to 13.90 higher)
QoL-Social functioning SF-36, high is good outcome. Scale from: 0 to 100.	97 (2 studies) 12-15 weeks	⨁⨁⊖⊖ LOW1,2 due to indirectness, imprecision		The mean QoL-social functioning in the control groups was 78.85	The mean QoL-social functioning in the intervention groups was 4.61 higher (0.87 lower to 10.09 higher)
QoL-Mental health SF-36, high is good outcome. Scale from: 0 to 100.	232 (3 studies) 12-15 weeks	⨁⨁⊖⊖ LOW1,2 due to indirectness, imprecision		The mean QoL-mental health in the control groups was 72.9	The mean QoL-mental health in the intervention groups was 1.55 higher (2.14 lower to 5.23 higher)
QoL-Role-emotional SF-36, high is good outcome. Scale from: 0 to 100.	37 (1 study) 15 weeks	⨁⊖⊖⊖ VERY LOW1,2,4 due to risk of bias, indirectness, imprecision		The mean QoL-role-emotional in the control groups was 62.7	The mean QoL-role-emotional in the intervention groups was 8.7 higher (13.34 lower to 30.74 higher)
QoL-Vitality SF-36, high is good outcome. Scale from: 0 to 100.	234 (3 studies) 12-15 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to indirectness, imprecision		The mean QoL-vitality in the control groups was 55.15	The mean QoL-vitality in the intervention groups was 1.44 higher (3.27 lower to 6.16 higher)
QoL-Physical functioning SF-36, high is good outcome. Scale from: 0 to 100.	38 (1 study) 15 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to indirectness, imprecision		The mean QoL-physical functioning in the control groups was 77	The mean QoL-physical functioning in the intervention groups was 2.3 higher (9.74 lower to 14.34 higher)
QoL-Role-physical functioning	37	⨁⊖⊖⊖		The mean QoL-role-physical	The mean QoL-role-physical

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4 alone	Risk difference with Combined T4 and T3 (95% CI)
SF-36, high is good outcome. Scale from: 0 to 100.	(1 study) 15 weeks	VERY LOW1,2,4 due to risk of bias, indirectness, imprecision		functioning in the control groups was 64.1	functioning in the intervention groups was 3.4 lower (26.02 lower to 19.22 higher)
QoL-Bodily pain SF-36, high is good outcome. Scale from: 0 to 100.	37 (1 study) 15 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to indirectness, imprecision		The mean QoL-bodily pain in the control groups was 60.4	The mean QoL-bodily pain in the intervention groups was 2.7 higher (10.85 lower to 16.25 higher)
Depression Cases by HADS/BDI	650 (2 studies) 3-4 months	⊕⊖⊖⊖ VERY LOW1,2,4 due to risk of bias, indirectness, imprecision	RR 0.94 (0.6 to 1.49)	111 per 1000	7 fewer per 1000 (from 44 fewer to 54 more)
Depression BDI, high is poor outcome. Scale from: 0 to 63.	82 (2 studies) 3 months	⊕⊕⊖⊖ LOW1,2 due to indirectness, imprecision		The mean depression in the control groups was 7.3	The mean depression in the intervention groups was 1.77 lower (3.58 lower to 0.03 higher)
Depression (change scores) SCL-90, high is poor outcome. Scale from: 0 to 64.	174 (2 studies) 15 weeks	⊕⊕⊖⊖ LOW1,2 due to indirectness, imprecision		The mean depression (change scores) in the control groups was -6.2	The mean depression (change scores) in the intervention groups was 2.5 higher (0.05 lower to 5.04 higher)
Depression GHQ-28, high is poor outcome. Scale from: 0-21	60 (1 study) 4 months	⊕⊕⊕⊖ MODERATE1 due to indirectness		The mean depression in the control groups was 3.7	The mean depression in the intervention groups was 0.1 lower (1.66 lower to 1.46 higher)
Symptom scores TSQ, high is poor outcome. Scale from: 0 to 36.	697 (1 study) 3 months	⊕⊕⊕⊖ MODERATE1 due to indirectness		The mean symptom scores in the control groups was 11.62	The mean symptom scores in the intervention groups was 0.08 higher (0.5 lower to 0.66 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4 alone	Risk difference with Combined T4 and T3 (95% CI)
cases	(3 studies) 12-16 weeks	MODERATE1 due to indirectness	2.86 (1.54 to 5.32)	87 per 1000	162 more per 1000 (from 47 more to 376 more)

1 Downgraded by 1 or 2 increments because the majority of the evidence included an indirect or very indirect population respectively
2 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
3 Downgraded by 1 or 2 increments because the point estimate and or the confidence intervals varied widely across studies, unexplained by subgroup analysis
4 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 4: Clinical evidence summary: T4 high dose versus T4 titrated dose

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4 titrated dose	Risk difference with T4 high dose (95% CI)
QoL-General health SF-36, high is good outcome. Scale from: 0 to 100.	50 (1 study) 12 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean QoL-general health in the control groups was 50	The mean QoL-general health in the intervention groups was 1 higher (2.71 lower to 4.71 higher)
QoL-Social functioning SF-36, high is good outcome. Scale from: 0 to 100.	50 (1 study) 12 months	⨁⊖⊖⊖ VERY LOW2,3 due to risk of bias, imprecision		The mean QoLsocial functioning in the control groups was 67	The mean QoL-social functioning in the intervention groups was 12 higher (6.1 lower to 30.1 higher)
QoL-Emotional well-being SF-36, high is good outcome. Scale from: 0	50 (1 study) 12 months	⨁⨁⨁⊖ MODERATE3 due to risk of bias		The mean QoL-emotional well-being in the control groups was 50	The mean QoL-emotional well-being in the intervention groups was 1 higher (0.87 lower to 2.87 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with T4 titrated dose	Risk difference with T4 high dose (95% CI)
to 100.					
QoL-Role limits due to emotional well-being SF-36, high is good outcome. Scale from: 0 to 100.	50 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW2,3 due to risk of bias, imprecision		The mean QoL-role limitis due to emotional well-being in the control groups was 62	The mean QoL-role limits due to emotional well-being in the intervention groups was 9 higher (36.51 lower to 54.51 higher)
QoL-Energy SF-36, high is good outcome. Scale from: 0 to 100.	50 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW2,3 due to risk of bias, imprecision		The mean QoL-energy in the control groups was 61	The mean QoL-energy in the intervention groups was 1 lower (6.06 lower to 4.06 higher)
QoL-Physical functioning SF-36, high is good outcome. Scale from: 0 to 100.	50 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW2,3 due to risk of bias, imprecision		The mean QoLphysical functioning in the control groups was 69	The mean QoL-physical functioning in the intervention groups was 3 higher (5.65 lower to 11.65 higher)
QoL- Role limits due to physical functioning SF-36, high is good outcome. Scale from: 0 to 100.	50 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW2,3 due to risk of bias, imprecision		The mean QoL- role limits due to physical functioning in the control groups was 60	The mean QoL- role limits due to physical functioning in the intervention groups was 9 higher (1.11 to 16.89 higher)
QoL-Pain SF-36, high is good outcome. Scale from: 0 to 100.	50 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW2,3 due to risk of bias, imprecision		The mean QoL-pain in the control groups was 64	The mean QoL-pain in the intervention groups was 5 higher (9.42 lower to 19.42 higher)
Cardiac events	50 (1 study) 6 months	⊕⊕⊖⊖ LOW3,4 due to risk of bias, imprecision	Not estimable	0 per 1000	Not estimable[4]

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 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 T4 titrated dose	Risk difference with T4 high dose (95% CI)
at very high risk of bias 4 Zero events in either arm					

Table 5: Clinical evidence summary: Natural thyroid extract versus T4

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 Natural thyroid extract (95% CI)
Depression BDI, high is poor outcome. Scale from: 0 to 63.	70 (1 study) 4 months	⨁⨁⨁⨀ MODERATE1 due to indirectness		The mean depression in the control groups was 4.61	The mean depression in the intervention groups was 0.4 lower (1.99 lower to 1.19 higher)
Symptom scores TSQ, high is poor outcome,. Scale from: 0 to 36.	70 (1 study) 4 months	⨁⨁⨀⨀ LOW1,2 due to indirectness, imprecision		The mean symptom scores in the control groups was 13.16	The mean symptom scores in the intervention groups was 1.4 lower (3.61 lower to 0.81 higher)
TSH suppression (<0.5 µIU/mL) cases	70 (1 study) 4 months	⨁⨁⨁⨀ MODERATE1 due to indirectness	Not estimable		Not estimable[3]
1 Downgraded by 1 or 2 increments because the majority of the evidence included an indirect or very indirect population respectively 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 each arm					

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 Unit costs

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

Table 6: UK costs of hypothyroidism treatment

Drug	Daily dose	Cost - Month	Cost - annual
Levothyroxine (T_4)	100µg (a)	£1.34	£16.03
Liothyronine (T_3)	20µg (b)	£280.48	£3,365.82
Combination T_3 and T_4	Different ratios used 1:10, 1:5, 1:4, 1:3, 1:2 (c)	e.g. ratio 1:3, 50µg of T_4 and 17µg T_3 = £281.82	£3,381.85
Natural thyroid extract (pack size 180 capsules)	6 capsules (d)	£40.56	£486.67

Source: BNF, Date, December 2017[19] (BMJ Group and the Royal Pharmaceutical Society of Great Britain)(BMJ Group and the Royal Pharmaceutical Society of Great Britain)

(a) Maintenance dose 100-200mcg once daily

(b) Initially 10–20 micrograms daily; increased to 60 micrograms daily in 2–3 divided doses (60mcg annual cost = £10,097)

(c) Dose regime depends on the initial levothyroxine dose, varied in the clinical trials, T_3 ranged between 5µg to 20µg

(d) Online prices, amazon, different brands vary in cost, this is the most ordered brand. 6 capsules daily is the maintenance dose

1.6 Evidence statements

1.6.1 Clinical evidence statements

1.6.1.1 Levothyroxine and liothyronine vs levothyroxine alone

No clinically important difference was identified for health-related quality of life (1 study, Very low quality), quality of life- general health (2 studies, Very low quality), quality of life-mental health (3 studies, Low quality), quality of life- vitality (3 studies, Very low quality), quality of life-physical functioning, bodily pain (1 study, very low quality), depression-cases (2 studies, Very low quality), depression-BDI (2 studies, Low quality), depression-SCL-90 (2 studies, Low quality), depression-GHQ-28 (1 study, Moderate quality), symptom scores (1 study, Moderate quality).

There was a clinically important benefit of combined levothyroxine and liothyronine for quality of life-social functioning (2 studies, Low quality) and quality of life-role-emotional (1 study, Very low quality).

There was a clinically important harm of combined levothyroxine and liothyronine for quality of life-role physical functioning (1 study, Very low quality) and TSH suppression (3 studies, Moderate quality).

No evidence was identified for other outcomes.

1.6.1.2 Levothyroxine high dose vs levothyroxine titrated dose

No clinically important difference was identified for quality of life-general health, energy, physical functioning (1 study, Very low quality), quality of life- emotional well-being (1 study, Moderate quality) and cardiac events (1 study, Low quality).

There was a clinically important benefit of levothyroxine at a high dose for quality of life-social functioning, role limits due to emotional well-being and role limits due to physical functioning, pain (1 study, Very low quality).

No evidence was identified for other outcomes.

1.6.1.3 Natural thyroid extract vs levothyroxine

No clinically important difference was identified for depression, TSH suppression (1 study, Moderate quality) and symptom scores (1 study, Low quality).

No evidence was 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

Mortality and quality of life were agreed by the Committee to be the critical outcomes for this review. Important outcomes included cardiovascular morbidity, heart disease, arrhythmias, osteoporosis, impaired cognitive function, depression, experience of care, healthcare contacts, symptom scores, growth and TSH suppression.

1.7.1.2 The quality of the evidence

The most widely reported outcome across studies included in this review was depression. The majority of studies also reported quality of life. A limited number of studies reported symptom scores and cardiac events. TSH suppression was reported occasionally and the defined value below which TSH was suppressed varied across studies.

There was no evidence on mortality or any other outcome.

Overall, the quality of the evidence varied from very low to moderate. The levothyroxine and liothyronine vs levothyroxine alone comparison had the largest number of participants

compared to the other comparisons. Within this comparison, evidence ranged from very low to moderate quality. It was downgraded for indirectness, due to the non-treatment naïve population and imprecision. Evidence was generally also downgraded for risk of bias and occasionally for inconsistency. Natural thyroid extract vs levothyroxine comparison had the highest quality of evidence across comparisons. The evidence quality ranged from low to moderate; it was generally downgraded for indirectness due to the non-treatment naïve population and imprecision in the measurement. The high vs titrated levothyroxine dose comparison had the smallest number of participants and the lowest quality of evidence. The evidence quality ranged from very low to low and it was generally downgraded for risk of bias due to baseline differences and issues with outcome reporting and for imprecision.

1.7.1.3 Benefits and harms

Combined levothyroxine and liothyronine vs levothyroxine alone

There was evidence of a clinically important benefit of combined levothyroxine and liothyronine in terms of two aspects of quality of life, although both outcomes came from short-term follow-up studies. A clinically important harm was associated with the combined use of levothyroxine with liothyronine compared to levothyroxine monotherapy in terms of one aspect of quality of life and TSH suppression. There was no clinically important difference between the two treatments in terms of general health-related quality of life and five different aspects of quality of life. Furthermore, no clinically important difference was seen in either depression or symptom scores. Overall the committee agreed that the evidence was generally suggestive of combined therapy having no important effect on quality of life and the small and contradictory benefits and harms in subdomains of quality of life were more likely to reflect the low quality of the underlying evidence.

The committee noted that some people do not appear to achieve sufficient response to levothyroxine and agreed that it is possible that in this group the addition of liothyronine may have greater benefit than in the general population alone. However, there were no studies exclusively in the population of people who had failed to respond sufficiently to levothyroxine.

The committee were aware that the use of combination therapy is a critical issue in hypothyroidism. Based on the evidence available and the high costs of liothyronine the committee could not recommend its use. However the committee agreed that it is plausible in some people who are not responding to levothyroxine that combination therapy may be beneficial. Without RCT evidence to support this hypothesis, the committee agreed it was not appropriate to recommend the use of liothyronine even in this subpopulation however they made a high priority research recommendation for trials conducted in this subpopulation to allow for firmer guidance in the future.

Levothyroxine high starting dose vs levothyroxine titrated dose

There was a clinically important benefit of high-starting levothyroxine dose compared to titrated in four quality of life domains (social functioning, role limits due to emotional well being, role limits due to physical functioning and pain) but no difference in four different quality of life domains. There was an absence of cardiac events associated with both dosing strategies. This comparison was from a single, relatively small study with outcomes reported at the end of follow-up. The Committee noted that the greatest benefit of the high starting dose is likely to be during the early weeks of intervention, although the study did not report outcomes in this time period.

The Committee agreed that the available evidence was sufficient to make recommendations for starting with a high dose, in the population selected for the trial. The Committee agreed that it may be appropriate to still start with a low titrated dose in people with cardiovascular disease, where there may still be concerns that the higher dose could cause exacerbations of underlying cardiac disease.

Natural thyroid extract vs levothyroxine

There was no clinically important difference across the outcomes of depression and symptom scores for this comparison. No TSH suppression was evident in participants treated with natural thyroid extract or levothyroxine. There was consensus among Committee members that there was insufficient evidence to recommend natural thyroid extract, especially given its status as an unlicensed medication in the UK. The Committee also agreed that, in the absence of clear harm, there was insufficient evidence to make a strong recommendation against the use of natural thyroid extract.

1.7.2 Cost effectiveness and resource use

There was no health economic evidence identified for this question. The committee considered the costs of the different drugs in combination with the clinical evidence to make a judgement regarding likely cost effectiveness.

It was recognised by the committee that levothyroxine (T4) is an inexpensive treatment for hypothyroidism (cost £16 per year for a daily dose of 100µg). The anticipated cost of liothyronine (T3) is £3,365 per year for a daily dose 20µg and for the combination treatment of T3/T4 is £3,381 per year for 50µg T4 and 17µg T3. Given the clinical evidence was inconsistent in terms of whether combination T3/T4 conferred any benefits in terms of quality of life over T4 monotherapy and suggested potential for clinically important harm in terms of TSH suppression. Hence, the committee concluded that T3 should not be routinely offered with or without levothyroxine as it is unlikely to be cost effective compared to T4 monotherapy. The committee acknowledged that the quality of the clinical evidence was poor and felt that a research recommendation would be the most appropriate given the need for good quality evidence that assesses the clinical and cost effectiveness of using T3 alone and T3/T4 combinations opted for a research recommendation.

In conclusion offering levothyroxine as first line is considered to be cost effective and in line with current practice.

The committee agreed that starting levothyroxine (T4) at a high dose is likely to be cost effective, as it has shown benefit over using a titrated dose; given that the individual is unlikely to suffer from any cardiac complications. This will ensure adequate control of symptoms and prompt achievement of treatment targets, leading to gain in quality of life, compared to titrating the dose over a period of time, for a small increase in the same cost.

Natural thyroid extract is also higher cost than T4 monotherapy (£486.67 per year) .The clinical evidence did not show benefit for using natural thyroid extract, which is currently unlicensed in the UK. There was also no data relating to its safety. Given the higher cost and given the lack of evidence to support its clinical efficacy and safety, the committee felt that this intervention is agreed there was no evidence to support it being unlikely to be cost effective.

No clinical evidence was identified for any other intervention in this review, hence; the committee felt agreed that it is not possible to draw any conclusion regarding their clinical and cost effectiveness.

1.7.3 Other factors the Committee took into account

The committee discussed how people and healthcare professionals adjust the dose of levothyroxine in response to thyroid symptoms. The committee agreed that there may be some benefit to some people of changes in levothyroxine dose even when their TSH is in the reference range, as the reference range is based on average population values. However they also noted that the vague nature of thyroid symptoms may make it easy to misattribute other symptoms to thyroid disease which will not respond to levothyroxine dose changes. The committee agreed that this can be a challenging area for healthcare professionals and

people with thyroid disease but that while they were aware that healthcare professionals do alter levothyroxine doses even when TSH is within the reference range, they could not make specific recommendations to titrate more subtly than to the reference range, based on the evidence available.

References

1. Abu-Helalah M, Law MR, Bestwick JP, Monson JP, Wald NJ. A randomized double-blind crossover trial to investigate the efficacy of screening for adult hypothyroidism. Journal of Medical Screening. 2010; 17(4):164-9

2. Akintola AA, Jansen SW, van Bodegom D, van der Grond J, Westendorp RG, de Craen AJ et al. Subclinical hypothyroidism and cognitive function in people over 60 years: a systematic review and meta-analysis. Frontiers in Aging Neuroscience. 2015; 7:150

3. Angermayr L, Clar C. Iodine supplementation for preventing iodine deficiency disorders in children. Cochrane Database of Systematic Reviews 2004, Issue 2. Art. No.: CD003819. DOI: 10.1002/14651858.CD003819.pub2.

4. Appelhof BC, Fliers E, Wekking EM, Schene AH, Huyser J, Tijssen JG et al. Combined therapy with levothyroxine and liothyronine in two ratios, compared with levothyroxine monotherapy in primary hypothyroidism: a double-blind, randomized, controlled clinical trial. Journal of Clinical Endocrinology and Metabolism. 2005; 90(5):2666-74

5. Balázs C. The effect of selenium therapy on autoimmune thyroiditis. Orvosi Hetilap. 2008; 149(26):1227-1232

6. Bunevicius R, Jakuboniene N, Jurkevicius R, Cernicat J, Lasas L, Prange AJ, Jr. Thyroxine vs thyroxine plus triiodothyronine in treatment of hypothyroidism after thyroidectomy for Graves' disease. Endocrine. 2002; 18(2):129-33

7. Carle A, Faber J, Steffensen R, Laurberg P, Nygaard B. Hypothyroid Patients Encoding Combined MCT10 and DIO2 Gene Polymorphisms May Prefer L-T3 + L-T4 Combination Treatment - Data Using a Blind, Randomized, Clinical Study. European Thyroid Journal. 2017; 6(3):143-151

8. Cerbone M, Capalbo D, Wasniewska M, Alfano S, Raso GM, Oliviero U et al. Effects of L-thyroxine treatment on early markers of atherosclerotic disease in children with subclinical hypothyroidism. European Journal of Endocrinology. 2016; 175(1):11-19

9. Clyde PW, Harari AE, Getka EJ, Shakir KM. Combined levothyroxine plus liothyronine compared with levothyroxine alone in primary hypothyroidism: a randomized controlled trial. JAMA. 2003; 290(22):2952-8

10. Cooper DS, Halpern R, Wood LC, Levin AA, Ridgway EC. L-Thyroxine therapy in subclinical hypothyroidism. A double-blind, placebo-controlled trial. Annals of Internal Medicine. 1984; 101(1):18-24

11. Fadeyev VV, Morgunova TB, Melnichenko GA, Dedov, II. Combined therapy with L-thyroxine and L-triiodothyronine compared to L-thyroxine alone in the treatment of primary hypothyroidism. Hormones. 2010; 9(3):245-52

12. Fadeyev VV, Sytch J, Kalashnikov V, Rojtman A, Syrkin A, Melnichenko G. Levothyroxine replacement therapy in patients with subclinical hypothyroidism and coronary artery disease. Endocrine Practice. 2006; 12(1):5-17

13. Fan Y, Xu S, Zhang H, Cao W, Wang K, Chen G et al. Selenium supplementation for autoimmune thyroiditis: A systematic review and meta-analysis. International Journal of Endocrinology. 2014; 2014(904573)

14. Grozinsky-Glasberg S, Fraser A, Nahshoni E, Weizman A, Leibovici L. Thyroxine-triiodothyronine combination therapy versus thyroxine monotherapy for clinical hypothyroidism: meta-analysis of randomized controlled trials. Journal of Clinical Endocrinology and Metabolism. 2006; 91(7):2592-9

15. Hoang TD, Olsen CH, Mai VQ, Clyde PW, Shakir MK. Desiccated thyroid extract compared with levothyroxine in the treatment of hypothyroidism: a randomized, double-blind, crossover study. Journal of Clinical Endocrinology and Metabolism. 2013; 98(5):1982-90

16. Ineck BA, Ng TM. Effects of subclinical hypothyroidism and its treatment on serum lipids. Annals of Pharmacotherapy. 2003; 37(5):725-30

17. Joffe RT, Brimacombe M, Levitt AJ, Stagnaro-Green A. Treatment of clinical hypothyroidism with thyroxine and triiodothyronine: a literature review and metaanalysis. Psychosomatics. 2007; 48(5):379-84

18. Joffe RT, Sawka AM, Marriott MJ, MacQueen GM, Gernstein HC. Does substitution of T4 with T3 plus T4 for T4 replacement improve depressive symptoms in patients with hypothyroidism? Annals of the New York Academy of Sciences. 2004; 1032:287-8

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

20. Kachouei A, Rezvanian H, Amini M, Aminorroaya A, Moradi E. The Effect of Levothyroxine and Selenium versus Levothyroxine Alone on Reducing the Level of Anti-thyroid Peroxidase Antibody in Autoimmune Hypothyroid Patients. Advanced Biomedical Research. 2018; 7:1

21. Kong WM, Sheikh MH, Lumb PJ, Naoumova RP, Freedman DB, Crook M et al. A 6-month randomized trial of thyroxine treatment in women with mild subclinical hypothyroidism. American Journal of Medicine. 2002; 112(5):348-54

22. Kraut E, Farahani P. A Systematic Review of Clinical Practice Guidelines' Recommendations on Levothyroxine Therapy Alone versus Combination Therapy (LT4 plus LT3) for Hypothyroidism. Clinical & Investigative Medicine - Medecine Clinique et Experimentale. 2015; 38(6):E305-13

23. Li X, Meng Z, Jia Q, Ren X. Effect of L-thyroxine treatment versus a placebo on serum lipid levels in patients with sub-clinical hypothyroidism. Biomedical Reports. 2016; 5(4):443-449

24. Ma C, Xie J, Huang X, Wang G, Wang Y, Wang X et al. Thyroxine alone or thyroxine plus triiodothyronine replacement therapy for hypothyroidism. Nuclear Medicine Communications. 2009; 30(8):586-93

25. Mahmoodianfard S, Vafa M, Golgiri F, Khoshniat M, Gohari M, Solati Z et al. Effects of Zinc and Selenium Supplementation on Thyroid Function in Overweight and Obese Hypothyroid Female Patients: A Randomized Double-Blind Controlled Trial. Journal of the American College of Nutrition. 2015; 34(5):391-9

26. Mainenti MR, Vigario PS, Teixeira PF, Maia MD, Oliveira FP, Vaisman M. Effect of levothyroxine replacement on exercise performance in subclinical hypothyroidism. Journal of Endocrinological Investigation. 2009; 32(5):470-3

27. Martins RM, Fonseca RH, Duarte MM, Reuters VS, Ferreira MM, Almeida C et al. Impact of subclinical hypothyroidism treatment in systolic and diastolic cardiac function. Arquivos Brasileiros de Endocrinologia e Metabologia. 2011; 55(7):460-7

28. McDermott MT. Does combination T4 and T3 therapy make sense? Endocrine Practice. 2012; 18(5):750-7

29. Meier C, Staub JJ, Roth CB, Guglielmetti M, Kunz M, Miserez AR et al. TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism: a double blind, placebo-controlled trial (Basel Thyroid Study). Journal of Clinical Endocrinology and Metabolism. 2001; 86(10):4860-6

30. Monzani F, Caraccio N, Kozakowa M, Dardano A, Vittone F, Virdis A et al. Effect of levothyroxine replacement on lipid profile and intima-media thickness in subclinical hypothyroidism: a double-blind, placebo- controlled study. Journal of Clinical Endocrinology and Metabolism. 2004; 89(5):2099-106

31. Monzani F, Di Bello V, Caraccio N, Bertini A, Giorgi D, Giusti C et al. Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: a double blind, placebo-controlled study. Journal of Clinical Endocrinology and Metabolism. 2001; 86(3):1110-5

32. 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

33. Nygaard B, Jensen EW, Kvetny J, Jarlov A, Faber J. Effect of combination therapy with thyroxine (T4) and 3,5,3'-triiodothyronine versus T4 monotherapy in patients with hypothyroidism, a double-blind, randomised cross-over study. European Journal of Endocrinology. 2009; 161(6):895-902

34. Nystrom E, Caidahl K, Fager G, Wikkelso C, Lundberg PA, Lindstedt G. A double-blind cross-over 12-month study of L-thyroxine treatment of women with 'subclinical' hypothyroidism. Clinical Endocrinology. 1988; 29(1):63-75

35. Panicker V, Saravanan P, Vaidya B, Evans J, Hattersley AT, Frayling TM et al. Common variation in the DIO2 gene predicts baseline psychological well-being and response to combination thyroxine plus triiodothyronine therapy in hypothyroid patients. Journal of Clinical Endocrinology and Metabolism. 2009; 94(5):1623-9

36. Parle J, Roberts L, Wilson S, Pattison H, Roalfe A, Haque MS et al. A randomized controlled trial of the effect of thyroxine replacement on cognitive function in community-living elderly subjects with subclinical hypothyroidism: the Birmingham Elderly Thyroid study. Journal of Clinical Endocrinology and Metabolism. 2010; 95(8):3623-32

37. Pinchera A, Santini F. Is combined therapy with levothyroxine and liothyronine effective in patients with primary hypothyroidism? Nature Clinical Practice: Endocrinology & Metabolism. 2005; 1(1):18-19

38. Rayman MP, Thompson AJ, Bekaert B, Catterick J, Galassini R, Hall E et al. Randomized controlled trial of the effect of selenium supplementation on thyroid function in the elderly in the United Kingdom. American Journal of Clinical Nutrition. 2008; 87(2):370-8

39. Reuters VS, Almeida Cde P, Teixeira Pde F, Vigario Pdos S, Ferreira MM, Castro CL et al. Effects of subclinical hypothyroidism treatment on psychiatric symptoms, muscular complaints, and quality of life. Arquivos Brasileiros de Endocrinologia e Metabologia. 2012; 56(2):128-36

40. Rink T, Schroth HJ, Holle LH, Garth H. Effect of iodine and thyroid hormones in the induction and therapy of Hashimoto's thyroiditis. Nuklearmedizin. 1999; 38(5):144-149

41. Roos A, Linn-Rasker SP, van Domburg RT, Tijssen JP, Berghout A. The starting dose of levothyroxine in primary hypothyroidism treatment: a prospective, randomized, double-blind trial. Archives of Internal Medicine. 2005; 165(15):1714-20

42. Ross DS. Bone density is not reduced during the short-term administration of levothyroxine to postmenopausal women with subclinical hypothyroidism: a randomized, prospective study. American Journal of Medicine. 1993; 95(4):385-8

43. Ruggeri RM, Trimarchi F, Biondi B. L-Thyroxine replacement therapy in the frail elderly: A challenge in clinical practice. European Journal of Endocrinology. 2017; 177(4):R199-R217

44. Samuels MH, Kolobova I, Niederhausen M, Janowsky JS, Schuff KG. Effects of Altering Levothyroxine (L-T4) Doses on Quality of Life, Mood, and Cognition in L-T4 Treated Subjects. The Journal of Clinical Endocrinology & Metabolism. 2018; 103(5):1997-2008

45. Samuels MH, Kolobova I, Niederhausen M, Purnell JQ, Schuff KG. Effects of Altering Levothyroxine Dose on Energy Expenditure and Body Composition in Subjects Treated With LT4. Journal of Clinical Endocrinology and Metabolism. 2018; 103(11):4163-4175

46. Saravanan P, Simmons DJ, Greenwood R, Peters TJ, Dayan CM. Partial substitution of thyroxine (T4) with tri-iodothyronine in patients on T4 replacement therapy: results of a large community-based randomized controlled trial. Journal of Clinical Endocrinology and Metabolism. 2005; 90(2):805-12

47. Sawka AM, Gerstein HC, Marriott MJ, MacQueen GM, Joffe RT. Does a combination regimen of thyroxine (T4) and 3,5,3'-triiodothyronine improve depressive symptoms better than T4 alone in patients with hypothyroidism? Results of a double-blind, randomized, controlled trial. Journal of Clinical Endocrinology and Metabolism. 2003; 88(10):4551-5

48. Schmidt U, Nygaard B, Jensen EW, Kvetny J, Jarlov A, Faber J. Peripheral markers of thyroid function: the effect of T4 monotherapy vs T4/T3 combination therapy in hypothyroid subjects in a randomized crossover study. Endocrine Connections. 2013; 2(1):55-60

49. Shatynska-Mytsyk I, Rodrigo L, Cioccocioppo R, Petrovic D, Lakusic N, Compostella L et al. The impact of thyroid hormone replacement therapy on left ventricular diastolic function in patients with subclinical hypothyroidism. Journal of Endocrinological Investigation. 2016; 39(6):709-713

50. Siegmund W, Spieker K, Weike AI, Giessmann T, Modess C, Dabers T et al. Replacement therapy with levothyroxine plus triiodothyronine (bioavailable molar ratio 14 : 1) is not superior to thyroxine alone to improve well-being and cognitive performance in hypothyroidism. Clinical Endocrinology. 2004; 60(6):750-7

51. Smith RN, Taylor SA, Massey JC. Controlled clinical trial of combined triiodothyronine and thyroxine in the treatment of hypothyroidism. British Medical Journal. 1970; 4(5728):145-8

52. Stott DJ, Rodondi N, Kearney PM, Ford I, Westendorp RGJ, Mooijaart SP et al. Thyroid hormone therapy for older adults with subclinical hypothyroidism. New England Journal of Medicine. 2017; 376(26):2534-2544

53. Teixeira PF, Reuters VS, Ferreira MM, Almeida CP, Reis FA, Melo BA et al. Treatment of subclinical hypothyroidism reduces atherogenic lipid levels in a placebo-controlled double-blind clinical trial. Hormone and Metabolic Research. 2008; 40(1):50-5

54. Toulis KA, Anastasilakis AD, Tzellos TG, Goulis DG, Kouvelas D. Selenium supplementation in the treatment of Hashimoto's thyroiditis: a systematic review and a meta-analysis. Thyroid. 2010; 20(10):1163-73

55. Turker O, Kumanlioglu K, Karapolat I, Dogan I. Selenium treatment in autoimmune thyroiditis: 9-month follow-up with variable doses. Journal of Endocrinology. 2006; 190(1):151-6

56. Valizadeh M, Seyyed-Majidi MR, Hajibeigloo H, Momtazi S, Musavinasab N, Hayatbakhsh MR. Efficacy of combined levothyroxine and liothyronine as compared with levothyroxine monotherapy in primary hypothyroidism: a randomized controlled trial. Endocrine Research. 2009; 34(3):80-9

57. van ZEJ, Albusta AY, Fedorowicz Z, Carter B, Pijl H. Selenium supplementation for Hashimoto's thyroiditis. Cochrane Database of Systematic Reviews 2013, Issue 6. Art. No.: CD010223. DOI: 10.1002/14651858.CD010223.pub2.

58. Villar HCCE, Saconato H, Valente O, Atallah ÁN. Thyroid hormone replacement for subclinical hypothyroidism. Cochrane Database of Systematic Reviews 2007, Issue 3. Art. No.: CD003419. DOI: 10.1002/14651858.CD003419.pub2.

59. Walsh JP, Shiels L, Lim EM, Bhagat CI, Ward LC, Stuckey BG et al. Combined thyroxine/liothyronine treatment does not improve well-being, quality of life, or cognitive function compared to thyroxine alone: a randomized controlled trial in patients with primary hypothyroidism. Journal of Clinical Endocrinology and Metabolism. 2003; 88(10):4543-50

60. Wasniewska M, Corrias A, Aversa T, Valenzise M, Mussa A, De Martino L et al. Comparative evaluation of therapy with L-thyroxine versus no treatment in children with idiopathic and mild subclinical hypothyroidism. Hormone research in pdiatrics. 2012; 77(6):376-81

61. Weetman AP. T4 versus T3 and T4: Is it a real controversy? Hormone Research. 2007; 67(SUPPL. 1):128-131

62. Wichman J, Winther KH, Bonnema SJ, Hegedus L. Selenium supplementation significantly reduces thyroid autoantibody levels in patients with chronic autoimmune thyroiditis: A systematic review and meta-analysis. Thyroid. 2016; 26(12):1681-1692

63. Wiersinga WM. L-T4 and L-T3 combined treatment vs L-T4 alone. Annales d Endocrinologie. 2007; 68(4):216-9

64. Wiersinga WM. Therapy of Endocrine disease: T4+T3 combination therapy: is there a true effect? European Journal of Endocrinology. 2017; 177(6):R287-R296

65. Wiersinga WM, Duntas L, Fadeyev V, Nygaard B, Vanderpump MP. 2012 ETA Guidelines: The Use of L-T4 + L-T3 in the Treatment of Hypothyroidism. European Thyroid Journal. 2012; 1(2):55-71

66. Winther KH, Bonnema SJ, Cold F, Debrabant B, Nybo M, Cold S et al. Does selenium supplementation affect thyroid function? Results from a randomized, controlled, double-blinded trial in a Danish population. European Journal of Endocrinology. 2015; 172(6):657-67

67. Winther KH, Wichman JE, Bonnema SJ, Hegedus L. Insufficient documentation for clinical efficacy of selenium supplementation in chronic autoimmune thyroiditis, based on a systematic review and meta-analysis. Endocrine. 2017; 55(2):376-385

68. Yu L, Zhou L, Xu E, Bi Y, Hu X, Pei X et al. Levothyroxine monotherapy versus levothyroxine and selenium combination therapy in chronic lymphocytic thyroiditis. Journal of Endocrinological Investigation. 2017; 40(11):1243-1250

69. Zhao T, Chen B, Zhou Y, Wang X, Zhang Y, Wang H et al. Effect of levothyroxine on the progression of carotid intima-media thickness in subclinical hypothyroidism patients: a meta-analysis. BMJ Open. 2017; 7(10):e016053

Appendices

Appendix A: Review protocols

Table 7:

ID	Field	Content
I	Review question	What is the clinical and cost effectiveness of using levothyroxine [L-T4], liothyronine [L-T3], combination of L-T4 and L-T3, thyroid extracts, and iodine and selenium supplementation to treat primary hypothyroidism?
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 treat hypothyroidism
IV	Eligibility criteria – population / disease / condition / issue / domain	People diagnosed with primary hypothyroidism (TSH greater than upper limit of context specific normal range, T3/T4 below lower limit of context specific normal range)
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	T3 T4 – initiation at high dose T4 – initiation via gradual titration Combination of T3 & T4 Natural thyroid extract (mammalian only) Iodine supplementation Selenium supplementation Placebo
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	Any of above vs any other, in isolation or combination
VII	Outcomes and prioritisation	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Cardiovascular morbidity - ischemic heart disease, heart failure (dichotomous) • Arrhythmias (dichotomous) • Osteoporosis (dichotomous) • Impaired cognitive function (dichotomous) • Depression (dichotomous) • Patient/family/carer experience of care (continuous) • Healthcare contacts (rates/dichotomous) • Symptom scores (continuous) • Growth (continuous) • TSH suppression (dichotomous)

		Minimum duration as for the minimum duration for inclusion of studies unless specified.
VIII	Eligibility criteria – study design	• RCTs only • Blinded comparisons prioritised, non-blinded comparisons only considered if blinded unavailable on an intervention by intervention basis • Minimum treatment duration of 3 months • Crossover studies included
IX	Other inclusion / exclusion criteria	• Including Europe based studies only for selenium supplementation to maintain representative selenium status in trial populations to UK population • Studies in areas/populations of severe iodine deficiency excluded for iodine supplementation • Studies in pregnant women excluded • Studies in people with hypothyroidism post-cancer treatment excluded
X	Proposed sensitivity / subgroup analysis, or meta-regression	**Stratifications** • Age – young children (0-4), children and young people (4-18), adults (>18-65), older adults (>65) • Treatment stage – naïve/general (non-naïve, downgraded for indirectness), second line (remain symptomatic despite previous treatment, as defined by studies) • TSH at initiation of treatment – TSH 2.5-<5 U/ml, 5-<10 U/ml, 10 or more U/ml (only applicable to treatment naïve) • DiO_2 genotype – CC rs225014 vs non-CC **Subgroup analyses** • Age subdivisions (18-50, 50-65, 65-80, >85) • T4 treatment strategy (liquid vs pill, daily vs weekly) • Children on dietary restrictions vs general diet
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 [to add link to history page of the guideline after publication] 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 number	Not registered

1
2

Table 8: 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).[32] **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.

1
2

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. [Add cross reference after publication]

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
Embase (OVID)	1974 – 07 January 2019	Exclusions Randomised controlled trials Systematic review 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.	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.	((iodine or selenium) adj2 supplement*).ti,ab.
28.	(desiccated adj3 (thyroid or hormone* or extract or extracts)).ti,ab.
29.	(thyroid adj2 (extract or extracts)).ti,ab.
30.	(natural adj4 thyroid).ti,ab.
31.	(natural adj3 (extract or extracts)).ti,ab.
32.	armour*.ti,ab.
33.	(thyroxine or levothyroxine or liothyronine or triiodothyronine or tri-iodothyronine).ti,ab.
34.	Thyroxine/ or Triiodothyronine/
35.	(T3 or T4).ti,ab.
36.	(TSH or thyroid stimulating hormone or thyrotropin).ti,ab.
37.	or/27-36
38.	26 and 37
39.	randomized controlled trial.pt.
40.	controlled clinical trial.pt.
41.	randomi#ed.ti,ab.
42.	placebo.ab.
43.	randomly.ti,ab.
44.	Clinical Trials as topic.sh.
45.	trial.ti.
46.	or/39-45
47.	Meta-Analysis/
48.	exp Meta-Analysis as Topic/
49.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
50.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
51.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
52.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
53.	(search* adj4 literature).ab.
54.	(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.
55.	cochrane.jw.
56.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
57.	or/47-56
58.	38 and (46 or 57)

1 **Embase (Ovid) search terms**

1.	exp thyroid disease/
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.	((iodine or selenium) adj2 supplement*).ti,ab.
26.	(desiccated adj3 (thyroid or hormone* or extract or extracts)).ti,ab.
27.	(thyroid adj2 (extract or extracts)).ti,ab.
28.	(natural adj4 thyroid).ti,ab.
29.	armour*.ti,ab.
30.	*thyroxine/ or *levothyroxine/ or *liothyronine/ or *triiodothyronine/
31.	(thyroxine or levothyroxine or liothyronine or triiodothyronine or tri-iodothyronine).ti,ab.
32.	(T3 or T4).ti,ab.
33.	(TSH or thyroid stimulating hormone or thyrotropin).ti,ab.
34.	*thyrotropin/
35.	or/25-34
36.	24 and 35
37.	random*.ti,ab.
38.	factorial*.ti,ab.
39.	(crossover* or cross over*).ti,ab.
40.	((doubl* or singl*) adj blind*).ti,ab.
41.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
42.	crossover procedure/
43.	single blind procedure/

44.	randomized controlled trial/
45.	double blind procedure/
46.	or/37-45
47.	systematic review/
48.	meta-analysis/
49.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
50.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
51.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
52.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
53.	(search* adj4 literature).ab.
54.	(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.
55.	cochrane.jw.
56.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
57.	or/47-56
58.	36 and (46 or 57)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Thyroid Diseases] explode all trees
#2.	hyperthyroid*:ti,ab
#3.	hypothyroid*:ti,ab
#4.	thyrotoxicosis:ti,ab
#5.	(thyroid near/3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)):ti,ab
#6.	(or #1-#5)
#7.	((iodine or selenium) near/2 supplement*):ti,ab
#8.	(desiccated near/3 (thyroid or hormone* or extract or extracts)):ti,ab
#9.	(thyroid near/2 (extract or extracts)):ti,ab
#10.	(natural near/4 thyroid):ti,ab
#11.	(natural near/3 (extract or extracts)):ti,ab
#12.	armour*:ti,ab
#13.	(thyroxine or levothyroxine or liothyronine or triiodothyronine or tri-iodothyronine):ti,ab
#14.	MeSH descriptor: [Thyroxine] explode all trees
#15.	MeSH descriptor: [Triiodothyronine] explode all trees
#16.	(T3 or T4):ti,ab
#17.	(TSH or thyroid stimulating hormone or thyrotropin):ti,ab
#18.	(or #7-#17)
#19.	#6 and #18

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 9: Database date parameters and filters used

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

1 **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)

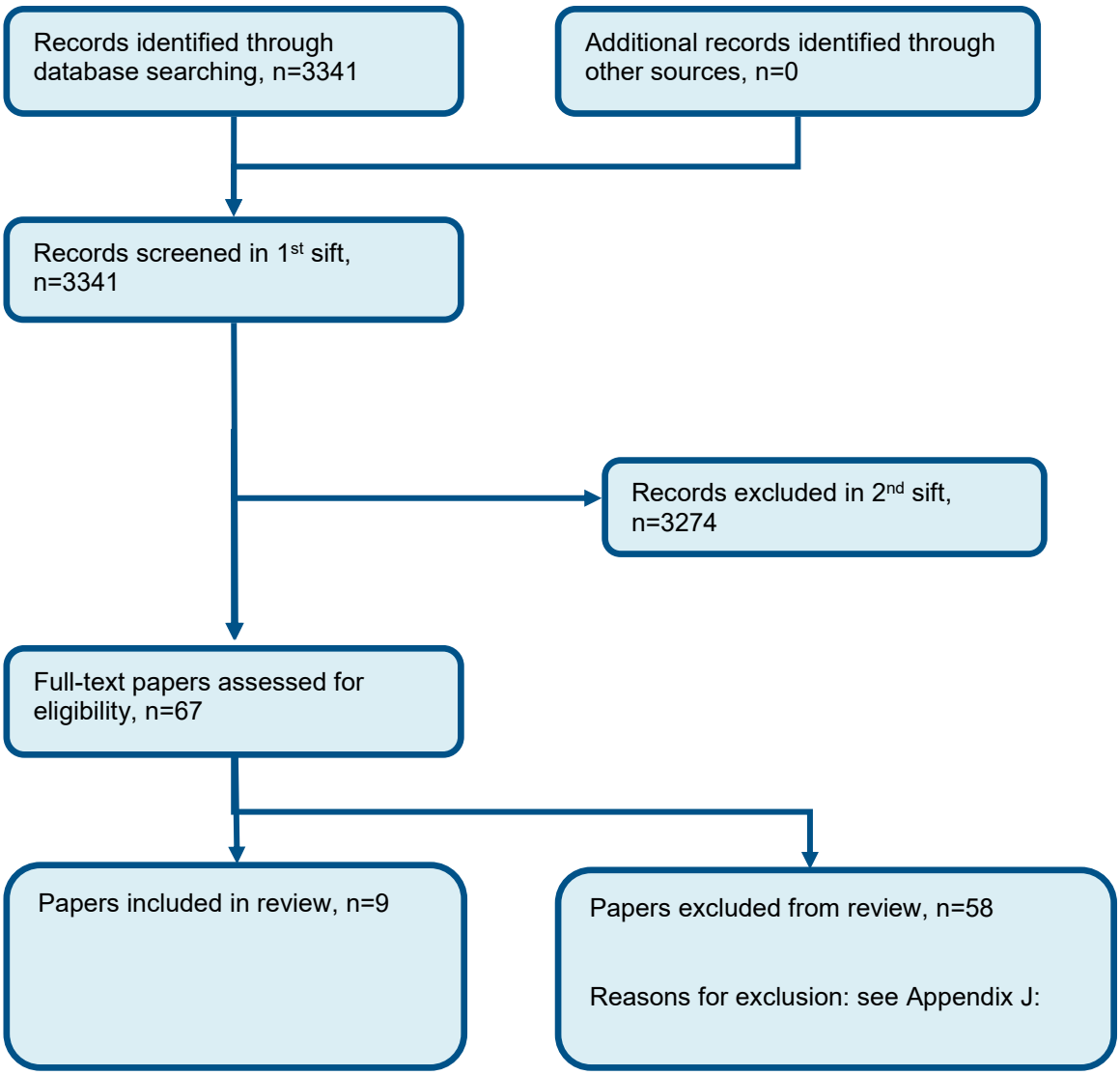
1 **NHS EED and HTA (CRD) search terms**

#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

2

Appendix C: Clinical evidence selection

Figure 1: Flow chart of clinical study selection for the review of management of hypothyroidism



Appendix D: Clinical evidence tables

Study	Appelhof 2005[4]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=141)
Countries and setting	Conducted in Netherlands; Setting: Academic medical centre
Line of therapy	2nd line
Duration of study	Intervention time: 15 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Screening visit
Stratum	Overall
Subgroup analysis within study	Not applicable:
Inclusion criteria	Between 18 and 70 years of age, adequate dose of LT4 replacement therapy for primary autoimmune hypothyroidism for ≥6 months. Adequate dose defined as resulting in serum TSH between 0.11 and 4.0 µU/ml as measured the morning before LT4 intake
Exclusion criteria	history of congenital hypothyroidism, hypethyroidism, thyroidectomy, l-therapy or thyroid cancer; angina pectoris, paroxysmal supraventicular tachycardia, or any serious unstable medical condition; being pregnant or within 6 months postpartum, insufficient understanding of the Dutch language
Recruitment/selection of patients	General practices records
Age, gender and ethnicity	Age - Mean (SD): 48.38 (9.61). Gender (M:F): Define. Ethnicity: Not reported
Further population details	
Extra comments	100% Autoimmune hypothyroidism
Indirectness of population	Serious indirectness: Non-naive to T4 treatment
Interventions	(n=93) Intervention 1: Combined T4 and T3. T4:usual dose minus 25 µg/d; T3:dose required to achieve a 10:1 or a 5:1 T4 to T3 ratio (two separate study arms). Duration 15 weeks. Concurrent medication/care: -. Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive Further details: 1. T4 dosing: 2. T4 formulations: (n=48) Intervention 2: T4 only - T4 - high dose start, usual dose. Duration 15 weeks. Concurrent medication/care: -.

	Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Academic or government funding (Academic Medical Centre Anton Meelmeijer Fund)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINED T4 AND T3 versus T4 - HIGH DOSE START

Protocol outcome 1: Quality of life
- Actual outcome: QoI-Vitality at 15 weeks; Group 1: mean 7.25 (SD 19.59); n=90, Group 2: mean 8.3 (SD 18.5); n=45; Rand-36-Vitality 0-100 Top=High is good outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Other 1 - Low;
Indirectness of outcome: Serious indirectness, Comments: Continuous outcome; Group 1 Number missing: 3, Reason: side effects, illness unrelated to medication, personal time constraints; Group 2 Number missing: 3, Reason: side effects
- Actual outcome: QoI-Mental Health at 15 weeks; Group 1: mean 5.7 (SD 17.12); n=90, Group 2: mean 5.4 (SD 16.1); n=45; RAND-36-Mental health 0-100 Top=High is good outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Other 1 - Low;
Indirectness of outcome: Serious indirectness, Comments: Continuous outcome; Group 1 Number missing: 3, Reason: side effects, illness unrelated to medication, personal time constraints; Group 2 Number missing: 3, Reason: side effects

Protocol outcome 2: Depression
- Actual outcome: Depression at 15 weeks; Group 1: mean -3.6 (SD 7.2); n=90, Group 2: mean -6.2 (SD 8.1); n=45; SCL-90-Depression 0-64 High is poor outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Other 1 - Low;
Indirectness of outcome: Serious indirectness, Comments: Continuous outcome; Group 1 Number missing: 3, Reason: side effects, illness unrelated to medication, personal time constraints; Group 2 Number missing: 3, Reason: side effects

Protocol outcome 3: TSH suppression at end of treatment
-Actual outcome: TSH <0.11 µU/ml at 15 weeks; Group 1: 38/90, Group 2: 7/45
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Other 1 - Low; Indirectness of outcome: No indirectness; Group 1 Number missing: 3, Reason: side effects, illness unrelated to medication, personal time constraints; Group 2 Number missing: 3, Reason: side effects

Protocol outcomes not reported by the study	Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Symptom scores ; Growth

Study	Clyde 2003[9]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=46)
Countries and setting	Conducted in USA; Setting: Military treatment facility
Line of therapy	2nd line
Duration of study	Intervention time: 4 months
Method of assessment of guideline condition	Unclear method of assessment/diagnosis: Not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	between ages 18 and 65, receiving treatment from primary hypothyroidism for at least 6 months, including a stable dose of levothyroxine for at least 3 months
Exclusion criteria	taking suppressive doses of thyroid hormone, pregnancy, cardiac disease or medical problems significantly affecting renal or liver function, taking corticosteroids, amiodarone, carafate, cholestyramine, or more than 325 mg/d of iron
Recruitment/selection of patients	via advertisements
Age, gender and ethnicity	Age - Mean (range): 24-65. Gender (M:F): 8 / 36. Ethnicity: Not stated
Further population details	
Extra comments	Condition caused by 70% Autoimmune thyroiditis
Indirectness of population	Serious indirectness: non-naive to treatment
Interventions	(n=23) Intervention 1: Combined T4 and T3. T4: usual dose minus 50µg/d ;T3 15 µg/d (7.5 µg twice daily). Duration 4 months. Concurrent medication/care: previous history of T4 . Indirectness: Serious indirectness; Indirectness comment: treatment non-naive, 10 patients required dose adjustment at 5 weeks to monitor TSH Further details: 1. T4 dosing: 2. T4 formulations: (n=23) Intervention 2: T4 only - T4 - high dose start. usual dose minus 50µg plus 25µg twice daily. Duration 4 months. Concurrent medication/care: previous history of T4. Indirectness: Serious indirectness; Indirectness comment: treatment non-naive, 8 patients required dose adjustment at 5 weeks to monitor TSH Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Other (Clinical Investigation Program of the National Naval Medical Centre, Bathesda, Md.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINED T4 AND T3 versus T4 - HIGH DOSE START

Protocol outcome 1: Quality of life
- Actual outcome: Hypothyroid Health-related quality of life at After treatment (4 months); Group 1: mean 15 (SD 26); n=21, Group 2: mean 19 (SD 18); n=20; Hypothyroid-specific Health-Related Quality-of-Life 29-145 Top=High is poor outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 2, Reason: 1 Drop-out due to lack of time for testing; Group 2 Number missing: 3, Reason: 1 Drop-out due to tremulousness, fatigue and poor work performance

Protocol outcome 2: Depression
- Actual outcome: Beck Depression Inventory: measuring degree of depressive symptoms (score >10= high) at After treatment (4 months); Group 1: 2/17, Group 2: 2/17
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - High, Measurement - High, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6, Reason: were not given opportunity to complete test; Group 2 Number missing: 6, Reason: were not given the opportunity to complete test

Protocol outcome 3: TSH suppression at end of treatment
-Actual outcome: TSH <0.20 µlU/L at 4 months; Group 1: 2/22, Group 2: 1/22; Comments: Dose adjustments at 5 weeks after review of TSH levels (Group 1: 10/22, Group 2: 8/22)
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1, Reason: adverse symptoms; Group 2 Number missing: 1, Reason: personal time constrains

Protocol outcomes not reported by the study	Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Symptom scores ; Growth

Study	Hoang 2013[15]
Study type	RCT (Patient randomised; Crossover: None reported)
Number of studies (number of participants)	1 (n=78)
Countries and setting	Conducted in USA; Setting: Tertiary care centre
Line of therapy	2nd line
Duration of study	Intervention time: 16 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Stable normal serum TSH verified before testing
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	beneficiaries of the military health care system between ages of 18-65, diagnosed with primary hypothyroidism, on stable L-T4 dose for at least 6 months
Exclusion criteria	pregnancy, coronary artery disease, chronic obstructive lung disease, malabsorption disorder, gastrointestinal surgeries, significant renal or liver dysfunction, seizure disorders, any active cancer, uncontrolled psychosis, psychotropic medications, corticosteroids, amiodarone, iron supplements sucralfate, proton pump inhibitors, cholestyramine
Recruitment/selection of patients	Patients enrolled in the military healthcare system
Age, gender and ethnicity	Age - Mean (range): 50.66 (23-65). Gender (M:F): 17/ 53. Ethnicity: Not reported
Further population details	
Extra comments	50% of patients had autoimmune hypothyroidism.
Indirectness of population	Serious indirectness: Treatment non-naive
Interventions	(n=78) Intervention 1: Combined T4 and T3. Each grain = 38µg L-T4; 9µg T3, Armour thyroid. For initial DTE dose, previous T4 dose was converted to DTE based on: 1mg DTE=1.667 µg L-T4. Titrated at 6 weeks to maintain TSH level 0.5- 3.0 µlU/mL. Duration 16 weeks. Concurrent medication/care: two patients on low-dose β-blocker therapy, potential treatment for hypertension, hyperlipidemia, type 2 diabetes. . Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive, L-T4 for at least 6 months, 2 patients treated with DTE before study Further details: 1. T4 dosing: 2. T4 formulations: (n=78) Intervention 2: T4 only - T4 - high dose start. usual dose. Duration 16 weeks. Concurrent medication/care: LT4 2 patients on low-dose ß-blocker therapy. Potentially treatment for hypertension, hyperlipidemia, type 2 diabetes.

	Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Study funded by industry (Walter Reed National Military Medical Centre Institutional Review Board)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: NATURAL THYROID EXTRACT versus T4 - HIGH DOSE START

Protocol outcome 1: Depression
- Actual outcome: Beck Depression Inventory score at End of each treatment period; Group 1: mean 4.41 (SD 4.71); n=70, Group 2: mean 4.81 (SD 4.89); n=70; Beck Depression Inventory (BDI) 0-63 Top=High is poor outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Other 1 - Low; Indirectness of outcome: Serious indirectness, Comments: Continuous score; Baseline details: Potentially baseline differences in BDI scores; Group 1 Number missing: 8, Reason: pregnancy, time conflicts, relocation; Group 2 Number missing: 8, Reason: pregnancy, time conflicts, relocation

Protocol outcome 2: Symptom scores
- Actual outcome: Thyroid Symptom Questionnaire score at End of each treatment; Group 1: mean 11.76 (SD 6.7); n=70, Group 2: mean 13.16 (SD 6.64); n=70; TSQ-36 0-36 Top=High is poor outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Other 1 - Low; Indirectness of outcome: -- ; Baseline details: Potentially baseline differences in TSQ scores; Group 1 Number missing: 8, Reason: pregnancy, time conflicts, relocation; Group 2 Number missing: 8, Reason: pregnancy, time conflicts, relocation

Protocol outcome 3: TSH suppression
-Actual outcome: TSH < 0.5 µlU/mL at End of treatment; Group 1: 0/70, Group 2: 0/70
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 8, Reason: pregnancy, time conflicts, relocation; Group 2 Number missing: 8, Reason: pregnancy, time conflicts, relocation

Protocol outcomes not reported by the study	Quality of life ; Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Growth

Study	Nygaard 2009[33]
Study type	RCT (Patient randomised; Crossover: No wash out)
Number of studies (number of participants)	1 (n=68)
Countries and setting	Conducted in Denmark; Setting: outpatients, endocrine clinic
Line of therapy	2nd line
Duration of study	Intervention time: 6 months
Method of assessment of guideline condition	Method of assessment /diagnosis not stated: Patients with known overt autoimmune hypothyroidism
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Overt, spontaneous hypothyroidism subjects with serum TSH levels > 20 mU/l, serum T4 < 60 nmol/l, positive TPO antibodies (>60 U/ml) at diagnosis, serum TSH 0.1-5.0 mU/l at screening, unaltered T4 substitution for at least 6 months at screening, 18-76 years
Exclusion criteria	Women pregnant or planning to be pregnant; patients with any other chronic disease, previous T3 treatment, active post partum subacute thyroiditis, hypothyroidism due to surgery or radioactive iodine treatment
Recruitment/selection of patients	from outpatient clinics of three centers, method not reported
Age, gender and ethnicity	Age - Mean (SD): Goup 1: 46.5 (13.1), Group 2: 47.6(12.3). Gender (M:F): 4 /55. Ethnicity: Not stated
Further population details	
Extra comments	Patients with overt autoimmune hypothyroidism .
Indirectness of population	Serious indirectness: non-naive to T4 treatment
Interventions	(n=68) Intervention 1: T4 only - T4 - high dose start. usual dose. Duration 12 weeks. Concurrent medication/care: T4 . Indirectness: No indirectness Further details: 1. T4 dosing: 2. T4 formulations: (n=68) Intervention 2: Combined T4 and T3. usual-50 µg T4 and 20 µg T3. Duration 12 weeks. Concurrent medication/care: usual stable T4 6 months prior treatment . Indirectness: No indirectness Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Other (The Agnes and Knut Mork's Foundation)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINED T4 AND T3 versus T4 - HIGH DOSE START Protocol outcome 1: Quality of life - Actual outcome: SF-36: General health at after each treatment; Group 1: mean 66 (SD 22.28); n=59, Group 2: mean 72 (SD 19.97); n=59; SF-36 0-100 Top=High is good outcome Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness; Group 1 Number missing: 4, Reason: drop-out/excluded patients excluded from analysis; Group 2 Number missing: 5, Reason: drop-out/excluded patients excluded from analysis - Actual outcome: SF-36: Social Functioning at after each treatment; Group 1: mean 85 (SD 19.97); n=59, Group 2: mean 90 (SD 13.83); n=59; SF-36 0-100 Top=High is good outcome Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Subgroups - Low, Other 1 - Low; Indirectness of outcome: -- ; Group 1 Number missing: 4, Reason: excluded or drop-out patients were excluded from analysis; Group 2 Number missing: 5, Reason: excluded or drop-out patients were excluded from analysis - Actual outcome: SF-36: Mental Health at after each treatment; Group 1: mean 76 (SD 15.36); n=59, Group 2: mean 80 (SD 13.06); n=59; SF-36 0-100 Top=High is good outcome Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Subgroups - Low, Other 1 - Low; Indirectness of outcome: -- ; Group 1 Number missing: 4, Reason: excluded or drop-out patients were excluded from analysis; Group 2 Number missing: 5, Reason: excluded or drop-out patients were excluded from analysis - Actual outcome: SF-36: Vitality at after each treatment; Group 1: mean 59 (SD 23.81); n=59, Group 2: mean 65 (SD 20.74); n=59; SF-36 0-100 Top=High is good outcome Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Subgroups - Low, Other 1 - Low; Indirectness of outcome: -- ; Group 1 Number missing: 4, Reason: excluded or drop-out patients were excluded from analysis; Group 2 Number missing: 5, Reason: excluded or drop-out patients were excluded from analysis Protocol outcome 2: Depression - Actual outcome: Beck Depression Inventory (BDI) (score 0-63, 0 best) at after each treatment; Group 1: mean 7.6 (SD 6.14); n=59, Group 2: mean 5.7 (SD 5.38); n=59; BDI 0-63 Top=High is poor outcome Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Subgroups - Low, Other 1 - Low; Indirectness of outcome: Serious indirectness, Comments: Analysis method does not match protocol; Baseline details: Differences in FT4, Anti-TPO, T4 dose between participants may exist; Group 1 Number missing: 4, Reason: excluded or drop-out patients were excluded from analysis; Group 2 Number missing: 5, Reason: excluded or drop-out patients were excluded from analysis	
Protocol outcomes not reported by the study	Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Symptom scores ; Growth

Study	Roos 2005[41]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=50)
Countries and setting	Conducted in Netherlands; Setting: Hospital
Line of therapy	1st line
Duration of study	Intervention + follow up: 48 weeks
Method of assessment of guideline condition	Method of assessment /diagnosis not stated: Clinical score of hypothyroidism was completed on each visit (every 4 weeks during the first 24 weeks of treatments and every 12 weeks thereafter)
Stratum	Naive - TSH >10
Subgroup analysis within study	Not applicable
Inclusion criteria	first diagnosed, untreated primary autoimmune hypothyroidism (serum thyrotropin level>4.2 mIU/L and FT4 level<0.78 ng/dL
Exclusion criteria	history of cardiac disease, taking cardiac medication such as β-blockers
Recruitment/selection of patients	Consecutive patients
Age, gender and ethnicity	Age - Mean (range): 47 (25-86). Gender (M:F): 11/39. Ethnicity: Not stated
Further population details	
Indirectness of population	No indirectness
Interventions	(n=25) Intervention 1: T4 only - T4 - high dose start. 1.6 µg/kg. Duration 48 weeks. Concurrent medication/care: No other medication. Indirectness: No indirectness Further details: 1. T4 dosing: 2. T4 formulations: (n=25) Intervention 2: T4 only - T4 - titrated dose start. Started on 25 µg, titrated every 4 weeks by 25µg until 24 weeks and every 12 weeks from then onwards according to Ft4 and serum thyrotropin levels. Duration 48 weeks. Concurrent medication/care: No other medication. Indirectness: No indirectness Further details: 1. T4 dosing: 2. T4 formulations:
Funding	No funding
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: T4 - HIGH DOSE START versus T4 - TITRATED DOSE START	

Protocol outcome 1: Quality of life
- Actual outcome: Quality of Life- Physical functioning at 48 weeks post start of treatment; Group 1: mean 72 (SD 15.61); n=25, Group 2: mean 69 (SD 15.61); n=25; RAND 36-Item Health Survey Questionnaire-Physical functioning 0-100 Top=High is good outcome
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25; Group 2 Number missing: 25
- Actual outcome: Quality of Life- Role limits due to physical functioning at 48 weeks post start of treatment; Group 1: mean 69 (SD 14.23); n=25, Group 2: mean 60 (SD 14.23); n=25; RAND 36-Item Health Survey-Role limits due to physical functioning 0-100 Top=High is good outcome
Risk of bias: All domain -Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Baseline scores differed between groups; Group 1 Number missing: 25; Group 2 Number missing: 25
- Actual outcome: Quality of life- Social functioning at 48 weeks post start of treatment; Group 1: mean 79 (SD 32.65); n=25, Group 2: mean 67 (SD 32.65); n=25; RAND 36-Item Health Survey- Social functioning 0-100 Top=High is good outcome
Risk of bias: All domain - Very high, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Large difference in baseline scores between groups; Group 1 Number missing: 25; Group 2 Number missing: 25
- Actual outcome: Quality of life-Emotional well-being at 48 weeks post start of treatment; Group 1: mean 51 (SD 3.37); n=25, Group 2: mean 50 (SD 3.37); n=25; RAND 36-Item Health Survey Questionnaire- Emotional well-being 0-100 Top=High is good outcome
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25; Group 2 Number missing: 25
- Actual outcome: Quality of life- Role limits due to emotional well-being at 48 weeks post start of treatment; Group 1: mean 71 (SD 82.09); n=25, Group 2: mean 62 (SD 82.09); n=25; RAND 35-Item Health Survey 0-100 Top=High is good outcome
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Large difference in baseline scores between groups; Group 1 Number missing: 25; Group 2 Number missing: 25
- Actual outcome: Quality of life-Pain at 48 weeks post start of treatment; Group 1: mean 69 (SD 26.01); n=25, Group 2: mean 64 (SD 26.01); n=25; RAND 36-Item Health Survey 0-100 Top=High is good outcome
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25; Group 2 Number missing: 25
- Actual outcome: Quality of life- Energy at 48 weeks post start of treatment; Group 1: mean 60 (SD 9.12); n=25, Group 2: mean 61 (SD 9.12); n=25; RAND 36-Item Health Survey 0-100 Top=High is good outcome
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25; Group 2 Number missing: 25
- Actual outcome: Quality of life-General Health at 48 weeks post start of treatment; Group 1: mean 51 (SD 6.7); n=25, Group 2: mean 50 (SD 6.7); n=25
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25; Group 2 Number missing: 25

Protocol outcome 2: Ischemic heart disease

- Actual outcome: Cardiac events at 24 weeks post start of treatment; Group 1: 0/25, Group 2: 0/25 Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25; Group 2 Number missing: 25	
Protocol outcomes not reported by the study	Mortality ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Depression ; Experience of care ; Healthcare contacts ; Symptom scores ; Growth

Study	Saravanan 2005[46]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=697)
Countries and setting	Conducted in United Kingdom
Line of therapy	Adjunctive to current care
Duration of study	Intervention + follow up: 3 month treatment + 12 month follow up
Method of assessment of guideline condition	Method of assessment /diagnosis not stated: 70% primary hypothyroidism
Stratum	Overall: -
Subgroup analysis within study	Not stratified but pre-specified: Baseline T3, T4, TSH
Inclusion criteria	Age 18-75; T4 dose >100mh/d; TSH level recorded in the last 15 months and known to be within the local laboratory reference range; no T4 dose adjustment in the last 3 months.
Exclusion criteria	History of myocardial infraction, unstable angina or heart failure in the past 3 months; thyroid cancer or secondary hypothyroidism, cholestyramine use, use of antidepressants in the previous 3 months or amiodarone in the previous 12 months.
Recruitment/selection of patients	Patients from 28 family practices
Age, gender and ethnicity	Age - Mean (SD): Intervention: 57.08 (11.31), Control: 57.60 (10.8). Gender (M:F): 16:84. Ethnicity:
Further population details	
Indirectness of population	Serious indirectness: TSH within local laboratory reference range
Interventions	(n=344) Intervention 1: Combined T4 and T3. T4 usual dose minus 50 mg/d; T3: 10 mg/d. Duration 3 months. Concurrent medication/care: -. Indirectness: No indirectness Further details: 1. T4 dosing: Daily (-). 2. T4 formulations: Pill (-). (n=353) Intervention 2: T4 only - T4 - high dose start. usual dose. Duration 3 months. Concurrent medication/care: -. Indirectness: No indirectness Further details: 1. T4 dosing: Daily 2. T4 formulations: Pill
Funding	Study funded by industry (South West NHS R&D Goldshield Pharmaceuticals PLC.)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINED T4 AND T3 versus T4 - HIGH DOSE START	

Protocol outcome 1: Depression - Actual outcome: HADS at 3 months; Group 1: 30/308, Group 2: 32/308; Comments: Numbers at risk were estimated form available data Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: -- ; Group 1 Number missing: 36, Reason: Participants declined to continue with medication; Group 2 Number missing: 45, Reason: Participants declined to continue with medication Protocol outcome 2: Symptom scores - Actual outcome: TSQ at 3 months; MD; 0.08 (95%CI -0.5 to 0.65) 0-36 Top=High is poor outcome, Comments: Comparison between groups at 3 months; Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: -- ; Group 1 Number missing: 36, Reason: Participants declined to continue with medication; Group 2 Number missing: 45, Reason: Participants declined to continue with medication	
Protocol outcomes not reported by the study	Quality of life ; Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Growth

Study	Sawka 2003[47]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=40)
Countries and setting	Conducted in Canada; Setting: McMaster University Medical Centre laboratory
Line of therapy	2nd line
Duration of study	Intervention time: 15 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: TSH concentrations, free T4 and T3 measured at screening and randomization
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	women and men aged 25 to 75 years with an established diagnosis of primary hypothyroidism, use of stable unchanged dose of levothyroxine for 6 months before randomization, baseline TSH concentration within normal limits, evidence of depressive symptoms as defined by a score of more than 5 on the 30-item General Health Questionnaire on 2 occasions, at least 2 weeks apart.
Exclusion criteria	a history of hyperthyroidism, thyroidectomy, or thyroid cancer; a diagnosis of mood disorder predating the hypothyroidism; taking concurrent medication that may affect mental state (including psychotropic medications, β-blockers, systemic glucorticoids, or lithium); concurrent medical illness that may affect mental state or that required active treatment (including type 1 diabetes mellitus or insulin-requiring type 2 diabetes mellitus); inability to complete questionnaires or fertile women not using reliable birth control methods.
Recruitment/selection of patients	outpatients and public advertisements
Age, gender and ethnicity	Age - Mean (SD): Intervention: 45.0 (10.1); Control: 49.5 (11.8). Gender (M:F): 4/36. Ethnicity: Not stated
Further population details	
Extra comments	100% thyroiditis
Indirectness of population	Serious indirectness: Treatment non-naive
Interventions	(n=20) Intervention 1: Combined T4 and T3. T4: 50% usual dose; T3: 25 µg/d (adjusted to keep goal TSH within normal range: 0.52 - 5.0 mU/L). Duration 15 weeks. Concurrent medication/care: stable L-T4 for minimum six months prior study. Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive Further details: 1. T4 dosing: 2. T4 formulations:

	(n=20) Intervention 2: T4 only - T4 - high dose start. T4: usual dose and placebo. Duration 15 weeks. Concurrent medication/care: stable L-T4 for minimum six months prior study. Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINED T4 AND T3 versus T4 - HIGH DOSE START

Protocol outcome 1: Quality of life
- Actual outcome: Quality of life-Physical functioning at End of treatment; Group 1: mean 79.3 (SD 14.9); n=20, Group 2: mean 77 (SD 21.9); n=18; The Medical Outcomes Study (MOS) health status questionnaire- Physical functioning 0-100 Top=High is good outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness; Group 1 Number missing: 0; Group 2 Number missing: 2, Reason: Did not undergo measurement
- Actual outcome: Quality of life- Role-physical at End of treatment; Group 1: mean 60.7 (SD 35.1); n=20, Group 2: mean 64.1 (SD 34.9); n=17; MOS-Role-physical 0-100 Top=High is good outcome
Risk of bias: All domain - High, Selection – Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness; Baseline details: Large baseline difference favoring T4 group; Group 1 Number missing: 0; Group 2 Number missing: 3, Reason: did not undergo measurement, side effects, no explanation
- Actual outcome: Quality of life-Bodily pain at End of treatment; Group 1: mean 63.1 (SD 21.8); n=20, Group 2: mean 60.4 (SD 20.2); n=17; MOS-Bodily pain 0-100 Top=High is good outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness; Group 1 Number missing: 0; Group 2 Number missing: 3, Reason: did not undergo measurement, side effects, no explanation
- Actual outcome: Quality of life-General Health at End of treatment; Group 1: mean 59 (SD 15.4); n=20, Group 2: mean 68.6 (SD 17.5); n=18; Mos-General Health 0-100 Top=High is good outcome
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Baseline difference in scores favoring T4 group; Group 1 Number missing: 0; Group 2 Number missing: 2, Reason: did not undergo measurement
- Actual outcome: Quality of life-Vitality at End of treatment; Group 1: mean 50.7 (SD 14.4); n=20, Group 2: mean 51.3 (SD 21.9); n=18; MOS-Vitality 0-100 Top=High is good outcome
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 2, Reason: did not undergo measurement
- Actual outcome: Quality of life-Social functioning at End of treatment; Group 1: mean 75.9 (SD 14.3); n=20, Group 2: mean 72.7 (SD 21.5); n=18; MOS- Social functioning 0-100 Top=High is good outcome
Risk of bias: All domain - High, Selection – Very high , Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low,

Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Large baseline difference favoring T4 group; Group 1 Number missing: 0; Group 2 Number missing: 2, Reason: Did not undergo measurement - Actual outcome: Quality of life-Role-emotional at End of treatment; Group 1: mean 71.4 (SD 30.3); n=20, Group 2: mean 62.7 (SD 37); n=17; MOS- Role-emotional 0-100 Top=High is good outcome Risk of bias: All domain - High, Selection – Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Large baseline difference favoring T4 group; Group 1 Number missing: 0; Group 2 Number missing: 3, Reason: did not undergo measurement, side effects, no explanation - Actual outcome: Quality of life-Mental Health at End of treatment; Group 1: mean 63.3 (SD 16.6); n=20, Group 2: mean 69.8 (SD 20.4); n=18; MOS-Mental health 0-100 Top=High is good outcome Risk of bias: All domain - High, Selection –Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Baseline difference in scores favoring T4 group; Group 1 Number missing: 0; Group 2 Number missing: 2, Reason: Did not undergo measurement Protocol outcome 2: Depression - Actual outcome: SCL-90, Depressive symptoms at End of treatment; Group 1: mean 0.69 (SD 0.64); n=20, Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: Serious indirectness, Comments: Continuous score; Group 1 Number missing: 0; Group 2 Number missing: 1, Reason: did not undergo measurement for that outcome	
Protocol outcomes not reported by the study	Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Symptom scores ; Growth

Study	Siegmund 2004[50]
Study type	RCT (Patient randomised; Crossover: No washout)
Number of studies (number of participants)	1 (n=23)
Countries and setting	Conducted in Germany; Setting: secondary care
Line of therapy	2nd line
Duration of study	Intervention time: 3 months
Method of assessment of guideline condition	Method of assessment /diagnosis not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	hypothyroidism, stable long-term T4 replacement therapy
Exclusion criteria	hepatitis B, HIV positive, consuming more than 40 g of alcohol per day
Recruitment/selection of patients	outpatients
Age, gender and ethnicity	Age - Range: 23-69. Gender (M:F): 5/21. Ethnicity: Not specified
Further population details	
Extra comments	92% surgery or radioactive iodine therapy. Inclusion/exclusion criteria not specified
Indirectness of population	Serious indirectness: Treatment non-naive
Interventions	(n=26) Intervention 1: Combined T4 and T3. T4: usual dose-5%; T3: dose required to achieve a 14:1 T4 to T3 ratio. Duration 12 weeks. Concurrent medication/care: 11 subjects were on β-adrenoreceptor blocking drugs, ACE inhibitors and diuretics. Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive Further details: 1. T4 dosing: 2. T4 formulations: (n=26) Intervention 2: T4 only - T4 - high dose start. usual dose. Duration 12 weeks. Concurrent medication/care: 11 subjects were on β-adrenoreceptor blocking drugs, ACE inhibitors and diuretics. Indirectness: Serious indirectness; Indirectness comment: Treatment non-naive Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Other author(s) funded by industry (Henning-Berlin (Medical equipment and devices/ Health care supplies))
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINED T4 AND T3 versus T4 - HIGH DOSE START	

Protocol outcome 1: Depression - Actual outcome: Mood states-severity of depressive symptoms at 3 months post treatment; Group 1: mean 5.5 (SD 5.7); n=23, Group 2: mean 6.9 (SD 6.7); n=23; Beck Depression Inventory (BDI) 0-63 Top=High is poor outcome Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Other 1 - Low; Indirectness of outcome: Serious indirectness, Comments: Continuous outcome; Baseline details: Unknown comparability of baseline mood state; Group 1 Number missing: 3, Reason: withdrawn for personal reasons, surgical treatment for a disk prolapse, atrial fabrilation with absolute arrhythmia in association with TSH suppression below zero after treatment; Group 2 Number missing: 3, Reason: withdrawn for personal reasons, surgical treatment for a disk prolapse, atrial fabrilation with absolute arrhythmia in association with TSH suppression below zero after treatment Protocol outcome 2: TSH suppression at end of treatment -Actual outcome: TSH <0.02 µU/l at 3 months; Group 1: 8/23, Group 2: 2/23 Risk of bias: All domain - Low, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - High, Other 1 - Low; Indirectness of outcome: No indirectness; Baseline details: Unknown comparability of baseline mood state; Group 1 Number missing: 3, Reason: withdrawn for personal reasons, surgical treatment for a disk prolapse, atrial fabrilation with absolute arrhythmia in association with TSH suppression below zero after treatment; Group 2 Number missing: 3, Reason: withdrawn for personal reasons, surgical treatment for a disk prolapse, atrial fabrilation with absolute arrhythmia in association with TSH suppression below zero after treatment	
Protocol outcomes not reported by the study	Quality of life ; Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Symptom scores ; Growth

Study	Valizadeh 2009[56]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=71)
Countries and setting	Conducted in Iran; Setting: Outpatients
Line of therapy	2nd line
Duration of study	Intervention time: 4 months
Method of assessment of guideline condition	Method of assessment /diagnosis not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age between 18 and 60 years, on adequate dose of LT4 (resulting in normal level TSH 0.3-5.0 mIU/mL) for primary hypothyroidism for at least 6 months preceding recruitment including a stable dose for at least 3 months.
Exclusion criteria	Taking suppressive doses of thyroxine, antiobesity chemicals, amiodarone, corticosteroids, ferrous sulfate or psychiatric pharmaceuticals; cardiac diseases or medical problems that would significantly affect renal or liver function; psychiatric disorders; pregnancy
Recruitment/selection of patients	not specified
Age, gender and ethnicity	Age - Mean (SD): Intervention: 39.2(11.2); Control: 38.8(11.7). Gender (M:F): 12/48. Ethnicity: Iranian
Further population details	
Extra comments	76.6% Autoimmune thyroiditis
Indirectness of population	Serious indirectness: Treatment non-naive
Interventions	(n=36) Intervention 1: Combined T4 and T3. T4: usual dose-50µg; T3: 12.5µg/d. Duration 4 months. Concurrent medication/care: T4 for at least 6 months prior study. Indirectness: Serious indirectness; Indirectness comment: non-naive to T4 treatment Further details: 1. T4 dosing: 2. T4 formulations: (n=35) Intervention 2: T4 only - T4 - high dose start. usual dose-50µg + 50µg/d in study capsule; adjusted for normal TSH. Duration 4 months. Concurrent medication/care: T4 for at least 6 months prior study. Indirectness: Serious indirectness; Indirectness comment: non-naive to T4 treatment Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINED T4 AND T3 versus T4 - HIGH DOSE START Protocol outcome 1: Depression - Actual outcome: Psychological state: Depression at baseline and 4 months after treatment; Group 1: mean -0.5 (SD 2.1); n=30, Group 2: mean 0 (SD 2.1); n=30; GHQ-28-depression subscale 0-21 Top=High is poor outcome Risk of bias: All domain - Low, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Other 1 - Low; Indirectness of outcome: Serious indirectness, Comments: continuous outcome; Group 1 Number missing: 6, Reason: withdrawal due to pregnancy, palpitation, digestive problems; Group 2 Number missing: 5, Reason: withdrawal due to digestive problems	
Protocol outcomes not reported by the study	Quality of life ; Mortality ; Ischemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Experience of care ; Healthcare contacts ; Symptom scores ; Growth

Appendix E: Forest plots

E.1 Primary hypothyroidism - combined T4 + T3 vs T4 only

Figure 1: Quality of life (hypothyroidism QoL, 29-45, high is poor outcome, 4 months)

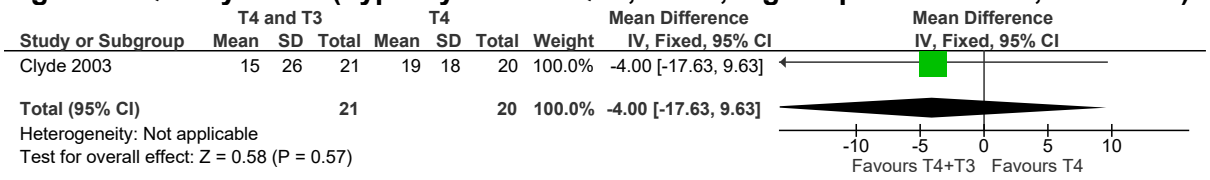


Figure 2: Quality of life: general health (SF-36, 0-100, high is good outcome, 12-15 weeks)

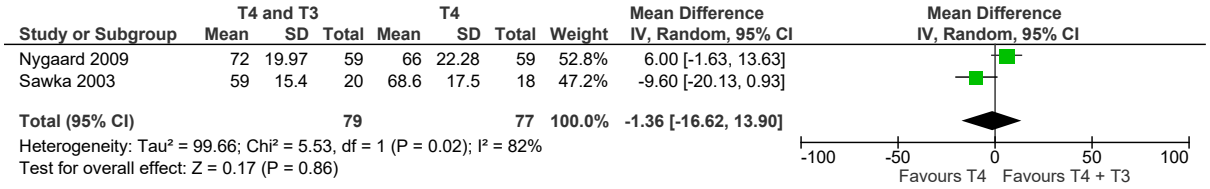


Figure 3: Quality of life: social functioning (SF-36, 0-100, high is good outcome, 12-15 weeks)

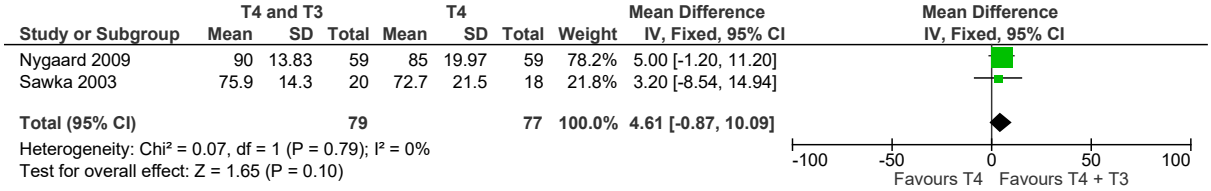


Figure 4: Quality of life: mental health (SF-36, 0-100, high is good outcome, 12-15 weeks)

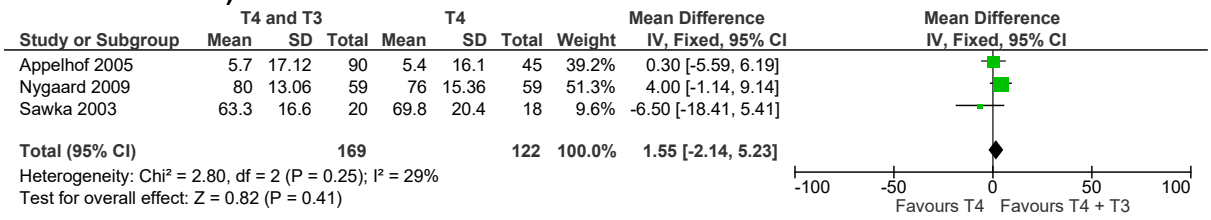
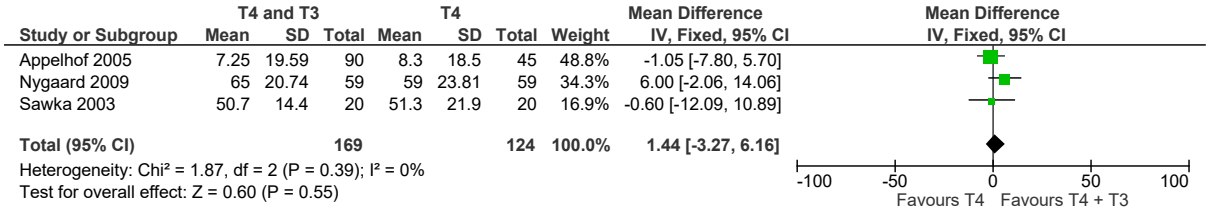
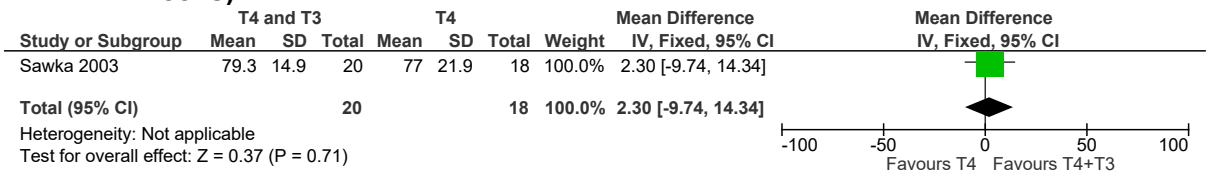


Figure 5: Quality of life: vitality (SF-36, 0-100, high is good outcome, 12-15 weeks)



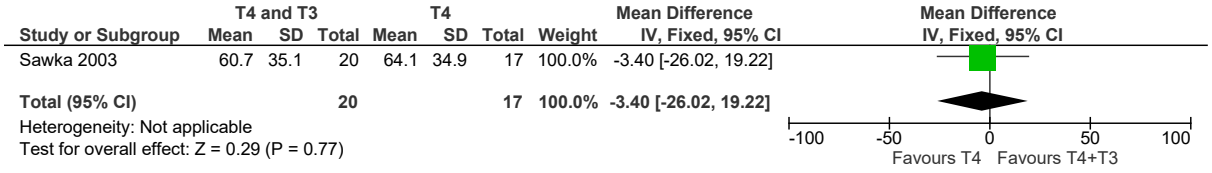
1

Figure 6: Quality of life: physical functioning (SF-36, 0-100, high is good outcome, 15 weeks)



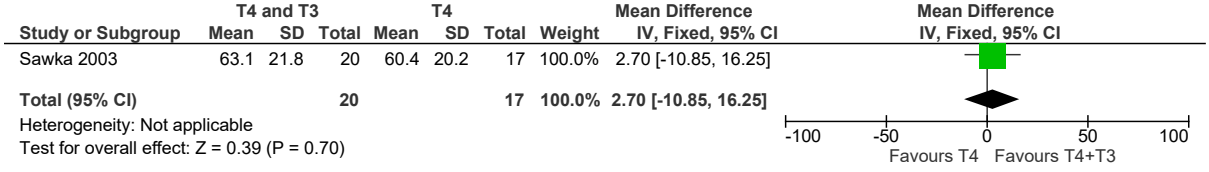
2

Figure 7: Quality of life: role limits due to physical functioning (SF-36, 0-100, high is good outcome, 15 weeks)



3

Figure 8: Quality of life: bodily pain (SF-36, 0-100, high is good outcome, 15 weeks)



4

Figure 9: Quality of life: role limits due to emotional problems (SF-36, 0-100, high is good outcome, 15 weeks)

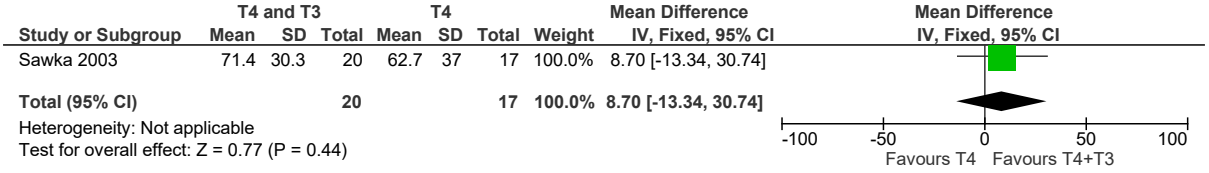


Figure 10: Depression (cases by HADS/BDI, 3-4 months)

Study or Subgroup	T4 and T3 Events	T4 and T3 Total	T4 Events	T4 Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Clyde 2003	2	17	2	17	5.9%	1.00 [0.16, 6.30]
Saravanan 2005	30	308	32	308	94.1%	0.94 [0.58, 1.50]
Total (95% CI)		**325**		**325**	**100.0%**	**0.94 [0.60, 1.49]**
Total events	32		34			

Heterogeneity: Chi² = 0.00, df = 1 (P = 0.95); I² = 0%
Test for overall effect: Z = 0.26 (P = 0.80)

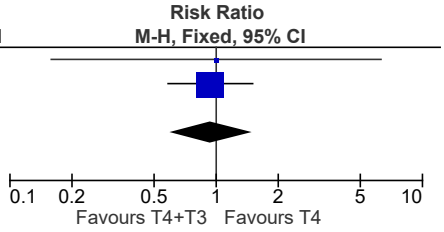


Figure 11: Depression (BDI, 0-63, high is poor outcome, 3months)

Study or Subgroup	T4 and T3 Mean	T4 and T3 SD	T4 and T3 Total	T4 Mean	T4 SD	T4 Total	Weight	Mean Difference IV, Fixed, 95% CI
Nygaard 2009	5.7	5.38	59	7.6	6.14	59	74.9%	-1.90 [-3.98, 0.18]
Siegmund 2004	5.5	5.7	23	6.9	6.7	23	25.1%	-1.40 [-4.99, 2.19]
Total (95% CI)			**82**			**82**	**100.0%**	**-1.77 [-3.58, 0.03]**

Heterogeneity: Chi² = 0.06, df = 1 (P = 0.81); I² = 0%
Test for overall effect: Z = 1.93 (P = 0.05)

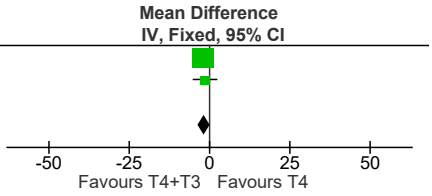


Figure 12: Depression- change score (SCL-90 depression, 0-64, high is poor outcome, 15 weeks)

Study or Subgroup	T4+T3 Mean	T4+T3 SD	T4+T3 Total	T4 Mean	T4 SD	T4 Total	Weight	Mean Difference IV, Fixed, 95% CI
Appelhof 2005	-3.6	7.2	90	-6.2	8.1	45	82.7%	2.60 [-0.20, 5.40]
Sawka 2003	11	10.2	20	9	9.3	19	17.3%	2.00 [-4.12, 8.12]
Total (95% CI)			**110**			**64**	**100.0%**	**2.50 [-0.05, 5.04]**

Heterogeneity: Chi² = 0.03, df = 1 (P = 0.86); I² = 0%
Test for overall effect: Z = 1.92 (P = 0.05)

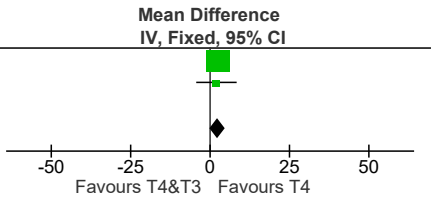


Figure 13: Depression (GHQ-28,high is poor outcome, 4 months)

Study or Subgroup	T4 and T3 Mean	T4 and T3 SD	T4 and T3 Total	T4 Mean	T4 SD	T4 Total	Weight	Mean Difference IV, Fixed, 95% CI
Valizadeh 2009	3.6	2.6	30	3.7	3.5	30	100.0%	-0.10 [-1.66, 1.46]
Total (95% CI)			**30**			**30**	**100.0%**	**-0.10 [-1.66, 1.46]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.13 (P = 0.90)

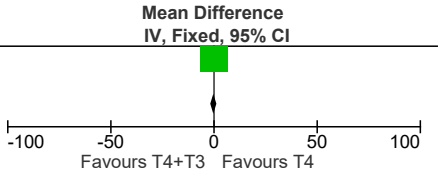


Figure 14: Symptom scores (TSQ, 0-36, high is poor outcome, 3 months)

Study or Subgroup	Mean Difference	SE	T4 and T3 Total	T4 Total	Weight	Mean Difference IV, Fixed, 95% CI
Saravanan 2005	0.08	0.2959	344	353	100.0%	0.08 [-0.50, 0.66]
Total (95% CI)			**344**	**353**	**100.0%**	**0.08 [-0.50, 0.66]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.27 (P = 0.79)

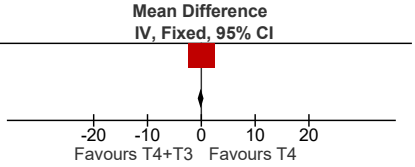
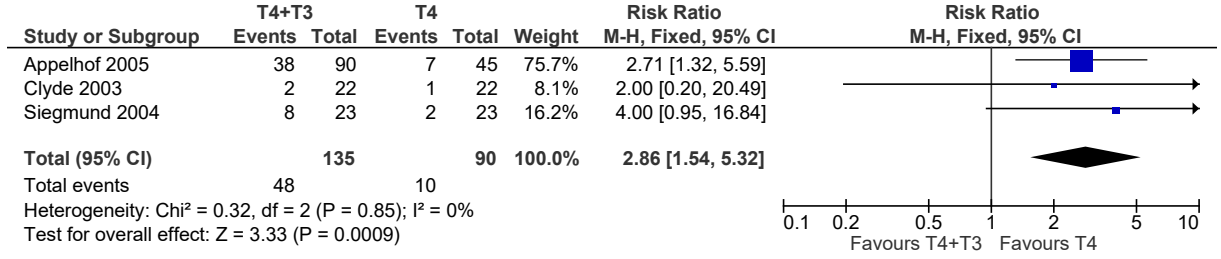


Figure 15: TSH suppression below normal (cases at 12-16 weeks)



E.2 Primary hypothyroidism - T4 high dose vs T4 titrated dose

Figure 16: Quality of life: general health (SF-36, 0-100, high is good outcome, 12 months)

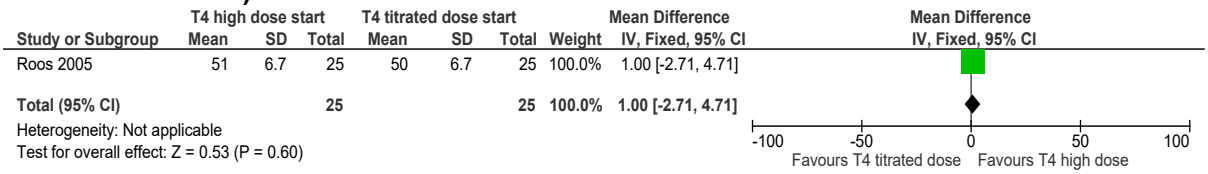


Figure 17: Quality of life: social functioning (SF-36, 0-100, high is good outcome, 12 months)

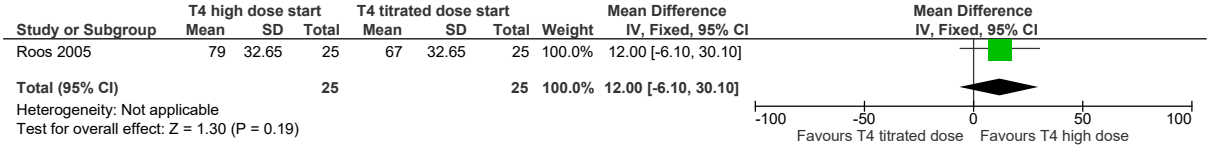


Figure 18: Quality of life: emotional well-being (SF-36, 0-100, high is good outcome, 12 months)

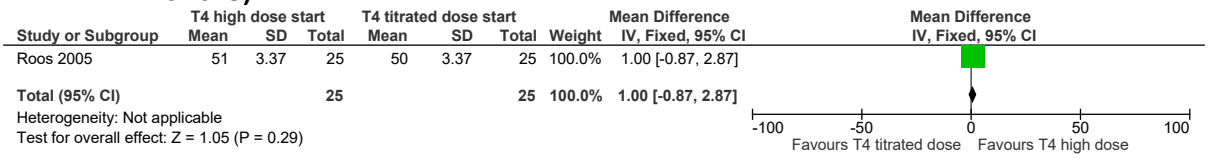


Figure 19: Quality of life: role limits due to emotional well-being (SF-36, 0-100, high is good outcome, 12 months)

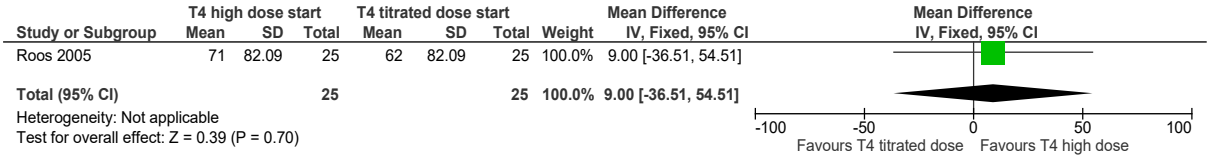


Figure 20: Quality of life: energy (SF-36, 0-100, high is good outcome, 12 months)

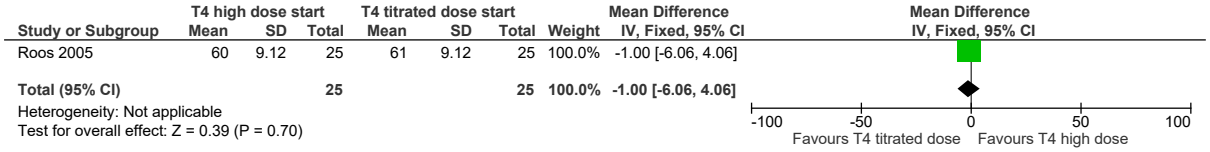


Figure 21: Quality of life: physical functioning (SF-36, 0-100, high is good outcome, 12 months)

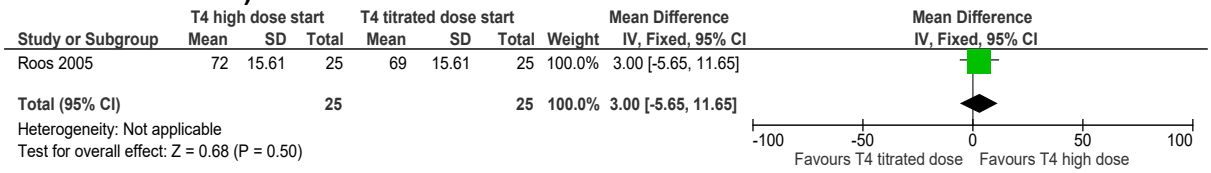


Figure 22: Quality of life: role limits due to physical functioning (SF-36, 0-100, high is good outcome, 12 months)

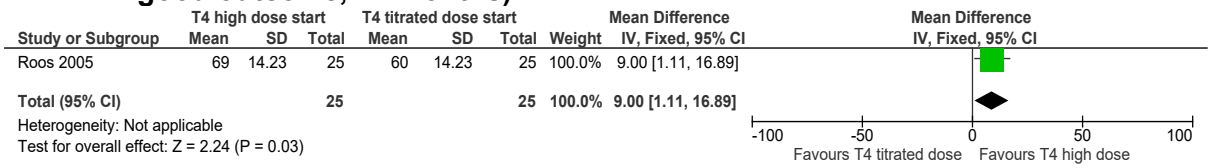
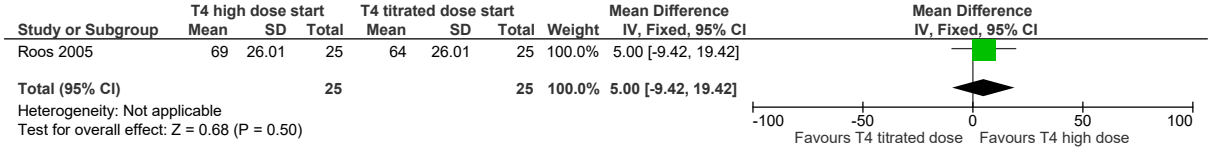
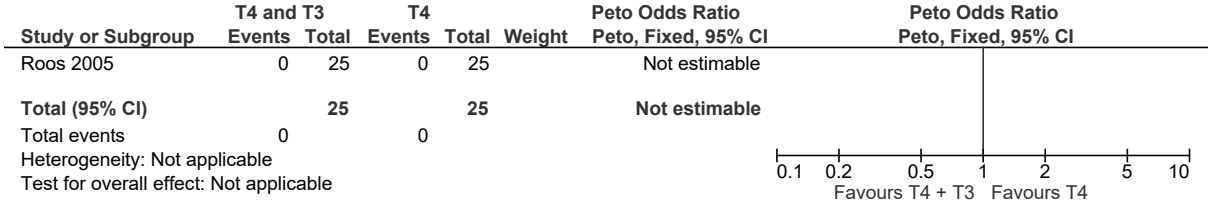


Figure 23: Quality of life: pain (SF-36, 0-100, high is good outcome, 12 months)



1

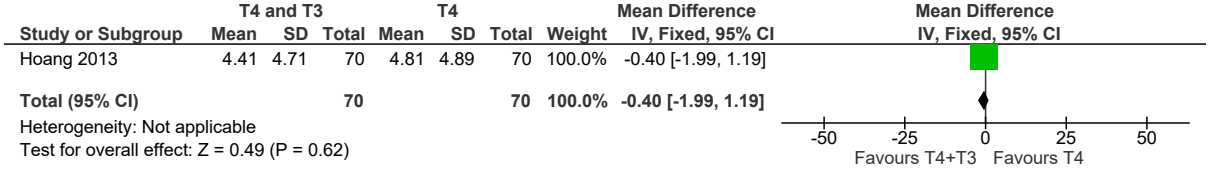
Figure 24: Cardiac events (6 months)



2

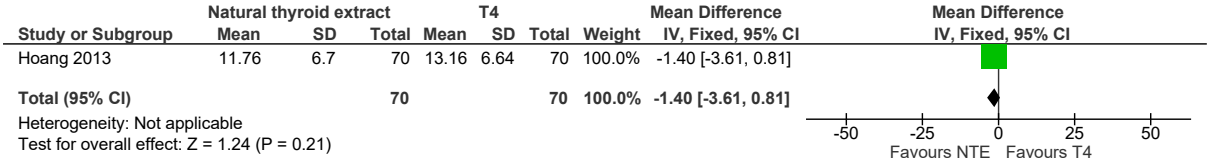
E.3 Primary hypothyroidism – natural thyroid extract vs T4

Figure 25: Depression (BDI, 0-63, high is poor outcome, 4 months)



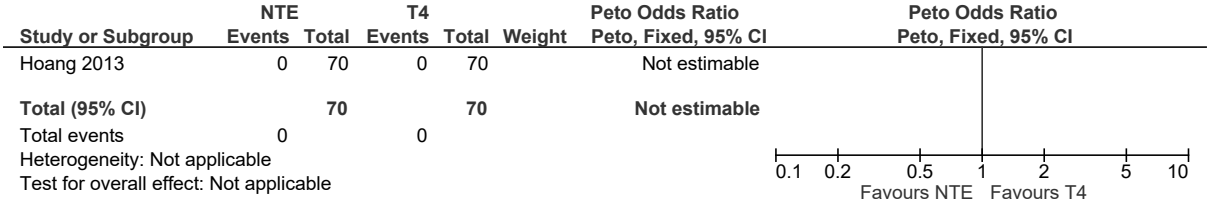
3

Figure 26: Symptom scores (TSQ, 0-36, high is poor outcome, 4 months)



1

Figure 27: TSH suppression below reference (<0.5 µIU/mL)



2

3

Appendix F: GRADE tables

Table 10: Clinical evidence profile: T4 +T3 vs T4 only

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Combined T4 and T3	Control	Relative (95% CI)	Absolute		
QoL-Disease specific (follow-up 4 months; measured with: hypo-specific HR-QoL, high is poor outcome; range of scores: 29-145)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	none	21	20	-	MD 4 lower (17.63 lower to 9.63 higher)	⊕OOO VERY LOW	CRITICAL
QoL-General health (follow-up 12-15 weeks; measured with: SF-36; high is good outcome; range of scores: 0-100)												
2	randomised trials	no serious risk of bias	serious[3]	serious[1]	very serious[2]	none	79	77	-	MD 1.36 lower (16.62 lower to 13.90 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Social functioning (follow-up 12-15 weeks; measured with: SF-36, high is good outcome; range of scores: 0-100)												
2	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	79	77	-	MD 4.61 higher (0.87 lower to 10.09 higher)	⊕⊕OO LOW	CRITICAL
QoL-Mental health (follow-up 12-15 weeks; measured with: SF-36, high is good outcome; range of scores: 0-100)												
3	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	169	122	-	MD 1.55 higher (2.14 lower to 5.23 higher)	⊕⊕OO LOW	CRITICAL
QoL-Role-emotional (follow-up 15 weeks; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	serious[4]	no serious inconsistency	serious[1]	very serious[2]	none	20	17	-	MD 8.7 higher (13.34 lower to 30.74 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Vitality (follow-up 12-15 weeks; measured with: SF-36, high is good outcome; range of scores: 0-100)												
3	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	none	169	124	-	MD 1.44 higher (3.27 lower to 6.16 higher)	⊕OOO	CRITICAL

											VERY LOW	
QoL-Physical functioning (follow-up 15 weeks; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	none	20	18	-	MD 2.3 higher (9.74 lower to 14.34 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Role-physical functioning (follow-up 15 weeks; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	serious[4]	no serious inconsistency	serious[1]	very serious[2]	none	20	17	-	MD 3.4 lower (26.02 lower to 19.22 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Bodily pain (follow-up 15 weeks; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	none	20	17	-	MD 2.7 higher (10.85 lower to 16.25 higher)	⊕OOO VERY LOW	CRITICAL
Depression (follow-up 3-4 months; assessed with: Cases by HADS/BDI)												
2	randomised trials	serious[4]	no serious inconsistency	serious[1]	very serious[2]	none	32/325 (9.8%)	11.1%	RR 0.94 (0.6 to 1.49)	7 fewer per 1000 (from 44 fewer to 54 more)	⊕OOO VERY LOW	IMPORTANT
Depression (follow-up 3 months; measured with: BDI, high is poor outcome; range of scores: 0-63)												
2	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	82	82	-	MD 1.77 lower (3.58 lower to 0.03 higher)	⊕⊕OO LOW	IMPORTANT
Depression (change scores) (follow-up 15 weeks; measured with: SCL-90, high is poor outcome; range of scores: 0-64)												
2	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	110	64	-	MD 2.5 higher (0.05 lower to 5.04 higher)	⊕⊕OO LOW	IMPORTANT
Depression (follow-up 4 months; measured with: GHQ-28, range of scores: 0-21; high is poor outcome)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	30	30	-	MD 0.1 lower (1.66 lower to 1.46 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Symptom scores (follow-up 3 months; measured with: TSQ, high is poor outcome; range of scores: 0-36)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	344	353	-	MD 0.08 higher (0.5 lower to 0.66 higher)	⊕⊕⊕O MODERATE	IMPORTANT
TSH suppression (<0.11 µU/ml, <0.02 mU/l, <0.20 mIU/L) (follow-up 12-16 weeks; assessed with: cases)												

3	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	48/135 (34.8%)	8.7%	RR 2.86 (1.54 to 5.32)	162 more per 1000 (from 47 more to 376 more)	⊕⊕⊕O MODERATE	IMPORTANT

[1] Downgraded by 1 or 2 increments because the majority of the evidence included an indirect or very indirect population respectively
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
[3] Downgraded by 1 or 2 increments because the point estimate and or the confidence intervals varied widely across studies, unexplained by subgroup analysis
[4] 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 11: Clinical evidence profile: T4 high dose vs T4 titrated dose

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**T4 high dose**	**T4 titrated dose**	**Relative (95% CI)**	**Absolute**		
QoL-General health (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	25	25	-	MD 1 higher (2.71 lower to 4.71 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Social functioning (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	25	25	-	MD 12 higher (6.1 lower to 30.1 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Emotional well-being (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	25	-	MD 1 higher (0.87 lower to 2.87 higher)	⊕⊕⊕O MODERATE	CRITICAL
QoL-Role limitis due to emotional well-being (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	25	25	-	MD 9 higher (36.51 lower to 54.51 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Energy (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												

1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	25	25	-	MD 1 lower (6.06 lower to 4.06 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Physical functioning (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	25	25	-	MD 3 higher (5.65 lower to 11.65 higher)	⊕OOO VERY LOW	CRITICAL
QoL- Role limits due to physical functioning (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	25	25	-	MD 9 higher (1.11 to 16.89 higher)	⊕OOO VERY LOW	CRITICAL
QoL-Pain (follow-up 12 months; measured with: SF-36, high is good outcome; range of scores: 0-100)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[2]	none	25	25	-	MD 5 higher (9.42 lower to 19.42 higher)	⊕OOO VERY LOW	CRITICAL
Cardiac events (follow-up 6 months)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	0/25 (0%)	0%	-	not estimable[4]	⊕⊕OO LOW	IMPORTANT

[1] Downgraded by 1 increment if the majority of the evidence was ar 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] 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.
[4]Zero events in either arm

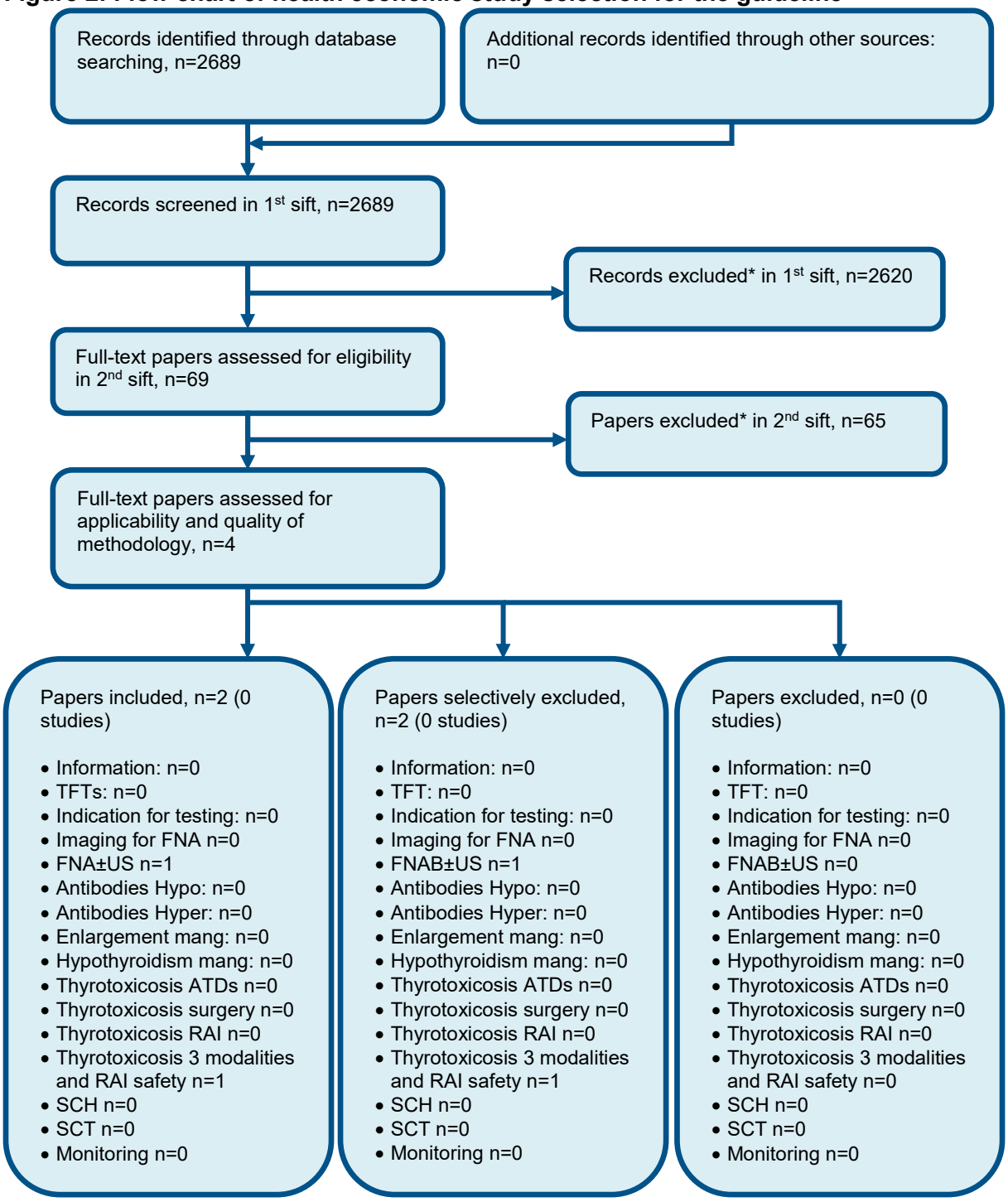
Table 12: Clinical evidence profile: Natural thyroid extract vs T4

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Natural thyroid extract	T4	Relative (95% CI)	Absolute		
Depression (follow-up 4 months; measured with: BDI , high is poor outcome; range of scores: 0-63)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	70	70	-	MD 0.4 lower (1.99 lower to 1.19 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Symptom scores (follow-up 4 months; measured with: TSQ, high is poor outcome,; range of scores: 0-36)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	70	70	-	MD 1.4 lower (3.61 lower to 0.81 higher)	⊕⊕OO LOW	IMPORTANT
TSH suppression (<0.5 µIU/mL) (follow-up 4 months; assessed with: cases)												
1	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	0/70 (0%)	0%	-	not estimable[3]	⊕⊕⊕O MODERATE	IMPORTANT

[1] Downgraded by 1 or 2 increments because the majority of the evidence included an indirect or very indirect population respectively
[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 each arm

Appendix G: Health economic evidence selection

Figure 2: 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.

1

2

Appendix H: Health economic evidence tables

None

Appendix I: Health economic analysis

None

Appendix J: Excluded studies

J.1 Excluded clinical studies

Table 11: Studies excluded from the clinical review

Study	Exclusion reason
Abu-helalah 2010[1]	No usable outcomes
Akintola 2015[2]	Not review population. Systematic review is not relevant to review question or unclear PICO. Systematic review: study designs inappropriate. Incorrect interventions
Angermayr 2004[3]	Not review population. Not guideline condition. Supplementation study in iodine deficient country
Balázs 2008[5]	Not in English
Bunevicius 2002[6]	Less than minimum duration
Carle 2017[7]	No outcome matching protocol reported
Cerbone 2016[8]	Not review population
Cooper 1984[10]	Not review population
Fadeyev 2006[12]	Not guideline condition. Not review population
Fadeyev 2010[11]	No usable outcomes matching protocol
Fan 2014[13]	Not guideline condition. no usable outcomes matching protocol. Not review population
Grozinsky-glasberg 2006[14]	References checked
Ineck 2003[16]	Not review population. no usable outcomes matching protocol
Joffe 2004[18]	Synopsis only
Joffe 2007[17]	References checked
Kachouei 2018[20]	No usable outcomes
Kong 2002[21]	Not review population
Kraut 2015[22]	References checked
Li 2016[23]	Not guideline condition. No usable outcomes matching protocol. Not review population
Ma 2009[24]	References checked
Mahmoodianfard 2015[25]	Incorrect interventions. No usable outcomes
Mainenti 2009[26]	Not review population. Inappropriate comparison. no usable outcomes matching protocol
Martins 2011[27]	Not review population
Mcdermott 2012[28]	References checked
Meier 2001[29]	Not review population. no usable outcomes matching protocol
Monzani 2001[31]	Not review population. Not guideline condition
Monzani 2004[30]	Not review population. Not guideline condition. no usable outcomes matching protocol
Nystrom 1988[34]	Not review population
Panicker 2009[35]	No usable outcomes
Parle 2010[36]	Not review population
Pinchera 2005[37]	Synopsis only
Rayman 2008[38]	Not review population
Reuters 2012[39]	Not review population
Rink 1999[40]	Not in English

Study	Exclusion reason
Ross 1993[42]	Not review population. Inappropriate comparison
Ruggeri 2017[43]	Incorrect interventions. Non-randomised studies. Inappropriate comparison
Samuels 2018[44]	Wrong comparison
Samuels 2018[45]	No additional outcomes to master publication (included)
Schmidt 2013[48]	No usable outcomes
Shatynska-mytsyk 2016[49]	Not guideline condition. Not review population. Inappropriate comparison
Smith 1970[51]	Less than minimum duration
Stott 2017[52]	Not review population
Teixeira 2008[53]	Not guideline condition. Not review population. No usable outcomes to match protocol
Toulis 2010[54]	References checked
Turker 2006[55]	No usable outcomes. Not review population
Van 2013[57]	References checked
Villar 2007[58]	Not review population
Walsh 2003[59]	Less than minimum duration
Wasniewska 2012[60]	Incorrect interventions. Non-randomised study. Inappropriate comparison. Not review population
Weetman 2007[61]	References checked
Wichman 2016[62]	References checked
Wiersinga 2007[63]	References checked
Wiersinga 2012[65]	References checked
Wiersinga 2017[64]	References checked
Winther 2015[66]	Not review population
Winther 2017[67]	Not guideline condition. Not review population
Yu 2017[68]	No usable outcomes matching protocol
Zhao 2017[69]	No usable outcomes matching protocol. Not review population

J.2 Excluded health economic studies

None

Appendix K: Research recommendation

K.1 Research question: What is the clinical and cost effectiveness of using levothyroxine (T4) and liothyronine (T3) combination therapy vs T4 alone in the group of people with hypothyroidism whose symptoms have not responded sufficiently to T4 alone? Does DiO2 polymorphism affect the response to T4-T3 combination therapy?

Why this is important:

Although most people with hypothyroidism are successfully treated with T4 monotherapy, a small subgroup of patients do not feel well on T4 monotherapy despite taking an optimum dose. A number of randomised controlled trials (RCTs) of T4-T3 combination therapy vs T4 monotherapy suggest there is no benefit of the combination therapy in the general population of people with hypothyroidism. However, most of these studies had small sample size, used variable and often non-physiological doses of T3, and had a short duration of follow-up. Furthermore, in some of the blinded randomised controlled trials, patients preferred the combination therapy over T4 monotherapy. Therefore, it remains to be tested in well conducted large RCTs whether T3 given in a more physiological dose and formulation (for example, sustained release formulation) improves outcomes specifically in the population of people who do not respond well to T4 alone. Finally, a post-hoc analysis of an RCT has suggested that an insufficient response to T4 alone may be due to a polymorphism in the type 2 deiodinase (DiO_2) gene although this has not been replicated in further studies. There is no evidence from longitudinal RCTs on people failing to respond sufficiently to levothyroxine to assess whether combination therapy could benefit populations not responding to levothyroxine monotherapy and whether DiO_2 polymorphism could mediate the treatment response.

Whilst current national and international guidelines do not recommend routine use of T4-T3 combination in hypothyroidism, some of these guidelines suggest a trial of the combination therapy in some patients. The limitations in the currently available evidence and conflicting recommendations from different guidelines have led to a wide variation in clinical practice. Furthermore, a sharp increase in the cost of T3 in the UK in the recent years has led to some health authorities (CCGs) banning the NHS prescription of T3 within their localities, leading to a 'postcode lottery' of care. Therefore, there is an urgent need for high quality RCT examining the efficacy and cost-effectiveness of T4-T3 combination treatment in people with hypothyroidism who are not responding to levothyroxine monotherapy.

Criteria for selecting high-priority research recommendations:

PICO question	Population: People with hypothyroidism whose symptoms have not responded sufficiently to T4 monotherapy despite biochemical euthyroidism, subgrouped or stratified by DiO_2 polymorphism Intervention(s): Combination of T4 and T3 (sustained release) Comparison: T4 monotherapy

	Outcome(s): quality of life, symptom control, patient preference, thyroid function tests, adverse effects, cost, impact of DiO_2 polymorphism on the response to treatment
Importance to patients or the population	If T4-T3 combination therapy offers clinically important benefits over T4 monotherapy for people with hypothyroidism whose symptoms have not responded sufficiently to T4 monotherapy, and is cost-effective then it may be an important modality to enhance clinical outcomes in this population. If the utility of DiO_2 polymorphism in predicting response to the T4-T3 combination therapy is confirmed, it could help to identify subgroup of patients likely to benefit from the combination therapy. If the combination therapy is shown not to be beneficial, it will help to stop an unnecessary use of a costly drug, liothyronine, reducing the economic burden to the NHS.
Relevance to NICE guidance	This research will reduce the existing uncertainty regarding the clinical and cost-effectiveness of T4-T3 combination therapy and enable future guidelines to clearly recommend for or against the use of combination therapy in the subgroup of people with hypothyroidism whose symptoms have not responded sufficiently to T4 monotherapy.
Relevance to the NHS	A clear recommendation for or against T4-T3 combination therapy will offer clinicians clearer guidance on whether it should be used in people with hypothyroidism whose symptoms have not responded sufficiently to T4 monotherapy, and whether DiO_2 polymorphism is useful in predicting patients who may benefit from the combination therapy.
National priorities	Hypothyroidism comes under the long-term condition directorate in the UK. A RCT would support a national evidence based approach to treatment of hypothyroidism.
Current evidence base	Although several RCTs of T4-T3 combination therapy vs T4 monotherapy have failed to show a clear benefit of the combination therapy, most of these studies were small, used variable and non-physiological doses of T3, and had short follow-up. In some of the blinded RCTs, patients preferred the combination therapy over T4 monotherapy. It remains uncertain whether T3 given in a more physiological dose and formulation (for example, sustained release formulation) improves outcomes in people with hypothyroidism not responding sufficiently to T4 monotherapy. A post-hoc analysis of an RCT has suggested that a polymorphism in the DiO_2 gene could predict the response to the combination therapy; however, this has not been replicated in further studies.
Equality	This recommendation will help to reduce the current variation in clinical practice and 'postcode lottery' of care in the UK.
Study design	Randomised controlled trial with corresponding health economic analysis.
Feasibility	The number of people with hypothyroidism (inadequately?) treated with T4 monotherapy each year will ensure adequate recruitment. The main challenge will be getting an access to a more physiological preparation in the form of sustained release T3

	tablets for the trial even though such preparations are well advanced in development. Patient recruitment should not be challenging.
Other comments	
Importance	Medium: The guidelines are unable to provide clear recommendations for T4-T3 combination therapy for people with hypothyroidism due to a lack of sufficient evidence. The research would inform future updates.

1

2

3