Author’s response to reviews

Title: Screening for thyroid dysfunction and treatment of screen-detected thyroid dysfunction in asymptomatic, community-dwelling adults: a systematic review

Authors:

Francesca Reyes Domingo (francesca.reyesdomingo@canada.ca)
Marc Avey (Marc.avey@canada.ca)
Marion Doull (Marion.doull@canada.ca)

Version: 1 Date: 31 Jul 2019

Author’s response to reviews:

Peer Review Comments Responses (manuscript edits in bold)

Reviewer #1

1. I'd suggest that the Abstract touch on findings for all the key questions.

So, on page 2, line 33, you might add "no studies on the treatment of screen-detected subclinical hyperthyroidism were found."

If you add this, then you'll need to specify "subclinical hypothyroidism" on line 34.

Similarly, on line 37, you might add a sentence or phrase on the harms that were found (or that few/none were found).

Thank you for your comment.
The following edits (in bold) have been made (line 30).

No studies were found on screening for TD, treatment of subclinical hyperthyroidism, or patients’ values and preferences for screening for TD.

Twenty-two studies (from 24 publications) on the treatment of TD in patients with screen-detected subclinical hypothyroidism were included.

Results from the included randomized controlled trials suggested no benefit of treatment for subclinical hypothyroidism for the large majority of outcomes.

2. On page 6, line 119, "TD affects...5% of people..." may be incorrect data.

The reference cited stated "5% of women and 3% of men"

Thank you for your comment,

however, we would like to keep the sentence as is.

The original paper does note “About 5% of U.S. adults report having thyroid disease or taking thyroid medication (1, 2).”

3. On page 10, lines 202-203, you wrote "suggests that TSH tests are potentially overused..."

I understand that you wrote "potentially overused," however, overuse would depend upon knowing that there were no clinical indications for the TSH test.
Thank you for your observation, we are making a judgement that many tests are potentially overused given that some primary care clinics report near 100% of their patients have a test on record and that primary care practitioners are screening for hypothyroidism in people without symptoms.

4. On page 12, line 243.

I think that should read "or" b) intermediate outcomes.

Thank you for your comment.

This edit has been made (line 300).

KQ3: Does treatment of screen-detected overt or subclinical TD improve (a) morbidity or mortality or (b) intermediate outcomes?

5. On page 13, lines 273-275. I was unclear about your population inclusion criteria of "no history of thyroid disease (except Hashimoto's thyroiditis, subclinical hypothyroidism or subclinical hyperthyroidism)"

I would think that for KQ1 (screening), the population included would be no history of any thyroid disease, and persons with "no history of thyroid disease (except Hashimoto's thyroiditis, subclinical hypothyroidism or subclinical hyperthyroidism)"might be the included population for KQ3 (and KQ4), treatment of screen-detected or subclinical TD.

If so, this should be clarified.

Thank you for your comment.
For KQ1 (screening) we included patients with subclinical hypothyroidism or subclinical hyperthyroidism when the study did not clearly describe enrolment of symptomatic patients.

We wanted to exclude patients who were identified as experiencing symptoms.

Line 332 (page 15) edited to:

Included studies involved patients without a history of thyroid disease (though studies of patients with Hashimoto’s thyroiditis, subclinical hypothyroidism or subclinical hyperthyroidism could have been included as long as patients were not clearly symptomatic).

Studies involving patients with uninvestigated nonspecific symptoms (e.g. fatigue, weight gain) and studies that did not clearly describe enrolment of symptomatic patients were also included.

6. On page 22, lines 464-465, you noted that "All of the...participants in Anderson 2016 had concomitant heart disease..."

If so, I wonder whether any outcomes that include this study (on pages 23-29) should be downgraded for indirectness (applicability of findings in a population with heart disease to the general population).

Thank you for your comment.

In GRADE tables, the explanation for Andersen 2016 ratings was “i. These findings are from studies addressing the key question on treatment. Although this is meant to provide indirect evidence to inform the CTFPHC's recommendation on screening for thyroid dysfunction, indirectness of the evidence to the question on screening will be addressed in the Evidence-to-Decision framework."
Therefore we will not rate down for indirectness in the GRADE tables.”

7. On pages 37 and 39, where you discuss the results for cholesterol and LDL cholesterol levels that were found in several RCTs of treating subclinical hypothyroidism, you note that "results were mixed" or that most studies "did not find a statistically significant difference..." I would note that most of the studies found lower levels of cholesterol and LDL levels in those treated, though most of these were not statistically significant.

As you note on page 53, lines 958-961, the issue here may be that "small sample sizes...may not have...sufficient power..." - rather than a lack of effectiveness.

I know that in the "Data Synthesis” section (p. 16) you stated that "due to clinical and methodological heterogeneity...a meta-analysis was not completed..." Still, I wonder if a meta-analysis might be appropriate for these intermediate outcomes, possibly by restricting the studies included in the meta-analysis to those that were more similar (eg. duration 6 months or longer, or 1 year or longer, etc...).

If you don't think a meta-analysis is appropriate, then a forest plot might be a good way to visually present this data to readers.

Related to this, on page 52, lines 937-946, you mention two systematic reviews with meta-analyses that found a small benefit of treating subclinical hypothyroidism on cholesterol and LDL levels.

It's my impression that the findings of this review are consistent with those two studies.

If you think they are not, then some discussion of these 2 reviews weaknesses (eg., do you think a meta-analysis was not warranted in those reviews), and why this review reached a different conclusion, would be helpful.
Thank you for your comment.

We did not conduct a meta-analysis due to the heterogeneity of the included trials (the trials used different definitions of subclinical hypothyroidism, different doses and dosing protocols were used, varying follow-up time frames, and there was a large variety in the settings of these trials).

Reviewer #2

8. Overall this is a well done comprehensive review that will serve a guideline panel well in deliberations about screening for thyroid dysfunction.

Specific comments/questions follow.

Line 273-275.

Eligibility criteria are confusing as worded; I cannot tell whether those listed in parentheses are explicitly included or excluded.

Similarly, the "no" preceding "history of thyroid disease" could be interpreted to extend to "uninvestigated nonspecific symptoms…"

Thank you for your comment.

The following edits (in bold) have been made (Line 332, page 15):

Included studies involved patients without a history of thyroid disease (though studies of patients with Hashimoto’s thyroiditis, subclinical hypothyroidism or subclinical hyperthyroidism could have been included as long as patients were not clearly symptomatic).
Studies involving patients with uninvestigated nonspecific symptoms (e.g. fatigue, weight gain) and studies that did not clearly describe enrolment of symptomatic patients were also included.


This sentence, as well as the category of excluded studies in the appendix suggests that if RCTs were available, observational studies were not used.

That does not appear to be the case for multiple outcome, and seems to contradict line 298.

Thank you for your comment.

This point has been clarified below:

(Old line 308, 371) A staged approach was used to identify the source of evidence for each outcome for KQ1-4, starting with study type providing the highest quality evidence –RCTs – followed by controlled observational studies (i.e., controlled observational studies were only included for outcomes/populations not already addressed via RCT evidence).

(Old line 298, now 358) Since we were interested in both the benefit of screening or treatment (KQ 1 and 3) and harms (KQ 2 and 4) we included both randomized controlled trials (RCTs) as well as controlled observational studies when RCT evidence was not available.

10. Line 457 (and line 709).

I find no forest plots in Additional file 2.
Thank you for your comment.

This has been corrected throughout.

The summary of the findings is described below.

Further details on the evidence, including forest plots, summary of findings tables and GRADE evidence profile tables for outcomes for KQ3a can be found in Additional file 2:

Evidence Set 1.

11. Line 496-505.

I am confused by the judgements made.

The phrase "moderate certainty that no statistically significant difference exists" in mortality for adults > age 65, appearing in the conclusions and in the evidence table seems to mask the large imprecision of the estimate with a HR that is clinically meaningful.

Perhaps your real conclusion is no statistically significant reduction in mortality?

Much tighter precision exists in the cohort based gender sub analysis; ruling out a change in mortality of 1 per 1000 is a high bar with a large sample size.

Thank you for your comment.
We agree that the imprecision is large here with the absolute effect of treatment ranging from 5 fewer deaths per 1000 to 60 more deaths per thousand.

The difference between placebo and treatment is small in absolute terms here as well (2 per 1000). That said the trial protocol is not powered to detect a difference for this outcome, and the imprecision crosses a clinical important threshold of $\geq 1$ death so we down rated by one level to moderate evidence; see footnote i in the evidence sets.

We believe our phrasing is consistent between outcomes and also consistent with the reviewer’s suggestions: “There is moderate certainty that no statistically significant difference exists between older adults treated and not treated for subclinical hypothyroidism on all-cause mortality.”

12. Line 528.

Inconsistency as well?

Thank you for your comment.

This was rated down for inconsistency for all-cause mortality (for Adults < 65 or <70 years of age), but not for ‘deaths due to cardiovascular diseases’ (line 594).

13 Discussion section:

line 902-905.

This statement seems to express a certainty that contradicts the described methods which include explicit predetermined clinically important differences.
Ignored here is the primary question of whether the precision and consistency of null findings are sufficient to conclude that no clinically significant difference exists.

It is perhaps left to a guideline panel to draw conclusions in this domain.

Thank you for your comment.

The reviewer raises an important point that defining clinically meaningful differences when results are null is difficult.

When interpreting results we relied on statistical significance since clinical meaningful difference could not be established.


Although "prospective" determination of clinically important differences is not possible retrospectively, I was somewhat surprised by the lack of estimates of clinically important benefits.

I agree it is hard.

The USPSTF made an attempt in the evidence review for the Behavioral Counseling to Promote a Healthy Lifestyle for Cardiovascular Disease Prevention in Persons With Cardiovascular Risk Factors (https://www.ncbi.nlm.nih.gov/books/NBK241535/#ch4.s2 - see section on Clinical Interpretation of Benefit and Harms Given Paucity of Direct Health Outcome Data).

I admit this is loose and may have exaggerated health effects.
Thank you for your comment.

We agree that prospective determination of clinical meaningful differences is difficult and we were unable to establish them for many of the included outcomes.

We thank the reviewer for the reference and look forward to additional research that enables us to determine these important thresholds.