Author’s response to reviews

Title: Time for a Reassessment of the Treatment of Hypothyroidism

Authors:
John Edward Midgley (jem.midgley@gmail.com)
Anthony Toft (toft41@hotmail.com)
Rolf Larisch (Rolf.Larisch@klinikum-luedenscheid.de)
Johannes Wolfgang Dietrich (ohannes.dietrich@ruhr-uni-bochum.de)
Rudolf Hoermann (rudolf.hoermann@gmail.com)

Version: 2 Date: 05 Oct 2018

Author’s response to reviews:

BEND-D-18-00171R1

Time for a Reassessment of the Treatment of Hypothyroidism

Response to Reviewers’ Comments
We are grateful to the reviewers for their thorough reviews, constructive criticism and specific suggestions that helped us to improve the manuscript. The manuscript has been revised accordingly addressing all issues raised. Please find a point to point reply to the reviewer’s comments below. We have also highlighted changes made in the revised manuscript.

Reviewer 1
Time for a reassessment of the treatment of hypothyroidism is a very sophisticated, articulate review of several excerpts from the literature which support and oppose the findings of the authors previously published research and opinions. It is clear that the authors consider current evidence based guidelines to be applied restrictively by the clinical community. Based upon previous pieces by the authors, their investigations support the use of combined LT4/LT3 and thyroid hormone extracts for the treatment of hypothyroidism. As the current guidelines do not recommend combination therapy, the authors make the case for utilizing their approach in the clinical setting. The paper flows nicely, is very articulate and provides unique and advanced techniques of analyzing data which provide a platform to interpret their previous research supporting the use of combination therapy in individual patients.
Specific comments:
Page 4 Line 36 Thyroid hormone extract products are "standardized" by adjusting the LT4 and LT3 content after HPLC measurement indicates that the batch in process is not within USP specifications. As these products are now pharmaceutically modified, some have stated that they are no longer "natural".

Response: This point is well taken, and it was not in our intention to emphasize or imply “natural” in this drug. We have reworded the sentence and provided more detail.

Page 4, lines 68ff. … even at this time some patients still regard it as the most satisfactory treatment of hypothyroidism for them (10,11). A policy was adopted by endocrinologists in the 1960s to replace thyroid extract with synthetic levothyroxine as the latter was then more consistent in its content (12-15). More recently, thyroid extracts have been standardized by modern HPLC techniques to maintain their content of thyroid hormones in different batches within USP specifications. Few clinical trials have been performed to compare the efficacy of the two products (15), and an exploratory RCT was conducted only in 2013 (16).


Response: We have added that information and referenced the trials. See above.

Page 5 Line 9 Perhaps the not only is hypothyroidism frequently diagnosed but often "over diagnosed" [mis- diagnosed] especially when subjects present with non-specific symptoms and their thyroid function test results are misinterpreted like single minimally "elevated" TSH, using non age adjusted expected ranges for TSH or FT3 especially in the setting of non-thyroidal illness.

Response: We agree that hypothyroidism is frequently “misdiagnosed” in both directions, and have expanded on this aspect.

Page 6, lines 106ff. For several reasons, this is an area of considerable uncertainty. Firstly, thyroid-related patient complaints overlap with a plethora of non-specific symptoms caused by other conditions and diseases (29-36). Thyroid tests are also more likely to be obtained in patients with unspecific symptoms (37-39). In these conditions, LT4 treatment may not be superior to placebo in symptom alleviation (40-42). Secondly, TSH is increasingly recognised to be less reliable as a definitive diagnostic tool than previously assumed (27). Not only is its reference interval not universally agreed on or adjusted for various influences, such as ethnicity, iodine supply, age, but the univariate statistical derivation of a TSH reference range is inherently ill-defined owing to its nature as a controlling element (43). Physiologically, stimulation by TSH raises thyroid hormones to a level appropriate to the optimal well-being of a person. Because TSH, FT4 and FT3 are interrelated through the operation of hypothalamic-pituitary-thyroid feedback regulation, integrated pairs of TSH and FT4 values define the so-called individual set points (43,44). Unlike a population-based univariate reference interval, set points are subject to multivariate normality and narrow homeostatic ranges (43). When plotting TSH against FT4 concentrations the resulting distribution in a healthy population does not describe the familiar rectangle, but a kite-shaped area (43). Accordingly, a TSH value can indicate true euthyroidism in an individual despite it slightly exceeding the upper reference limit, while a TSH measurement within that reference interval may represent a truly hypothyroid subject (43). Isolated TSH interpretation thereby becomes ambiguous, resulting in unacceptable diagnostic and therapeutic uncertainty surrounding a given TSH measurement when it approaches the TSH euthyroid range (43,44). As a consequence, this strategy divorces diagnostic disease definitions from treatment targets.
Rationally therefore, the triple roles of TSH as a screening test, diagnostic tool and therapeutic target require separate assessment. Diagnostic reliability for patients may be improved by reconstructing personal TSH-FT4 set points, depending on whether this novel approach can be confirmed in clinical trials (45).

Both the non-specific nature of complaints and inherent deficiencies in the diagnostic process raise an unsettling dilemma for patients and thyroid specialists alike.

Line 32 Perhaps much of the problem is assuming that subclinical hypothyroidism (SCHypo) actually IS associated with the non-specific symptoms commonly attributed to hypothyroidism. Or is it the opposite, that the symptoms that lead to the ordering of TFTs (see also Bould H et al. 2012 Family Practice 29:163-167) are not causally related to the minimal variances in TFTs which may be discovered in the search for organic disease. See also (Cathebras PJ et al. 1992 J Gen Intern Med; 7(3):276-86, Stadje R et al. 2016 BMC Family Practice 17:147, Carle A et al. Am J Med 2016;129(10):1082-92, Blum MR et al. 2016 Neuroendocrinology 103:291-299). Obviously, the utility of the diagnosis of SCHypo and initiation of treatment for symptom relief has recently been called into question, especially in the elderly as the non-specific symptoms experienced by so many are very common (See also Carle A et al. Am J Med 2016;129(10):1082-92). Unfortunately, some in the holistic community diagnose hypothyroidism without biochemical confirmation solely based on symptoms and initiate thyroid hormone of various varieties. Recently, it has been shown that in older individuals that TSH and FT4 levels do not correlate with measures of QOL, mood or cognition using age adjusted normal ranges derived from the NHANES data base (Samuels MH et al. 2017 Thyroid 26(9):1185-94). This finding indicates that symptoms do not correlate with age adjusted TSH [values as high as 9 mIU/ml. Additionally, if SCHypo symptoms were actually clinically relevant and responsive to thyroid hormone replacement, the studies by Jorde/Parle and Stott would have demonstrated some symptomatic change when thyroid hormones were provided (Jorde R et al. 2006 JCEM 91:145-53, Stott DJ et al. 2017 NEJM 376:2534-2544). All three of these studies defined the presence of SCHypo just based on TSH greater than a locally determined upper normal. None of these cutoffs exceeded the age adjusted cutoffs used by Samuels. Symptoms were minimal to age expected in their non-specificity, thyroid hormone replacement induced no change in their clinical state. This lack of therapeutic efficacy is likely due to the absence of hypothyroidism as the administration of thyroid hormone replacement has no beneficial effect in those who are euthyroid (Pollock MA et al. 2001 BMJ; 323 (Oct):891-895). So, if the primary care community over diagnoses hypothyroidism, we should not be surprised as specialists when we are referred patients in whom there are residual (non-thyroid dependent?) symptoms once thyroid hormone "normalizes" TSH.

Response: We believe various problems coincide to contribute to frequent misdiagnosis, namely a plethora of non-specific and non-thyroid-related symptoms, placebo-related treatment effects, the designation of subclinical hypothyroidism as a disease, and an imperfect diagnostic process over-reliant on TSH. Perceptions therefore vary by focal point and treatment may also be either wrongly withheld or wrongly administered. However, an inaccurate diagnostic process becomes even more unsettling in such a situation. We expand on these issues in the revised text. See above and Page 5, lines 105ff. As for TSH referencing, given underlying physiology that TSH raises the FT4 /FT3 levels into the “normal” range and in its absence secondary hypothyroidism ensues it follows that the interrelation of TSH and FT4 has to be considered. When TSH and FT4 values from a healthy population are depicted by an xy plot the enclosed reference area is kite-shaped (not rectangular) (43). This means that a TSH value of 7 may be truly hypothyroid for one person and still physiologically - unlike statistically - truly normal for another individual.

Amending diagnostic criteria should preferably avoid non-matching designations where a disease of subclinical “hypothyroidism” is diagnosed, yet treatment is not indicated.

Page 6 Line 6 I agree with the authors that before the era of TSH the severity of hypothyroidism
diagnosed with the techniques of the day was far more severe, the symptoms far more clearly caused by hypothyroidism and the results of thyroid hormone intervention (extract or LT4) for mare likely to relieve symptoms as this was the end point which was available. I offer the alternative explanation outlined above to nuance that with only marginal TSH elevations perhaps thee subjects are not sufficiently hypothyroid (maybe not at all) to respond so dramatically to thyroid hormone replacement. Response: Not only do we fully agree with that statement but we also suggested improved diagnostic criteria to better align them with treatment requirements. For instance, by separating the triple roles of TSH as a screening test, diagnostic tool and treatment target and personalising reference ranges.

Page 6, line 129. Set point reconstruction offers a more precise and personalised approach (43,55), if its practical utility can be proven in clinical trials.

Lines 16-19 Not sure what the authors are expressing here. Either needs a reference or a bit more transparency.
Response: ? need to identify the exact lines

Lines 29-43 I disagree here as I believe that data derived from observational studies were utilized to observe that clinical outcomes were not improved by excessive TSH suppression and therefore, given the prevailing thought that suppressed TSH levels were indeed associated with both atrial fibrillation and bone loss/fracture, it would be safer to back off on the degree that TSH needed to be suppressed from both perspectives. No one is denying that TSH can stimulate thyroid cancer growth, however is does not appear that an undetectable TSH is necessary in the vast majority of thyroid cancer patients.
Response: We may not have been sufficiently clear here and revised this paragraph.

Page 10, lines 241 ff. The cause of atrial fibrillation poses a complex problem, as its occurrence has been physiologically and statistically associated with both high and low FT3 concentrations (69). Thyrotoxicosis due to exogenous thyroid hormone intake and endogenous hyperthyroidism have different physiological roots. This traditional distinction should be noted because the interrelationships between TSH and thyroid hormones differ on LT4 treatment from those in thyroid health (24,56,57,59,62). This may explain why a prospective study measuring surrogate markers of thyroid tissue effects in athyreotic patients found a slightly suppressed TSH to be optimum for these patients rather than constituting overtreatment (57). However, this neither implies that TSH suppression is universally desirable, nor that a suppressed TSH is without risk (24). Rather TSH by itself, unaccompanied by measurements of FT4 and FT3, is an unsuitable risk measure in LT4-treated patients, displaying considerable inherent uncertainty in an individual about the risk - benefit ratio for TSH values close to the lower reference limit (27,62). This problem is paralleled in FT4 measurements, which also overlap significantly at the hypothyroid-euthyroid borderline, both in untreated states and even more so in LT4-treated patients (24,26).
Taken together, a combination of nonspecific complaints, statistical amalgamation bias and limited diagnostic performance of TSH test obfuscates the transition between diseased and healthy state and fosters disagreement of interpretation depending on the respective focal points.
Page 7, line 143. We clarified that this view has recently been revised, and TSH suppression is now deemed unnecessary for many thyroid cancer patients (49).

Lines 43-49 I believe that the reference cited has been superseded by more recent clinical data. (See also Carhill AA et al 2015 JCE&M 100(9):3270-3279).

Response: Thank you, we introduced this newer reference (49).

Page 7 Lines 19-26 Patients complaining of symptoms when they know they are on LT4 and are aware that the doses of LT4 have been reduced is NOT an accurate (objective) assessment of the need for more or less thyroid hormone. (See also Walsh JP et al. 2006 JCE&M 91(7):2624-30, Samuels MH et al. 2018 JCE&M 103(5):1997-2008).
Response: We elaborated on both the role of nonspecific complaints and diagnostic deficiencies. Page 5, lines 106ff, see above.

Lines 39-42 To interpret such data there would need to a be a blinded, controlled experience for the patients as the mere knowledge that they are hypothyroid, are on thyroid hormone and have chronic disease(s) generates symptoms independent of the thyroid function test results. See also (Quinque EM et al. Health and QOL 2013. 11:68, Wouters H et al. ENDO2018 OR34-1, 20 Mach 2018, Massolt ET et al. Clin Endo 2016; 85(5):781-788). This statement should be put into context.

Response: We agree, added more references to that effect and suggest an appropriate statistical approach to addressing patient expectancy in clinical trials.

Page 7, lines 164ff. The expression of dissatisfaction may be partly explained by raised awareness of the problem, based on unspecific subjective criteria, and the possible contribution of a lack in certainty of the diagnostic process discussed above (11,29-44). Patient expectations introduce a confounding influence on perceived outcomes (11,53,54). This is difficult to address, particularly since expectation bias extends to RCTs, regarded as the highest class of evidence in Evidenced-Based Medicine (53). A conflict arises between Evidenced-Based Medicine and FDA regulations, the latter mandating that drug evaluation is strictly done under conditions of actual use (53,55). A statistical remedy (R2R) has been proposed to adjust for expectation bias, but we are not aware of any thyroid-related analysis following such a rigorous protocol (53).

Line 49-50 The reference cited here has nothing to do with thyroid cancer treatment, is a relatively short-term study (only 6-month data) on treating Hashimoto's thyroiditis, concluding that there may be a component of the symptoms that are attributable to the Hashimoto's independent of the hypothyroidism. In addition of the n= 63 Completers TSH was still elevated in 24/63 (38%). The authors should find a more appropriate reference.

Response: This reference was introduced since relevant in the context of symptom specificity because - unlike in other studies relying on general QoL questionnaires - a newer validated thyroid-specific QoL instrument was used. We reworded the sentence.

Page 8, line 184ff. Similarly, in patients with autoimmune thyroiditis, LT4 treatment did not restore quality of life assessed with a validated state-of-the-art and thyroid-specific instrument to that of the healthy population in a large Danish open label study (51). It remains however questionable whether these patients received optimum treatment, since some patients did not have their TSH “normalised” and the pituitary hormone may also be an unreliable marker in this particular setting (51). Line 56 An excellent reference has also recently appeared which supports the drama of patent self-reporting. These reports are problematic as there seems to be no control on who the respondents are and if robots are producing a portion of the reports. This most recent survey supports that the patients are upset over LT4 treatment a bit more than they are over LT4/LT3 and thyroid hormone extract as a large proportion report that other treatments must be found. The patients also seem to be upset with the the physicians who prescribe these therapies. (See also Peterson et al. Thyroid 2018 28(6):707-721).

Response: We have included and referenced this large survey.

Page 7, line 136. This sentiment was confirmed by a large online survey of 12,146 hypothyroid patients conducted by the American Thyroid Association (11).

Page 8 Line 29 It should be made clear that the Winther study was done in an open label manner, all patients were aware of their diagnosis, all knew they were on LT4 and all had Hashimoto's thyroiditis which the authors seem to believe may be an independent factor to be considered when assessing patient symptoms on LT4. Again, it needs to be emphasized that upwards of 38% of the subjects in this study were still hypothyroid as assessed by elevated TSH at the time of the QOL assessments. So this
study can not be directly translated into clinical practice as none of us would assume that symptoms should be attributed to hypothyroidism treatment failure if the TSH were still elevated and sufficient time had not passed to allow for tissue T3 equilibrium to occur.

Response. We added a cautionary note.

Page 8, lines 187ff. It remains however questionable whether these patients received optimum treatment, since some patients did not have their TSH “normalised” and the pituitary hormone may also be an unreliable marker in this particular setting (51,62).

Line 56 FT3 seems to have been repeated here.

Response: Thank you, corrected.

Page 9 Line 26 I do not think most clinicians (primary readership) are familiar with the paradox. This paragraph is above the level of clinician understanding. If the point is that each individual is unique, how are we to assess each individual with the tools we currently have. Symptoms don't help, FT3 is not associated with symptoms in numerous studies, TSH being slightly elevated, high normal or low normal all seem to have the same outcomes. Is a clinician to do?

Response: Clinical researcher can no longer rely on traditional statistical standard tools, but are expected to use more robust statistical methods in personalised medicine. This issue goes beyond thyroidology and has been recognised in other fields (61). Retraction of rosiglitazone as a marketable drug is all but one prominent example where Simpson’s paradox proved to be very consequential (71). We recently demonstrated its direct relevance to the current debate (60).

We had no problem documenting a clear association of the presence or absence of patient complaints in a large retrospective sample of LT4-teated patients with thyroid carcinoma followed long-term including dose variation with serum FT3 concentrations (50). In this study, the association distorted by collider stratification bias (Fig. 2) required a multilevel model to be unmasked in the presence of group heterogeneity (50,60).

We briefly expanded on this, introducing new references 60,61,71.

Page 10 Line 9 Other well done studies reach the conclusion that there is no convincing relationship between Symptoms and individual TFTs. (Massolt ET et al. Clin Endo 2016; 85(5):781-788, Wouters HJ et al. 2017 Thyroid 27(2):147-155, Samuels MH et al. 2018 JCEM 103(5):1997-2008). The authors are asked to acknowledge these data.

Response: Although these studies were well done and have now been acknowledged in context, we believe there are underlying issues based on statistical and physiological principles, which can only be overcome by more robust statistical techniques (61).

Line 19 There are 4 metanalyses all looking at the same, but not completely overlapping studies which also comes to this conclusion. Perhaps for completeness the authors would consider including all of these reviews.

Response: We have included all of them (63-67).

Line 26 Might acknowledge that the end of study preference info is subjective and when in placebo-controlled trials, patients tend to prefer the study period which they perceive they were on the highest dose of LT4(Samuels MH et al. 2018 JCEM 103(5):1997-2008). In the Hoang study of course only 48% preferred the thyroid hormone extract phase while the majority (52%) either liked LT4 better or could not tell a difference.

Response: We briefly address adjusting for patient expectations, which is a major problem.

Page 7, lines 168ff. Patient expectations introduce a confounding influence on perceived outcomes
This is difficult to address, particularly since expectation bias extends to RCTs, regarded as the highest class of evidence in Evidenced-Based Medicine (53). A conflict arises between Evidenced-Based Medicine and FDA regulations, the latter mandating that drug evaluation is strictly done under conditions of actual use (53,55). A statistical remedy (R2R) has been proposed to adjust for expectation bias, but we are not aware of any thyroid-related analysis following such a rigorous protocol (53).

Line 42 Agree with the authors that patients with thyroid cancer seem to have symptoms and the presence of symptoms has a negative impact on their quality of life. This is a far cry from demonstrating that these symptoms are caused by some deficiency in thyroid hormone replacement therapy (which the Hedman et al. did not do. Again the work of Massolt ET et al. Clin Endo 2016; 85(5):781-788 concludes that co-morbidities are a better explanation of symptoms in these patients than an TFT. The authors should reflect an acknowledgment of these contradictory studies.

Response: In a large retrospective longitudinal analysis with LT4 dose variations in patients with thyroid carcinoma we observed a clear association of symptoms with biochemical tests, in particular FT3 (50). We offer an alternative explanation of statistical deficits (collider stratification bias), which are increasingly being recognised (61), and suggest improvements in trial design.

Given the physiology of an “optimum” set point, it follows that for instance the relationship between FT3 and Qol measures is u-shaped and therefore predictably missed by classic linear logistic regression (Larisch R, Schulte S, Hildenbrand G, Hoermann R. The role of thyroid hormones in anxiety and depression (Abstract). Nuklearmedizin. 2015;53:V162.).

Line 52 This may also be explained by the observation that patients tend to prefer doses of LT4 that they perceive to be higher (Samuels MH et al. 2018 JCEM 103(5):1997-2008).

Page 11 Line 5 I do think that TSH is a prognostic tool in predicting bone loss (Uzzan B, et al. J Clin Endocrinol Metab. 1996;81:4278-4289), fracture (Bauer DC, et al. Ann Intern Med. 2001;134:561-568, Flynn RW et al. 2010 Metab. 1996;81:4278-4289), fracture (Bauer DC, et al. Ann Intern Med. 2001;134:561-568, Flynn RW et al. 2010 JCEM 95:186-193, Blum MR et al. 2015 JAMA 313:2055-2065) and Afib (Sawin CT, et al N Engl J Med. 1994;331:1249-1252, Cappola, AR et al 2006 JAMA 295(9):1033-41, Flynn RW et al. 2010 JCEM 95:186-193). The fact that the degree of TSH suppression found in thyrotoxicosis is higher that that observed when the TSH is in the expected range is understandable to me. A higher degree of elevated circulating thyroid hormones sufficient to suppress TSH should have more impact on bone and cardiac outcomes than when the TSH is normal. In the Chaker study indeed FT4 (for the most part in the expected range) levels appear to predict the onset of Afib while TSH levels in the expected range do not. As the authors know, T3 or FT3 levels were not analyzed in this cohort and this study is focused on the 98% of subjects dependent on endogenous thyroid hormone production. The LT4 component in this study is quite small, but when the 720 LT4 users were analyzed separately, the 12 Afib events analyzed did not show a statistically significant risk (where the FT4 levels would be expected to be the highest observed). 720 subjects at risk is not trivial, if the FT4 levels generated by LT4 ingestion were a clear risk, I would have thought that a clinician might observe this from time to time. Statistical significance (not in this trial) does not always equate to clinical significance. The authors assertion here may be misconstrued by the reader and should be put in proper context.

Response: This requires very detailed examination. We discussed some of these aspects elsewhere (60). TSH is not a risk factor by itself, rather its association with “abnormal” thyroid hormone levels is. Importantly, those equilibria differ in health and LT4-treated disease. A suppressed TSH can be “normal” for certain athyreotic patients, not for all, and certainly not for those with an elevated FT3. Consequently, a prospective study has shown that a slightly suppressed TSH may be optimally “euthyroid” based on T3 tissue surrogate markers for thyroid carcinoma patients (Ito et al., ref. 57).
However, we generally agree with the reviewer and added a cautionary note that uncritical and unwarranted TSH suppression is not universally desirable and may have risks. We referenced a recent review (60), covering that aspect and relevant references. Page 10, lines 249 ff. However, this neither implies that TSH suppression is universally desirable, nor that a suppressed TSH is without risk (24). Rather TSH by itself, unaccompanied by measurements of FT4 and FT3, is an unsuitable risk measure in LT4-treated patients, displaying considerable inherent uncertainty in an individual about the risk - benefit ratio for TSH values close to the lower reference limit (27,62).

Line 53 I would say the opposite, combination therapies fail "the superiority test". Combination therapy is far more expensive less convenient (TID dosing closest to physiology with current preparations) and potentially supraphysiologic post absorption T3 levels when ingested in commonly used LT3 doses. So if it is not superior (n of 1 trials aside) why recommend this to patients when a safe, inexpensive, non-inferior product is available?

Response: We recommend combination therapy more discriminatorily. We opt for it primarily in symptomatic patients with a demonstrated inability to raise FT3 levels sufficiently (above median) despite LT4 dose escalation. We have expanded on the role of treatment response heterogeneity and potential candidates for combination therapy.

Page 11, line 270ff. However, heterogeneity of the observed treatment response and collider stratification bias require targeting homogenous subgroups and performing statistical latent class analysis (61). This may identify patients that preferentially benefit from the two modalities (61). TSH and FT3 dissociate under LT4 treatment, particularly in athyreotic patients where equilibria are formed between TSH and FT4/FT3 different from the healthy state (56,57). Poor T3 converters with persisting symptoms may thus be the most suitable candidates for trials of T3/T4 combinations. T3 addition may also avoid LT4 dose escalation resulting in T4 excess, as T4 has been implicated in non-genomic actions, not mediated via T3, such as actin-related cell migration (74).

Page 12 Line 9 I believe thyroid hormone production and physiology is different in rats.

Response: Apart from important physiological differences, there are similarities in that disturbed equilibria documented in the animals are equally observed in athyreotic humans. For that reason, TSH does not appear to be a reliable indicator of tissue euthyroidism when approaching its reference range, irrespective of species. When prospectively studying surrogate markers for tissue thyroid effects Ito and co-workers (57) reported that “TSH normalisation” was non-optimum, rather optimum outcomes based on these markers were associated with slightly suppressed TSH levels.

Line 22 In the US this would exclude thyroid hormone extracts as they are not FDA approved and have been removed from the list of approved drugs eligible for Medicare reimbursement under Part D of Medicare. Please make this clear in the conclusion.

Response: We thank the reviewer for pointing out this detail.

Page 12, line 288. We changed the wording to “currently available” and “prescribable.

Line 32 I agree, seek alternative explanations for the complaints before anchoring in on the thyroid as the cause of all complaints in those with well documented hypothyroidism.

Line 53 It is unclear what is meant by the term "unresolved hypothyroidism" when the symptoms are so non- specific and the differential so broad. Again anchoring on hypothyroidism as the root of all symptoms may be a dis- service to the patient.

Response: The question of perceived versus true hypothyroidism remains “unresolved” for various reasons including differences in perspectives and official disease definition in the absence of treatment necessity.
Page 12, line 297. This was replaced with “unresolved designation of hypothyroidism”.
Page 13 Line 1 I agree, the trend is prescribe more LT4 (labeling the patient as hypothyroid whether hypothyroidism has caused their symptoms or not), decreasing the TSH threshold generating a LT4 prescription thus likely over diagnosing "hypothyroidism as a cause of symptoms".

Response: Over- and under-diagnosis when facing non-specific complaints may be unsurprising, and the situation is further aggravated by the non-physiological use of univariate TSH reference ranges in a population and a discrepancy between applied diagnostic and therapeutic ranges.

Line 29 I agree, the medical profession seems to be taking a short cut, not listening to the patient, not considering the differential diagnosis of the complaints and giving out LT4 for the non-specific symptoms associated with a diagnosis of hypothyroidism as "hypothyroidism is easy to treat" and to move the patient along in busy primary care practices. I am not surprised when many "non-thyroid" symptoms persist. I do not equate this to a failure of LT4 to resolve the symptoms but rather a failure of LT4 to improve the QOL in euthyroid individuals as Dr. Pollack taught us many years ago.

Response: Nor do we. We just note that in such a complex situation there should be more consideration and more attention be given to patient symptoms and we should not be over-reliant on a single lab test (TSH) with demonstrated fallibility and pitfalls of its own.

Page 14 Line 36 Is this a research report (new data?) or a review/opinion piece?
We submitted this as a debate based on our own views and research. There are excellent review articles available on the subject and our intention was not to replicate their approach but to encourage a critical debate, emphasizing recent advances in thyroid physiology and statistical methods that have yet to find their way into clinical thinking and study planning.

Finally, the reviewer’s time and very detailed comments are very much appreciated. They made us aware of some misunderstandings and unintended possible implications and helped us considerably improve the manuscript.

Reviewer 2

REVIEWER COMMENTS FROM REPORT: This manuscript is a comment/debate on the ongoing controversy regarding substitution treatment in hypothyroid patients. The discussion is well-written but primarily represents the opinion of the authors. As such, I do think that it is an important addition to the debate.

REQUESTED REVISIONS:

ADDITIONAL REQUESTS/SUGGESTIONS:
Please see previous comment.
Response: We submitted this as a debate based on our own views and research. There are excellent review articles available on the subject and our intention was not to replicate their approach but to encourage a critical debate, emphasizing recent advances in thyroid physiology and statistical methods that have yet to find their way into clinical thinking and study planning. We appreciate the reviewer’s favourable view and comment on this.

Page 9, line 209 and Fig. 2. We replaced “amalgamated” by a simpler term “combined”.
Page 11, line 257. We added another reference explaining the Simpson’s paradox with a practical and consequential real-world example, the retraction of rosiglitazone as a marketable drug (71).

Page 11, line 278. The quote has been translated into English.