Professor Peter A Kopp

Editor

Thyroid

Dear Professor Kopp,

Please find attached our revised manuscript **Clinical parameters correlate better with levels of thyroid hormones than TSH – a systematic review and meta-analysis,** by Stephen Fitzgerald, Nigel Bean, Henrik Falhammer and Simon Tuke. We are grateful to you and the reviewers for the consideration of our manuscript and for the suggestions to improve and clarify our work.

We have made additions and changes to the manuscript to address the reviewers’ concerns. These additions and changes are marked in the text. In particular, in response to the reviewer concern that pregnancy outcomes were not included in our original analysis, we have updated and expanded the review. This larger analysis confirms the results of the original manuscript. In addition we have made other changes which we believe improve the paper- these too are highlighted.

Reviewer Comments

**Reviewer 1**:General comments:

This is an interesting study in which the authors have conducted a systematic review of the literature examining correlations between TSH or respective free thyroid hormone levels and other relevant clinical parameters. The authors have incorporated some standardized systematic review methods including searching multiple electronic databases, supplemented with cross-references of papers as well as what appears to be duplicate independent review of citations and papers (the later could be more clearly stated). It is not clear if data extraction and critical appraisal of the literature was performed in duplicate. The meta-analysis statistical methods are interesting but unconventional and a conventional weighted technique should be included in the report. The manuscript is generally well-written.

*Response: Thank you. The specific points are discussed below.*

Specific comments:

1. Was data extraction and critical appraisal of the studies performed independently by two reviewers? If so, how were any discrepancies resolved.

*Response:Data extraction was performed independently by two of the authors. Discrepancies were resolved by consensus, on the understanding that in the absence of consensus the default position was to include papers.*

1. The authors should tabulate the detailed correlation data from each study and report this in the manuscript. This would include the correlation estimates with 95% confidence intervals (if available), type of statistical test, and statistical significance, as well as sample size if different from that presented in Table 1.

*Response: The detailed correlation data have now been supplied as requested in the revised manuscript as supplemental data.*

1. A conventional weighted technique should be used for the meta-analyses (some examples of established techniques include those reported by Hedges and Olkin, 1985 or Schmidt and Hunter, 1990). Statistical heterogeneity of the meta-analyses should be reported. The description of the number studies reporting a statistically significant result and logistic regression model presented in the manuscript are interesting and can be retained, but do not really replace that of a traditional weighted statistic.

*Response: If we were solely interested in the predictive power of a single thyroid test, such as TSH, with regards to a single condition, such as Atrial Fibrillation, then we agree that it would be most appropriate to use a traditional meta-analysis. However, in this paper, we are considering a more general question - which thyroid test is most predictive of a wide class of conditions, using data using many different methods of statistical analysis. To recap, to perform a traditional meta-analysis we would need to compare like to like, that is we would need to select only those analyses that correspond to the same condition (eg Atrial Fibriliation or Breast Cancer), and using the same analytic methodology (eg Cox’s Proportional Hazard, or Pearson’s correlation coefficient) and then analyse the effect size of TSH and FT4 separately. We did consider the possibility of performing a traditional meta-analysis, as described above. Unfortunately, we found that once you count the number of papers considering the same condition and using the same methodology, there are very few condition-method pairs with at least two papers to compare, thus rendering meta-analysis incapable of making a useful contribution to answering our question. Also, even if we could perform such a meta-analysis on a range of conditions, we would still be left with the problem of combining those analyses to address out question as to which thyroid test is most predictive of the range of considered conditions.*

*As a result, we have developed a sophisticated statistical analysis using random effects models to address the issues of:*

* *the single cohort of people in a paper being considered across a range of conditions, thus introducing a layer of dependence;*
* *repeated analysis of such conditions in a paper by considering different statistical methodologies and by further restricting the population to remove any potential disturbances;*
* *repeated analysis of the above results by considering increasingly sophisticated models through the introduction of further covariates.*

*We are convinced that the usual concerns of using “vote counting” approaches are not valid in this occasion because each analysis of FT4 is controlled by exactly the same analysis of TSH on exactly the same cohort of patients, and so the results are self-controlled. However, in case this could be a concern, we also checked whether there was any association between the thyroid test (FT4, TSH, T3) and the P-values, sample size et cetera, and found nothing of interest.*

1. The correlational data is presented, but there are significant challenges in trying to apply such data in establishing a cut-point defining disease.

*Response:We agree. We have in the revised manuscript indicated that a cut-point can only be arbitrary; our work confirms that the effects of thyroid function are a continuum within and outside of the normal range.*

Reviewer: 2

Comments to Author

This is an increasing manuscript which seeks to demonstrate through a meta-analysis of cross-sectional and cohort studies that thyroid hormones correlated better with certain clinical phenotypes of thyroid status – esp. A-F, osteoporosis and cancer – than TSH. The authors conclude that TSH should not be the primary diagnostic parameter of thyroid status.

*Response: Thank you.*

1. The models appear to all presuppose a linear rather than U-shaped relationship between thyroid parameters and phenotype. Can alternative non-linear modelling be used to accommodate parameters such as cognitive decline, cardiovascular mortality and frailty

*Response: Our review relies on the relationships of the original papers and as such does not presuppose a linear relationship. Some of the relationships in the original papers certainly appear linear, but others such as fatty liver, frailty and mortality do appear to be U-shaped. At least one of the papers (Bano –frailty) explicitly documented a U shaped relationship.*

1. It should be noted that this is not an individual patient meta-analysis – the authors should discuss the potential weaknesses of not using this approach

*Response: In the revision we devote space to more discussion of the statistical methods and their potential weaknesses. The statistical methods have been carefully considered and we believe are appropriate for this work that covers many disparate outcomes. We have conducted many analyses to cover any potential weaknesses and to exclude as possible any artefactual influences on our results. In addition, the results of our analysis are not equivocal; they are so strong that we believe that they would withstand any additional analysis. We note furthermore that the conventional meta-analyses of individual outcomes done to date support our conclusions. Taken on their own even just these two meta-analyses stand as evidence against the TSH- based approach to thyroid function testing. See also the above response to reviewer 1 (point 3).*

1. Pregnancy outcomes are quite sensitive to thyroid function – it does not appear these were included. Can the authors please justify.

*Response: We initially did not study pregnancy to avoid its complexity- different physiology/outcomes of mother and baby etc. We have in the revision made an updated search and have now included pregnancy outcomes. It seems that the superiority of thyroid hormone over TSH levels,in terms of correlation with clinical parameters,is at least as strong with this group of parameters, confirming the overall result.*

1. To evaluate that analysis, it would be helpful to see more of the detail. For example, separate analysis/graphs for each of the major phenotypic outcomes should be included, perhaps in supplementary material.

*Response: This has now been added as suggested.In particular see the supplemental material.*

1. It would be easier to evaluate Table 1, if there were separate columns for study population size, age and cohort type, plus subtotals of numbers of subjects included for each parameter (e.g. A-F).

*Response: The table has been adjusted. Because there were so many sub-analyseswe did not include subtotals of number of subjects for each analysis;all of the different n values could not be accommodated. The n supplied gives an overall guide. All values of n are however, all available on the submitted supplement (point 2- reviewer 1).*

1. Please confirm that no subjects were on T4 therapy.

*Response*: *We have now included a paragraph on T4 therapy. Very few patients were on T4; their removal in the individual studies had no effect on the correlations; but they were too few to be analysed separately.*

1. The discussion is far too long (11 pages) and often repetitive and speculative. It should focus on the strengths and weaknesses of the approach and the potential relevance with only brief reference to “set point” theory, since this is speculative

*Response: We have adjusted and further referenced the Discussion, and in particular reduced repetition and the ‘set point’ discussion (the latter by over a half). Nevertheless we feel this subject has some importance as the set point concept is at the basis of the rationale used to justify the current dogma. A concern of ours was the potential problem of some readersbeing resistant to the conclusions of our paper on the basis of conflict with a traditional understanding of thyroid regulation. Similarly we have addressed many of the previous explanations/dismissals of what apparently seemed at the time to be an aberrant result- i.e. the superiority of thyroid hormone levels.*

*The discussion is thereby still longer than seen in typical single issue papers.*

*However, as this paper addresses a fundamental aspect of thyroid medicine its implications are broad. In the absence of any major rebuttal of our conclusions some of the principles of thyroid regulation, thyroid function testing and classification, and the research aspects of borderline thyroid function would appear to warrant re-consideration.Though we are prepared to edit the Discussion further as requested, we feel that these implications warrant introduction in our manuscript.We simultaneously wish to address any potential points of controversy in advance, and, indicate challenging implications.*

*We have also included the additional material suggested below.*

8. The discussion should include discussion of local regulation by deiodinases and transporters such that even serum levels of thyroid hormone are blunt estimates of intracellular levels.

9. Reverse causation is mentioned. Discussion of this should include reference to the use of genetic polymorphisms associated with thyroid hormone levels and the role of the Mendelian randomisation approach in determining the direction of causation.

*Response: In the revision we have included some discussion of thyroid hormone transporters, deiodinases and the fundamental importance of intra-cellular T3. We have also discussed the information that can be gleaned from study of genetic polymorphisms and Mendelian randomisation. The latter is included in an augmented discussion of reverse causation in turn supported by further analyses to isolate this potential problem.*

*We claim no expertise in these areas, but have not found any evidence to suggest that consideration of any of the above points challenges the conclusions of our work.*

10. Comment should be made on the potential impact of the wide range of different assays used in the papers analysed.

*Response: In the revision we have indicated our chosen statistical methods have has encompassed studies that differ from each other in many ways. We believe that this may be a strength of our study, i.e., we conclude that thyroid hormone levels are more likely than TSH levels to correlate with clinical parameters, regardless of the parameter, statistical method, population, assay etc. chosen. With particular regard to the wide range of assays used in the papers analysed, we would expect that, so long as the assay is valid, the correlation would be valid and consistent, i.e., we would expect the ranking of each individual in terms of any parameter to be generally consistent across different assays*.

We again thank the referees for improving our manuscript substantially. We hope that the Editor and reviewers find the above-suggested changes are in line with their intentions and will find our manuscript of sufficient quality to warrant publication.

Yours sincerely,

Stephen Fitzgerald MBBS. FRACP.

The Royal Adelaide Hospital