Reviewer: 1

Dear reviewer 1, we carefully read your comments and respectfully respond to them:

***This study is a systematic review and meta-analysis performed to identify how ‘incidental’ thyroid cancers were diagnosed, with a goal of estimating frequency of incidentally discovered thyroid malignancy. The authors included 17 studies, for a total of 4,470 patients, of which they report that half were identified incidentally through imaging studies, such as ultrasound, and that most were small papillary thyroid cancers.  
  
The authors define incidental as those who had a thyroid nodule harboring thyroid cancer found during use of imaging for reasons unrelated to thyroid nodular disorder or symptoms AND also as those cancers found when a thyroidectomy was performed for other indications. Non-incidental thyroid cancers were those identified because of an abnormality on PE, presumably a clinical indication to obtain a thyroid ultrasound. While I understand that the nodule was identified incidentally, presumably during the course of evaluation of the nodule, a thyroid cancer was identified, thus prompting surgery. Thus, the ‘incidental’ nature of this is slightly different than a true incidental histological finding. In addition, we do not know the true ‘n’ (realize this would be near impossible) of nodules identified, those that did not have surgery, etc. I would suggest the authors divide the categories into 3 – incidental nodules with malignancy, incidental cancers found on histology, and the existing non-incidental malignancies. I don’t think the first two can necessarily be grouped together***

* Because of your advice, we decided to add two new graphs presenting the frequencies of incidental findings by imaging (incidental nodules with malignancy) and histology (incidental cancers found on histology), non-incidental, and their definitions (figure 2). The other graph shows the analysis of imaging and histology separately and their corresponding subgroup analysis (figure 4).
* Regarding the third category (non-incidental malignancies), it is worthy clarifying the reasons why we reported its definition in the methods section and why we did not analysis it. Firstly our aim was to analysis the incidental diagnosis, and secondly we used this category as a guide in order to be sure that we were including the right information of incidental diagnosis.

***In addition, the authors state in the discussion that thyroid ultrasound may be overused, which I’m not sure is borne out by the data shown. They may be small cancers, but the authors do not show the data that evaluation wasn’t necessary – particularly in light of the fact that we don’t know the biopsy results/indications for surgery for those nodules identified on imaging (please see my above comment regarding how incidental is defined). In this case, is the problem the thyroid ultrasound or the preceding study? These happened to be the studies ordered for some other clinical reason – without knowing these, what is the relevance of which study identified the original nodule, which required formal evaluation with ultrasound?***

* We agree with your comment, the statement “ultrasound overused” is not borne out by the data shown, that is why we added your critics as part of our limitations (lines 345-352).
* As you mention knowing the preceding indications for ultrasound imaging is important to better understand the ultrasound overused, however, to assess these indications is few feasible, firstly because the ideal study to examine the indications to order a thyroid ultrasound is a prospective cohort study and the current published studies have not this design, and secondly because there could be thousands of indications that could generate the use of ultrasound, such as: examination of carotid arteries, sore throat pain, gynecological exam, hypothyroidism, etc. Moreover these indications vary depending on the country, type of hospital where one study was conducted, the variety of specialties that could exist in each hospital, the age and the sex of the patients, etc.

***Finally, the authors suggest renaming small thyroid cancers as ‘papillary lesions’ but again, the intent of this study does not demonstrate clinical outcomes of these patients, which makes this suggestion unsupported by evidence in this particular study.***

* We agree with your comment, the design of this systematic review is not proper to suggest renaming small thyroid cancers as papillary lesions, therefore we re-wrote lines 335-342 and changed the context of the idea.

***Other comments:  
  
1.      Regarding the Methods section, the first few lines seem unnecessary – lines 98-100.  
2.      For the Data Sources and searches, while the strategy is in the appendix, it may be helpful to include a brief summary in the text.  
3.      Study selection –  
a.      Line 117: which reviewers screened abstracts and titles?  
b.      Please clarify what data was looked at in each of the pilot studies and how these were able to help with feasibility of the study.  
c.      Line 123: who are the ‘both reviewers’ being referred to?  
4.      There is an abundance of figures in the appendix – would consider how to decrease these in number and/or if all are necessary. This correlates to the text on pages 188-193; while it shows the range of studies in which thyroid nodules were identified, I’m not sure what it means, clinically.  
5.      Line 177-178: please clarify if the papillary microcarcinomas identified were 26% of the 89% of papillary thyroid cancers (assume so) – or microcarcinomas represented 26% of ALL cancers identified?  
6.      Paragraph beginning page 11: it is not clear why the authors chose the 40% threshold for the figure/text.  
7.      Two minor typos:  
a.      Line 89: should be lesions  
b.      Line 113: ‘were’ excluded***

- Regarding to the other comments:

1. Although those lines may be seen unnecessary, we believe that being transparent is a cornerstone in the development of research, as the protocol shows the plan we performed before starting this study. We decided to keep the same lines as stated in the original manuscript (lines 99).
2. We agree with your comment and we added a brief summary of the research (lines 115-117).
3. We agree with your comment and:
   1. We added the authors that screened abstract and titles (line 122).
   2. We re-wrote lines 125-126 stating how pilot helped with the feasibility of the study.
   3. With the information added on lines 122, reviewers worked independently and duplicate, thus both reviewers refers to each pair of reviewers.
4. We agree with your comments and we decided to move out all the figures in the appendix to Github, with the coding scripts and excel file. Regarding to the clinical significant of the number of the studies we decided to re-phrase the paragraph (lines 203-206).
5. We agree with your comment and we re-wrote lines 195-197.
6. We agree with your comment and we decided to show the different proportions of the countries and delete the 40% threshold.
7. We agree with your comment and we re-wrote lines 90 and 117.

Reviewer: 2

***This is a systematic review and meta-analysis examining how thyroid cancer is detected trying to identify drivers of over-diagnosis.  
  
I think the authors are interested in addressing an important question and have tried to apply significant scientific rigor to their efforts, but unfortunately the data they are trying to analyze has significant limitations and can't provide the granularity of data necessary to provide the answers to their questions.  
  
The concerns I have with this paper have less to do with the methodologic/scientific rigor, but more with the quality of the data in which they are trying to apply it to and how these results can be interpreted and conclusions drawn.  
  
As I read this paper, it came across that the authors have their ideas of what is driving this problem and are looking to try to find a way to find data to support those ideas.  The only finding that the data clearly supports is that approximately 50% of thyroid cancers are found on imaging and are labeled as "incidental".  It really cannot support any causation for this problem, or really support many of the conclusions and suggestions made in the discussion.   The discussion of this paper is really more of just a summary of their opinions and believes about this issue and none of that is supported by the data presented in the paper.  
  
Here are some of my concerns:  
1.  How a lesion is described as "incidental" is highly variable.  Just because it was picked up on imaging, doesn't automatically imply it was incidental.  In most of these studies there is no analysis of why the study was ordered.  Was something felt on exam?  Was the patient complaining of something that led to the exam? There is no way in this type of analysis you can know if imaging was ordered appropriately or inappropriately and no way to know if these nodules were truly incidental findings or not.   The authors want to argue that a solution to this problem is to order less imaging, but there is really no data included that can demonstrate that this imaging was done for un-indicated or inappropriate reasons.   
Ex. If someone is getting an carotid US for TIA symptoms and they find a 2 cm thyroid nodule in the images.  It is an incidental finding, but they can't really pretend they didn't see it or fail to report it. The intention of the exam wasn't to look for thyroid nodules, but if they are seen, it is unethical to not report it. You can't really argue that it was unnecessary or inappropriate imaging.   I think for many of these lesions, you can't prevent their incidental detection on imaging, but you could change how we respond to them when they are found.  What are the criteria to do an FNA?  Why are FNAs being done on nodule <1 cm?***

***2.  The rates of incidental cancers also varies based upon rates of thyroidectomy for treatment of other conditions.  Again in this design it is impossible to say if the indications for surgery were inappropriate or in anyway impacted by the findings of a possible small thyroid cancer.  Utilization of thyroidectomy for benign conditions such as Graves' and Hashimoto's has increased and more incidental cancers are found in those patients.  It is impossible to imply from this data that rates of thyroidectomy are increasing because of these incidental findings and it is inappropriate to stay that increased utilization of thyroid surgery is "unnecessary" or "inappropriate".   It may be the opposite. Maybe more surgery is being done (which is completely appropriate for other indications) and hence more incidental cancers are found.***

***3.  The rates of incidental cancer varied between countries and studies, but I would be very hesitant to suggest that  Canada has the lowest rate of incidental cancers at 17% based on 1 study from that country.  The variability between studies is partly related to the country of origin and practice patterns, but also highly dependent upon definitions used and study design.  I am not sure that this study really has the ability to effectively discern what part of that variability can be attributed to country alone versus methodologic differences.  
  
4.  indications for US imaging would be critical to better understand this issue, and while they highlight a few studies in the discussion that examined this (poor correlation with PE findings and lack of appropriate indications), the data available in the current study simply does not provide any data to clarify our understanding of these issues.  
  
I think the authors are trying to answer an important question, but this is a question that can't be answered with the methods used.  I think the scientific findings are just the frequency of incidental cancers, which I think is well known and not really a novel finding.  The discussion highlights some of the factors that may be at play, but is really more of an opinion piece and isn't based upon any data presented.  I think the discussion is really more of a review article on the topic.***

Dear reviewer 2, we carefully read your comments and respectfully respond to them:

Thank you for your suggestions, we decided to do the following:

1. We have unintentionally missed to add two steps in the manuscript: firstly a paragraph in which we explain how the homogenization of the incidental and non-incidental definitions of the studies was done (lines 134-142), and secondly the supplementary table with the incidental and non-incidental definitions used by each study, with their corresponding indications for ultrasound imaging (Table A3).

Although we have extracted the indications for ultrasound imaging to better understand how it is affecting the proportion of incidental thyroid cancer, we could not know the specific indications with granularity because only 4 studies reported it. Moreover, few of these indications retrieved are clearly explained. Therefore an analysis whether an indication is appropriate or not, was impossible. However, even though we could not describe in detail whether or not the neck ultrasound indications were appropriate, we decided to add this information to generate a global overview of how the indications are contributing in the frequency of incidental diagnosis and to encourage other authors to report studies aiming this issue. This information in results (lines 251-26), discussion (lines 284-288), implications of research (lines 320-326), and limitations (lines 345-352).

1. ….
2. We understand the concern regarding Canada, and we also believe this is inaccurate, however we are only reporting what we found without trying to discern what part of the variability can be attributed to country alone, we just wanted to have an overview of what is happening in other countries with the published studies.
3. We agree with your comments. We responded this concern above.
4. Finally, we agree with you, our results do not generate strong conclusions, and we added this in the limitations (345-352). However, we believe that we were clear from the begging that this is a systematic review of frequency of incidental diagnosis, so this study add to the literature that we need to stop doing these type of studies (frequency of incidental diagnosis) and focus more on studies that could actually help elicit the incidental diagnosis. Moreover knowing that the ultrasound use is the major contributor of the incidental thyroid cancer, we can emphasize the strategies on this trigger to try to decrease the incidental diagnosis rate, as according to the ATA Guidelines 2015, incidental thyroid cancer detected on histopathologic examination of thyroid tissues removed for MNG or another benign disease does not need more treatment if total or near total thyroidectomy has been performed.

Reviewer: 3

Dear reviewer 3, we carefully read your comments and respectfully respond them:

***The topic is very interesting and clinically relevant. However, there are study design details that need clarification.***

***Abstract methods state "from conception to September 2018" but line 175-176 of the text states that studies included were from between 1991-2015. Please update abstract methods with dates of included studies (this update can be in lines 49-50).***

* We understand this confusion. We would like to clarify that librarian searched all the studies from conception to September 2018***.*** After we reviewed those studies, we found studies published until June 2018 and those studies were conducted between 1991 and 2015. As you recommended, we decided to add the years when the studies were conducted in abstract (line 55).

***Abstract: Define "incidental". As stated later in this reviewer's comments, consider clarifying when it is an "incidental imaging finding" v. "incidental histologic finding". Clinically this difference is very important; especially if there is a future goal of decreasing overdiagnosis.***

* We agree to your comments and we decided to define incidental imaging and histological in the abstract (lines 48-54)

***Line 112: Is it possible that the authors need to search for studies after September 2018? It seems hard to believe that the last study on this topic was in 2015. This has been a hot topic in recent years. The study seems dated if the last study included is from 2015.***

* Because of your comment, we decided to update the search and add the new articles identified to the analysis.

***Line 98: Please give details on the "protocol" that was developed.***

* We agree with your comment, being transparent is a cornerstone in the development of research, as the protocol shows the plan we performed before starting this study. However we do not think it is necessary to give details in the manuscript, but we decided to upload the protocol to Github, with the coding scripts and excel files, thus anyone can review it (line 99).

***Line 134-135: Perhaps i) nodule found related to imaging test unrelated to thyroid and ii) incidental histology finding of cancer should both be "incidental" but should be analyzed separately. These are two different processes. Both "incidental" findings are interesting and warrant studying... but not pooled together.***

* We agree with your comment and we decided to perform an analysis separately and add a figure (figure 5) and its explanation in lines 229-250. We did not removed the analysis pooled together, as the reader can have both versions.

***Line 187-189: The category is "incidental thyroid cancer diagnosis by imaging techniques" but line 189 talks about triggered by histologic exam. I really like this descriptive section but I think it needs a different header (update line 187).***

* We agree with your comments and we decided to modify the header, you can find it in line 208.

***Line 196: Again for "incidentally found": do the authors mean incidentally found with histologic evaluation or with imaging? It is clinically relevant to separate these two types of "incidental" discovery.***

* We agree with your comment, it is clinically relevant to separate these two types of incidental discovery, thus we decided to performed an analysis separately, as stated above. Regarding the “incidentally found” on line 218, we meant the overall incidental thyroid cancer diagnosis (pooled histology and imaging). We will keep the same line as the first version of the manuscript, as the reader can have both versions.