Author’s response to reviews

Title: Elastography for the Diagnosis of High-Suspicion Thyroid Nodules Based on the 2015 American Thyroid Association Guidelines: A Multicenter Study

Authors:

Hairu Li (lihairu0419@126.com)
Yulan Peng (yulanpeng@163.com)
Yan Wang (Msyanwang1998@126.com)
Hong Ai (hong-ai@163.com)
Xiaodong Zhou (zhouxd@fmmu.edu.cn)
Lichun Yang (yn_ylc@163.com)
Kun Yan (ydbz@sina.com)
Ying Xiao (Laoer182002@aliyun.com)
Lisha Liu (gxwgaoxw@sina.com)
Baoming Luo (bmluo2005@126.com)
Qiang Yong (cfp99@163.com)
Shuzhen Cong (shzhcong@163.com)
Shuangquan Jiang (jsq1982@163.com)
Xin Fu (136897780@qq.com)
Buyun Ma (ws_mby@126.com)
Yi Li (13651873180@163.com)
Xixi Zhang (740530962@qq.com)
Xue Gong (gongxuedier@126.com)
Haitao Chen (736842555@qq.com)
Wenying Liu (liuwy23@sina.com)
Ling Tang (45759019@qq.com)
Jin an Zhang (Reviewer 1):

1. Kawk's TI-RADS is the mainstream classification of malignant degree. In the study, the elastograms of the thyroid lesions were separated into 5 types. Is it more accurate to combine the two classification methods?

We thank the reviewer for the excellent suggestion. In fact, it is more accurate to combine the two classification methods. We added the results of the combination to Table 1 and the revised manuscript as follows: As Kawk’s TI-RADS is the mainstream classification of malignant degree, we analyzed the diagnostic performance of Kawk’s TI-RADS. Kwak’s TI-RADS score showed a sensitivity of 62.2%, a specificity of 70.2%, a PPV of 74.0%, an NPV of 57.7% and an AUC under the ROC curve of 0.69 in the overall assessment (Table 1, Figure 3). (Page 6, Lines 26-30)

The combination of the TI-RADS with ES led to a significant increase in the sensitivity and NPV (97.1% and 91.9%, respectively). (Page 7, Lines 17-20)

Another controversial topic is whether the combined application of US and ES may provide better results for thyroid nodule characterization. The results of the present research demonstrated that the sensitivity and NPV dramatically higher with the combination than with the TI-RADS and ES alone, and the accuracy was higher with the combination than with the TI-RADS alone but not with the ES alone, consistent with the study by Cantisani [35]. However, the specificity decreased after the two were combined. (Page 11, Lines 5-11)

2. All the patients gained final pathology results through surgery or FNA. What are the indications for surgery or for FNA, or just randomly allocated?

We thank the reviewer for the detailed review. We added the indications for surgery and FNA in the revised manuscript, which now reads as follows: Fine-needle aspiration (FNA) was performed as recommended by the 2015 ATA guidelines[3]. Patients with suspicious or malignant cytology were referred for surgery. Some of the patients who underwent thyroidectomy did not receive FNA due to certain reasons, such as highly suspicious malignancy or refused FNA. (Page 5, Lines 11-14)
3. Most of the units of measurement in this article are "cm". Page 11 line 32, the unit of measurement is "mm". Page 12 line 11, "cm" should be added at the end of the paper.
We apologize for this mistake. We changed the unit of measure on page 11 line 32 to “cm” (Page 11, Line 20), and “cm” has been added to the end of the paper (Page 12, Line 6).

4. It is suggested that the paper be modified by an English native speaker.
We apologize for the language issue. The revised manuscript has been edited by a native English speaker.

Luca Chiovato (Reviewer 2):
Major concerns:

1. The reference standard should be mentioned in the abstract
The reference standard has been added to the Abstract section as follows: Materials and Methods: Thyroid nodules with high suspicion characteristics based on the 2015 ATA guidelines were subjected to conventional ultrasound (US) and ultrasound strain elastography (USE) examinations at 12 hospitals from 4 geographic regions across China. Cytology/histology of thyroid nodules was used as a reference method. Receiver operating characteristic (ROC) curves were plotted to evaluate the diagnostic performance of the elasticity score (ES) and strain ratio (SR). Logistic regression analysis was used to determine the predictors of malignancy. (Page 1, Lines 14-15)

2. "Background": when describing the malignant risk of high-suspicious nodules, please include references for both ATA guidelines and Kawk's TI-RADS.
References for Kwak’s TI-RADS have been added to the revised manuscript as follows: According to the 2015 ATA guidelines, the malignant risk of high-suspicion thyroid nodules is >70-90%[3]. However, the ATA characteristics of high-suspicion thyroid nodules overlap with degrees 4a to 5 of Kwak’s Thyroid Imaging Reporting and Data System (TI-RADS), the malignancy of which ranges from 3.3% to 87.5%[13]. (Page 3, Line 8)

3. It is not clear how in a multicentric study involving 12 centers, the US and real-time elastographic examinations could be conducted and performed only by two operators. Please, comment.
We apologize for the vague description; there were two operators at each center, and relevant changes have been made in the revised manuscript as follows: Both conventional US and real-time USE were conducted and recorded by two skilled sonographers at each center with at least 5 years of conventional US experience and 1 year of USE experience. (Page 4, Line 11)

4. The Authors should clearly define the inclusion and exclusion criteria.
A detailed description of the inclusion and exclusion criteria has been added as follows: The inclusion criteria were solid hypoechoic nodules with one or more of the following features: irregular margins (specifically defined as infiltrative, microlobulated, or spiculated), microcalcifications, taller than wide, disrupted rim calcifications with a small extrusive hypoechoic soft tissue component, or evidence of extrathyroidal extension (ETE)[3]. The exclusion criteria were as follows: cystic nodules or cystic mixed nodules, nodules with a tumor size >3 cm, nodules with eggshell calcifications, and nodules located in the isthmus. (Page 3, Lines 19-26)

5. When describing the diagnostic performance of the USE (from ES1 to ES5) the Authors describe a statistical significant difference between two groups. Please indicate which groups.
The “two groups” indicate the malignant and benign groups. Relevant changes have been made in the revised manuscript. (Page 7, Line 6)

6. The number and the clinical features of subjects with nodules greater or lower than 1 cm should be included.

We thank the reviewer for the excellent suggestion. The number and clinical features of subjects with nodules ≥ 1 cm or ≤ 1 cm were added to the revised manuscript as follows: There were 877 nodules of < 1 cm and 568 nodules of ≥ 1 cm. The male/female ratios were 180/697 and 124/444, respectively, and the mean age of the patients with nodules < 1 cm was 45.58±11.001 years for those with nodules ≤ 1 cm and 44.01±12.268 years for those with nodules ≥ 1 cm. (Page 8, Lines 4-7)

Minor concerns:

1. Page 11 line 32: please check
   10 mm has been replaced with 1 cm. (Page 11, Line 20)

2. Several studies addressed the issue of the accuracy of elastography. The paper by Magri et al (Endocr Pract. 2015;21) should be considered.

We thank the reviewer for the excellent suggestion. The study by Magri et al has been cited in the revised manuscript, which now reads as follows: However, Magri et al showed that the strain index was significantly higher in malignant thyroid nodules than in benign thyroid nodules and displayed a good diagnostic performance[24]. (Page 10, Lines 11-13)

Ewelina Szczepanek-Parulska (Reviewer 3):

The paper should be corrected by a professional editing service or English native spaker - there are some stylistic and grammar mistakes i.e. p.2 line 23 - indicators/indications, the/which; line 56 - Kwak's/Kawk; line 58 discrepancy/discirpency; p. 4 line 9 - "they shoud were not changed..." - this sentence should be corrected; line 18 - the expression windpipe is rather colloquial - the term trachea should be used instead; line 25 the word classified should be used instead of separated; line 57 - "showed the absence of lesion" - correct the style. In "authors' contributions" the word „drafted" should be used instead of draft

We thank the reviewer for the detailed review, and the aforementioned mistakes have been changed in the revised manuscript.

In figure legends: 1. Nodules instead of nodule and - I have doubts regarding the histopathological term fibroadenoma? Plase check for accuracy the histopathological diagnosis - maybe colloid or follicular adenoma?

We apologize for the incorrect description. The histopathological diagnosis for Figure 1a was adenoma, and relevant changes have been made in the revised manuscript. (Page 18, Line 4)

In the Figure 1 - in the upper panel I would leave only „1903 thyroid focal lesions" without the second part of the sentence

The second part of the sentence in the upper panel of Figure 1 has been deleted.
Authors should also pay attention to the use of the word nodules - as it should be rather reserved for the lesions which are palpable - if the lesions are detected only by the means of ultrasound examination, the terms „focal lesions” or „lesions” should rather be used.

We thank the reviewer for the detailed review and professional advice. The reason we used the word “nodules” was that it is the word used in the 2015 ATA guidelines. In addition, in other papers published in Thyroid, the authors also used the word “nodules” much more than “focal lesions”. I will pay attention to the use of the words “nodules” and “lesions” in the future.

page 6 line 39 - do not use italics when presenting p values
All instances of “p” have been changed to “p”.

page 7 line 2 and several more times throughout the whole manuscript - please present extremely significant p values as p<0,001 not p=0,000
All instances of “p-0,000” have been changed to “p<0.001”.
Authors write that this was a multicenter study - there should be indicated how many and what centers participated. Authors also state that only two examinators performed an analysis - how this was possible in all those centers? The results were recorded and then re-analysed?
We apologize for the vague description; there were two operators at each center, and relevant changes have been made in the revised manuscript, which now reads as follows: Between March 2014 and September 2018, 1819 patients with 1903 suspicious nodules at 12 diagnostic centers from 4 geographic regions across China were evaluated by conventional US and USE (Figure 1). (Page 3 Line 18). Detailed information about each hospital can be found on the title page.

Both conventional US and real-time USE were conducted and recorded by two skilled sonographers at each center with at least 5 years of conventional US experience and 1 year of USE experience. (Page 4, Lines 10-11)

Do I understand correctly that authors included only suspicious or highly suspicious lesions? Authors should specify what was the characteristics of these lesions at the begining of the methods paragraph, when authors first use the term.

We thank the reviewer for the detailed review. The characteristics of the included lesions have been added to the revised manuscript as follows: The inclusion criteria were solid hypoechoic nodules with one or more of the following features: irregular margins (specifically defined as infiltrative, microlobulated, or spiculated), microcalcifications, taller than wide, disrupted rim calcifications with a small extrusive hypoechoic soft tissue component, or evidence of extrathyroidal extension (ETE). (Page 3, Lines 19-24)

The manuscript lacks strong indication about the strenghts of the manuscript and what makes it unique in the background of other similar studies concerning the same issue? Was it the number of the patients or lesions selection?
We thank the reviewer for the excellent suggestion. We emphasized the strengths of the manuscript in the Background section as follows: To our knowledge, this is the first multicenter study in China to investigate the diagnostic performance of strain elastography in differentiating malignant from benign thyroid nodules of high suspicion based on the 2015 ATA guidelines in a large population. (Page 3, Lines 11-14)

Authors may provide an attempt to compare the diagnostic performance of strain ratio sonoelastography and shear wave sonoelastography in the discussion section.
We thank the reviewer for the excellent suggestion. A comparison of the diagnostic performance of strain ratio sonoelastography and shear wave sonoelastography has been added to the Discussion section as follows: Although shear wave elastography (SWE) has drawn more attention in recent years
and has been reported to be more accurate than USE in some studies[30-32]. Tian et al’s and Hu et al’s meta-analysis showed that USE had a better diagnostic performance than SWE, with comparable specificity between methods[33, 34]. Even more recently, a prospective study comprising 243 nodules revealed that USE yielded the highest performance compared with the TI-RADS score and SWE[35]. The aforementioned data confirm that USE is still a promising diagnostic tool for discriminating malignant from benign thyroid nodules. (Page 10, Lines 28-32, Page 11, Lines 1-4).

I would omit the sentence stating in the discussion section that sonoelastography has the power to replace the FNA - it is still rather considered as a helpful option and another diagnostic technique to evaluate preoperatively and non-invasively thyroid lesions. However, still the FNA is a gold standard in the management of thyroid nodules.

We thank the reviewer for the detailed review. The sentence in the Discussion section stating that sonoelastography has the power to replace FNA has been deleted.


We thank the reviewer for the excellent suggestions. The above three studies have been added to the manuscript, and relevant descriptions have been added as follows: Although certain sonographic features, such as taller than wide, blurred margins, irregular borders, internal microcalcifications, hypoechogeticity and marked hypoechogeticity[4], are conventional indicators of malignancy, the sensitivity and specificity vary largely from 38.8% to 90.9% and 53.0% to 96.6%, respectively[5-7]. (Page 2, Line 20)

Elasticity, which is absent in conventional US, is the single feature with the best diagnostic performance[7], as well as a potent predictor of malignant thyroid nodules[21, 22].(Page 9, Lines 22-24).

