

Article information: <http://dx.doi.org/10.21037/gc-20-404>

**REBUTTAL LETTER FOR MANUSCRIPT GS-2019-CATP-12(GS-20-404)**

Dear professor Wei-En Fan,

We are really thankful for the critical assessments from 2 reviewers on our manuscript **GS-2019-CATP-12(GS-20-404)**. We have checked the paper and revised in line with the reviewer's comments. We responded to their concerns point by point. These comments have allowed us to improve our manuscript considerably and we hope the Editors will consider it for publication.

**Reviewer #1**

**This work analyzed Asian studies with the aim of investigating application of TBSRTC in Thyroid cytology practice in Asia. Forty-two Asian studies were included encompassing 84,953 aspirates. The authors reported a low resection rate and higher malignancy rate when compared with Western data. They also found heterogeneity among Asian studies with reference to frequency, resection rate and malignancy risk. The reason for this heterogeneity was suggested to be “differences in countries” and application of molecular testing, but what do the authors mean by “differences in countries” has not been specified.**

**Major comments**

**Comment 1: Results: The authors have shown (in Figure 1) differences in results among South-East Asian and East Asian countries. However, number of studies included from each Asian country has not been specified, along with the distribution of these countries between the two groups (South-East Asian and East Asian). What about other parts of Asia, were there studies from those areas too? What points were considered before subdividing the continent into these zones, as there seems to be an overlap when we use the terms “South-East Asian and East Asian”. What do the authors imply by “differences in countries” (Page 6, last paragraph), The probable reasons for the same also need to be discussed too.**

Reply 1: Firstly, South-East Asia and East Asia are two separate regions in Asia (illustrated figure). In Table 1, we specified the origin of each included study. We added a sentence in the Result (first paragraph) to describe the distribution of Asian regions.



United Nation, Department of Economic and Social Affairs. *World population prospects: the 2012 Revision* [internet]. 2012. Available from: <http://esa.un.org/wpp/>. Archived at: <http://www.webcitation.org/6b47BXGXx>.

As you can see in Fig. 2, there is a significant heterogeneity in RR of TBSRTC V and VI among the Asian countries. To avoid misunderstanding, we modified this phrase as “difference in study origin” and also modified the legend of Fig. 2. We believe that the difference in education influence, working experience, and variations in diagnostic criteria of each individual countries and institutions contribute to this heterogeneity.

**Comment 2: In Page 7, line 4, it has been mentioned that there were significant differences in RR and ROM between studies employing and not employing molecular testing. Although Table 3 does depict increase in RR and ROM in some of the Bethesda categories, but how were the results called “significant”, as this term implies statistically significant increase? Also in line 7, it has been inferred that the increase was mostly seen in the indeterminate categories, although this was not exactly the case for RR.**

Reply 2: We have explained in the footnote of Table 3 that significant results were determined based on the 95% confidence intervals of the two groups. If the 95% confidence intervals do not overlap with each other, it signifies a statistically significant result. Although there is an indirect way to calculate the p-value of the t-test between the two groups, this method likely causes bias because the number of FNAs in each included study are much different and t-test does not consider the weight of each included studies. We have calculated the p-value for each TBSRTC output. However, the p-value did not correctly reflect the true statistical status so we avoided using these values to avoid misleading results.

To avoid misunderstanding, we have rephrased the sentence as “The ROM significantly increased in the benign, AUS/FLUS, FN/SFN, and SM categories following the application of molecular testing”.

**Comment 3: Discussion: I will suggest the authors to discuss the various points in an ordered manner, for eg, on page 8 (line 10), the authors talk about frequency of AUS and FN, then two sentences later they jump to RR and ROM of indeterminate categories and then to ROM of benign category (Page 9, line 1). First frequencies of all categories may be discussed and compared with western meta-analyses, followed by RR and ROM, that too category-wise (category I through VI). This can be followed by discussion of heterogeneity across studies.**

Reply 3: We have rearranged these points in the Discussion logically as your suggestion.

**Comment 4: On Page 8, line 10, the authors mention the lower frequency of AUS/FLUS and FN/SFN found in their meta-analysis when compared with western studies. However, neither any references for the western literature nor explanation for the discrepancy has been provided. Also, it needs to be mentioned that how did the rate of AUS/FLUS compare with TBSRTC recommendations.**

Reply 4: We have added the references for the Western literature and explanation for the discrepancy regarding the frequency of AUS/FLUS and FN/SFN between the two regions in the Discussion (Discussion, second paragraph, Line 2-6).

**Comment 5: Page 8, line 12: The authors have quoted here an article instead of mentioning results of their study. Hence, the authors should also mention their own results while comparing these with that in the study. Also, the reason for the difference in resection rates and ROM has been mentioned as the guidelines followed in Japan for thyroid nodule management. However, there is no mention of the possible reasons for other countries included in the study to have a higher RR and lower ROM (Korean practice has been mentioned later- Page 9, lines 22-27. This can be moved here beside discussing protocol followed in other countries). What about difference between the different zones of Asia and how do they compare with the west?**

Reply 5: We have added the result from our study and the result of the reference study in the Discussion (second paragraph, Line 6-10).

We moved the discussion regarding Korean practice to this part. We did not discuss the practice of other countries here because they usually follow the Western guidelines (e.g., India, Singapore). The practice of these countries was discussed later in the manuscript regarding the heterogeneity discussion.

**Comment 6: Page 8, line 20: In this line, “on the other hand” should be replaced with “additionally” or “Also”. Here, the authors are bringing up the point of inter-observer variation amongst pathologists. A recent study has reported inter and intra-observer variation in interpretation of nuclear features among nine Asian pathologists (Liu Z, Bychkov A, Jung CK, et al. Interobserver and intraobserver variation in the morphological evaluation of noninvasive follicular thyroid neoplasm with papillary-like nuclear features in Asian practice. *Pathol Int.* 2019;69(4): 202-10.), in contrast to excellent inter-observer agreement seen in another study in which pathologists from California, Japan and UK participated (Thompson LDR, Poller DN, Kakudo K, et al. An International Interobserver Variability Reporting of the Nuclear Scoring Criteria to Diagnose Noninvasive Follicular Thyroid Neoplasm with Papillary-Like Nuclear Features: a Validation Study. *Endocr Pathol.* 2018;29(3):242-9.). These should also be referred to.**

Reply 6: We changed “on the other hand” to “additionally” and added these references to our manuscript.

**Comment 7: Page 9, lines 6-8: The phrase “In addition to the possible explanation regarding active surveillance in Asian countries,” should be deleted. The authors here give “Cyst fluid only” aspirates as a probable reason for increase in ROM of category II. However as per TBSRTC 2008, “Cyst fluid only” aspirates have to be categorized non-diagnostic/unsatisfactory. Do the authors have any references to quote where such aspirates have been called benign and turned out to be malignant on resection?**

Reply 7:

We deleted the phrase “In addition to the possible explanation regarding active surveillance in Asian countries”.

Besides, we agree with your point. According to the TBSRTC 2009 and 2017, cyst fluid with or without histiocytes and fewer than six groups of ten benign follicular cells was classified into ND/UNS group.

Contrary to TBSRTC recommendations, aspirates with cyst fluid only were tended to include in “Benign” category in Japan, Thailand, China (Discussion, third paragraph) and VietNam. Takada et al indicated that 14 (2%) were histologically confirmed to be malignant in 715 “cystic fluid only” cases. Therefore, it could be a reason for an increase in ROM of the benign category.

**Comment 8: Page 9, lines 9-10: The authors report low-moderate heterogeneity across studies, in contrast to a high level of heterogeneity published by Vuong HG et al (Vuong HG, Ngo HTT, Bychkov A, Jung CK, Vu TH, Lu KB, et al. Differences in surgical resection rate and risk of malignancy in thyroid cytopathology practice between Western and Asian countries: A systematic review and meta-analysis. Cancer cytopathology. 2019.). Is there any explanation?**

Reply 8: In this study, we performed an analysis for Asian studies only. Conversely, the study by Vuong HG et al. included articles from Asia and Western countries. We think this is the reason of the differences in heterogeneity degree between these two studies.

**Comment 9: Page 9, lines 13-14: Kindly explain what “difference in countries” is being talked about?**

Reply 9: As you can see in Fig. 2, there is a significant variation in RR of TBSRTC V and VI among the Asian countries. To avoid misunderstanding, we modify this phrase as “difference in study origin”. We believe that the difference in education influence, working experience, and variations in diagnostic criteria of each individual countries and institutions contribute to this heterogeneity.

**Comment 10: Page 9, line 17: It appears that Fig. X has been typed by mistake.**

Reply 10: We made a mistake here. It should be Fig. 2.

**Comment 11: Page 9, lines 17-21: How does lack of access to Pathology reports affect the RR? Also, if there is no access to reports, then why were these patients sent to large cities for work-up?**

Reply 11: In some Southeast Asian countries, many small hospitals are not able to perform cytology FNA technique due to lack of formal training and human resources. As a result, patients are usually transferred to large hospitals in big cities for these procedures and then the patient will return to the former hospital for surgical intervention. Electronic medical records are not available between these hospitals which causes a lot of difficulties to follow-up the patient following the FNA procedures. That explains why the resection rate is low in these countries due to missing surgical data. We have rephrased this paragraph to better explain our idea (Discussion, 4<sup>th</sup> paragraph).

**Comment 12: Page 9, line 22: The results of this study document a low heterogeneity across studies included for ROM of the AUS category and <25% I-squared value for other categories. This is dissimilar to the study by Vuong HG, who reported moderate to high degree of heterogeneity in most categories (Vuong HG, Ngo HTT, Bychkov A, Jung CK, Vu TH, Lu KB, et al. Differences in surgical resection rate and risk of malignancy in thyroid cytopathology practice between Western and Asian countries: A systematic review and meta-analysis. Cancer cytopathology. 2019.). Please suggest a reason.**

Reply 12: We have answered in reply 8. The sole reason for this discrepancy is that the present study only focused on Asian studies.

**Comment 13: Page 9, line 27: About JTA guidelines, repetition may be avoided or may be mentioned in short if relevant. Also, note that JTA had already been acronymed in the manuscript.**

Reply 13: We have removed the part regarding the JTA guidelines to avoid redundancy.

**Comment 14: Page 10, line 12: “Otherwise” may be replaced by “In contrast”.**

Reply 14: We changed “Otherwise” to “In contrast”.

**Comment 15: Page 11: The authors have concluded that active surveillance is the management protocol in Asia for indeterminate nodules, and pre-operative molecular testing is useful in thyroid nodule management. Although these are definite findings of their study and need to be mentioned, these have already been well-documented. The authors have not (in this section) answered their query regarding finding out of a reason behind the heterogeneity of results among Asian studies. Also, the heterogeneity seen among different zones of Asia has not been worked up and discussed well.**

Reply 15: We have modified the conclusion of the study and added the answer for the query regarding the reason behind the heterogeneity (Discussion, last paragraph).

**Comment 16: The abstract will also need modifications, following the above mentioned suggestions.**

Reply 16: We have modified the abstract as suggestion.

#### **Minor comments**

**Comment 17: Multiple typographical and grammatical mistakes are there throughout the manuscript, and need to be looked into.**

Reply 17: We have looked into and corrected the typographical and grammatical mistakes.

**Comment 18: References are not as per journal instructions.**

Reply 18: We have modified the reference style according to the journal instructions.

#### **Reviewer #2**

**Comment 1: Please add reference(s) to this sentence “*In addition to the possible explanation regarding active surveillance in Asian countries, another potential reason is that aspirates with “cystic fluid only” are included in this category*” which appeared in paragraph 3 of the discussion part.**

Reply 1: We have added references to the sentence (reference 59, 60, 66-68).