Peer Review File

Article Information: http://dx.doi.org/10.21037/atm-20-3236

Reviewer Comments:

This is an interesting study about a relevant clinical puzzle: How to detect esophageal reflux disease among subjects with chronic cough. I have the following comments and suggestions:

Major comments:
1. The authors think that the combination questionnaire (HARQ + GerdQ) has a better diagnostic value than either alone. However, the authors have not compared the AUC values statistically. This can be done, for example, with the DeLong test. It is possible that there is a statistically significant difference between the combined questionnaire and the HARQ, given the large sample. However, one can still ask whether there is a clinically significant difference: The sensitivity values were 77.19 % for the combination and the same, 77.19 % for HARQ. The specificity values were 79.82 for the combination and 77.06 for HARQ. I think that there is no clinically significant difference. If we ask very long questionnaires, for example in epidemiological studies, the subjects may get tired and do not fill in the questionnaire carefully. Therefore, all questionnaires should be as short as possible.
Response: We have run the DeLong test according to your comment. The result did show that there is a significant difference between the combined questionnaire and the HARQ (P=0.0006) (See below). As for its clinical significance, except for the sensitivity, the specificity, negative predictive value, and positive predictive value all increased by some extent, which indicates advantages of the combined questionnaire over the HARQ. HARQ and GerdQ are common and standard questionairres used in diagnosing GERC, which has been adopoted and confirmed by

many studies (1,2).

Pairwise comparison of ROC curves

HARQ ~ GerdQ	
Difference between areas	0.0330
Standard Error [a]	0.0336
95% Confidence Interval	-0.0329 to 0.0989
z statistic	0.980
Significance level	P = 0.3269
HARQ ~ United	
Difference between areas	0.0521
Standard Error [a]	0.0151
95% Confidence Interval	0.0225 to 0.0818
z statistic	3.447
Significance level	P = 0.0006
GerdQ ~ United	
Difference between areas	0.0851
Standard Error [a]	0.0223
95% Confidence Interval	0.0415 to 0.129
z statistic	3.822
Significance level	P = 0.0001

[a] DeLong et al., 1988

Changes in the text: We added the results of Delong's test (Page 13, line 397)

2. The author's way to manage the GERC differs from international standards. The authors rely on medicines but do not mention the non-pharmacological interventions for GERD: diet modification to promote weight loss in overweight or obese patients; head of bed elevation; avoiding meals within 3 hours of bedtime. These interventions have larger effect than the drugs. Thus, there may have been real GERC patients among those judged to have non-reflux cough on the basis of insufficient response to GERD therapy.

Response: Thanks for your comment. When managing GERC, we followed the 2015 Guideline for the Diagnosis and Treatment of Cough (3). There are indeed some non-pharmacological interventions for GERD, while they were not mentioned in the latest Lyon Consensus (4) in diagnosing GERD as well as other diagnosing guidelines. Thus, we did not mention them in our manuscript. However, it is very important to consider those non-pharmacological interventions when diagnosing GERD. We would consider these non-pharmacological interventions in our following studies.

3. One of the three criteria to define GERC was that cough was responsive to the anti-reflux therapy. How was the response measured? The response should be measured utilizing either a validated cough quality of life questionnaire (Leicester Cough Questionnaire for example) or objective cough counting.

Response: The response indicates that cough symptom score decreased by >50%. We are sorry that we might delete some criteria by mistake when revising the manuscript. We have added the cough symptom score decreased by >50% in the methods.

Changes in the text: Page 6, line 163

Minor comments:
1. Please give the ROC as well as the sensitivity and specificity values of GerdQ in the abstract
Response: ROC, sensitivity and specificity were added in the abstract.
Changes in the text: Page 3, line 58-59

2. HARQ includes questions about post-nasal drip and wheezing, alongside with the reflux questions. Is it problematic? GerdQ is more reflux-specific. Therefore, it is a little bit surprising that the AUC of HARQ was better than that of GerdQ to diagnose GERC.
Response: As we mentioned in the discussion section: GERC includes acid and non-acid GERC. some GERC patients may not have symptoms such as acid refluxes and heartburn. The GerdQ score in patients with non-acid GERC was significantly lower than that in acid GERC patients. The GerdQ score only has advantages in distinguishing acid and non-acid GERC. Compare with GerdQ, HARQ adds the evaluation of chronic cough. Thus, HARQ has an assessment of both chronic cough and gastroesophageal reflux symptoms. However, the results of DeLong test showed that there was no significant difference between AUC of HARQ and GerdQ (P=0.3269).

3. Please move the sentence 'This was a prospective clinical trial' from the chapter 'Patients' to the chapter 'Study design'.
Response: It has been moved to the Study design section.
Changes in the text: Page 6, line 176

4. Was this really a prospective study given the fact that not all patients underwent all procedures (like the histamine challenge)?
Response: Yes, according to the diagnostic the 2015 Guideline for the Diagnosis and Treatment of Cough, not all patients need to be confirmed the diagnosis with all the diagnostic procedures (like the histamine challenge).

5. Lines 77-84: In the list of inclusion criteria there are also exclusion criteria (no rales, no x-ray abnormalities for example). Please list them in the following sentence, which lists the exclusion criteria.
Response: We have moved them to the exclusion criteria.
Changes in the text: Page 5, line 139-142

6. Line 79. What do you precisely mean by etc? ("no symptoms like wheezing, hemoptysis, or fever etc.;.")
Response: We have revised the sentence. patients with any symptom like wheezing, hemoptysis, or fever should be excluded.
Changes in the text: Page 5, line 139-140

7. Line 96. Please open the abbreviation MII-pH when you use it for the first time.
Response: Yes, the multichannel intraluminal esophageal impedance and pH monitoring (MII-pH) is defined in the 1st paragraph of Introduction.
Changes in the text: Page 4, line 91

8. Line 100. What is cough variant syndrome?
Response: It should be cough variant asthma. It has been revised.
Changes in the text: Page 7, line 208-209

9. Line 103. Please open the term 'combination therapy' when you use it for the first time.
Response: We rearranged the methods section. The paragraph of Diagnostic criteria of GERC was moved to the front of the paragraph of Study Design. And we deleted the above mentioned sentence.
Changes in the text: Page 6-7, line 157-216

10. Statistical analysis: How did you deal the problem of multiple comparisons when comparing paired groups after ANOVA? (Bonferroni correction or something like that)
Response: ANOVA was conducted by using SPSS software, and we chose both LSD and Bonferroni adjustment in post host test, which showed the same results.

11. The paragraph on page 10, lines 221 – 238: You could shorten this paragraph a lot, too detailed information.
Response: We have shortened the paragraph.
Changes in the text: Page 11, line 315-324

12. Results, lines 261 – 268: The comparison of acid GERC and non-acid GERC is already written in the previous paragraphs. Please do not write the same results twice.
Response: It has been deleted.
Changes in the text: Page 12, line 365

13. Lines 273 - 288 (end of the results section). The text only refers to figures or tables, but does not summarize the main findings in the figures or tables. Please add.
Response: We have added the summary of the main findings.
Changes in the text: Page 13, line 379-399