Author’s response to reviews

Title: Quality analysis of smart phone sleep apps in China: Can apps be used to conveniently screen for obstructive sleep apnea at home?

Authors:

Zhaofeng Xu (474953549@qq.com)
Xin Luo (luoxingz@126.com)
Jianbo Shi (tsjbent@163.com)
Yin-yan Lai (Laiyy3@mail.sysu.edu.cn)

Version: 1 Date: 27 Jul 2019

Author’s response to reviews:

Dr Yinyan Lai

July 21, 2019

Dear Editors and Reviewers:

Thank you for your letter and for the reviewers’ comments concerning our manuscript entitled “Quality analysis of smart phone sleep apps in China: Can apps be used to conveniently screen for obstructive sleep apnea at home?” (ID: MIDM-D-19-00187). Those comments are all very valuable and very helpful for revising and improving our paper as well as the important guiding significance to our researchers. We have studied comments carefully and have made correction which we hope meet with approval. Revised portion are marked in yellow in the paper. The main corrections in the paper and the responds to reviewers’ comments are listed below.

Reviewer 1 Hyo Soung Cha

1) Response to comment “Line 153 evaluation procedure, what are the criteria for allocated points for scientific basis (16 items, 32 points), functionality and usability (11 items, 12 points), and accountability (9 items, 9 points)”
Response: We are very sorry that we didn’t make a clear declaration of our evaluation criteria. There are thousands of health apps in app stores. One of the purposes of health apps is to analyze sleep quality and even screening people for OSA. However, there are not any evaluation system to assess the quality of these apps until now. In previous studies which evaluated the quality of other health apps[1, 2], they developed quality assessment tools which including criteria in accountability, functionality, scientific basis and so on. So, in order to evaluate sleep apps in China we developed this evaluation scale. In Scientific basis part, according to the diagnosis criteria of OSA we designed 16 items. Since some apps with part of information related to one of these items, we coded 0 indicating “no information provided”, 1 indicating “part of the information provided”, 2 indicating “correct and complete information”. In functionality domain, when app with one of the functions described in evaluation scale, we will add one point for it.

When an apps could only be used with other devices, the usability of this app was low. So, in usability domain, 0 indicated “the app must connect to other equipment to work properly”, 1 indicated” the app can work independently”, and 2 indicated “the app can work independently and connect to other equipment”.

Silberg scale was wildly used to assess the accountability of health information, so we used silberg scale to evaluate the accountability of these sleep apps. There were 9 items in silberg scale, 1 point was awarded for the presence of each of the items, and the maximum score was 9 points.

2)Response to comment” It is uncertain whether a user's result of analysis of PSG information through apps accurately diagnose with OSA. As this paper is to evaluate the quality of apps for analyzing the information for screening OSA with convenience, it also should evaluate whether the result of OSA diagnosis with selected apps can be agreed with doctors' diagnosis. We suggest that the author should describe the accuracy of screening tests with apps, including the use of smart wearable devices, in relation to the sensitivity and specificity analysis”

Response: It’s really true as reviewer suggested that the sensitivity, specificity and accuracy are very important in analyzing the quality of sleep apps.

There were several studies focused on the sensitivity and specificity analysis of sleep apps or to monitor sleep using smartphone technologies. Hiroshi Nakano ect used a smartphone to collect the snoring sound during sleep time and compared it with the snoring time measure by PSG. They recruited 50 patients to join in the studies. The relevant index between respiratory disturbance index measured by smartphone and the apnea-hypopnea index measured by PSG was 0.94. The result proved that a smartphone can be used for monitoring snoring and OSA[3]. Another example was the ApneaApp, which was designed by Computer Science and Engineering, University of Washington. They used smartphone to detect breathing movement during sleep. ApneaApp was validated in a study with 37 volunteers[4, 5]. These apps or
smartphone technologies were designed by scientists but most of sleep apps were consumer sleep apps which were designed by companies. Those consumer sleep apps renewed frequently and might not be validated. There were no criteria to evaluated these consumer sleep apps. Some studies focused on the consumer sleep apps such as Sleep Cycle. The result showed that the sensitivity of Sleep Cycle was high but the specificity was low. The results reported by Sleep Cycle correlated poorly with PSG[6]. Other sleep apps such as Sleep Time was also compared with PSG, the sensitivity was 90% and the specificity was 50%[7]. Some apps used actigraph or wearable devices to collect information. Some studies compared the actigraph and PSG, the sensitivity was 96.5% but the specificity was 32.9%[8]. According to our studies, the Sleep Cycle (Rank No.5) and Sleep Time (Rank No.8) were top 10 apps among all 127 including apps. Because of the high sensitivity, with the help of these apps, people with snoring problem or sleep disorder might go to hospital for professionally advices given by sleep specialists and do further test such as PSG to make the specific diagnosis.

For sleep specialists, they should evaluate and verified those sleep apps and help more patients in the future. However, it was impossible to evaluated all sleep apps which provided by app store. In our study, we firstly downloaded all available apps and try them while processing evaluation procedure. For screening OSA, doctors need information about sleep structure, sleep stages, oxygen saturation, respiratory events (such as apnea, hypopnea ect.) and cardiac events. These items are including in our evaluation scale. While evaluating these apps, we would analyze whether the apps could provide the information about these items or not.

Therefore, one of the aims of our study was to create an evaluation scale which could easily be used by sleep specialists to quickly evaluate all available sleep apps and the sleep apps which would be designed in the future. That was the reason that we didn’t do the specificity and sensitivity in this study but as the reviewer suggested that the specificity, sensitivity and accuracy are necessary in evaluate the quality of these apps, we are now processing the comparison of these selected apps with standard PSG and we will show our result in the coming future.

3) Response to comment” The author should illustrate the previous studies that have done regarding this study's topic”

Response: We have rewritten the functionality and usability in Method part according to reviewers’ comments. Line 131-137 “Adrian A. Ong ect evaluated functionality of 51 selected sleep apps in 2016, they found that except analyzing sleep structure, some apps also could be used as movement trackers, sound recorders, smart alarm and so on. Other studies also described these additional functions of sleep apps. Users may download these apps because of their additional function. All apps need sensors to collect information during sleep times. Some apps
used accelerometer in a mobile phone while others require other wearable devices or sensors to monitor sleep” was added.

4) Response to comment “ denote abbreviation information for OSAHS ”
Response: We are sorry for our incorrect writing. Line 125 we have changed “OSAHS” into “OSA”.

5) Response to comment “denote reference in line 126 that is "we used functionality evaluation criteria reported in previous studies to appraise the functionality of several apps, and then adjusted the criteria according to the actual situation."”
Response: It is our negligence and we have corrected it. In Line 140 the references of “Previous studies” was added.

Reviewer 2 Adam Dunn

1)Response to comment “Abstract: I would have liked to have seen more detail in the methods to make it clearer how the score was defined. The conclusions could have been shorter to make room for longer methods.”
Response: We have rewritten this part according to the Reviewer’s suggestion. We are very sorry for our unclear report in abstract. We described more detail in methods to make it clear. Line 32-40 was rewritten.

2)Response to comment “If the range of diagnostic accuracies are not reported in the abstract, then "promising" may be a step too far.”
Response: We have rewritten the conclusion part according to the suggestion. Line 48-51 was rewritten.

3)Response to comment” Table 1. There is definitely not enough information in Table 1 and lines 117 to 122 to understand how the score was determined for the scientific basis. Just because an app mentions each of these items does not mean that it is evidence-based, does it? I don't
understand how an app can address each of the items that are related to a standard PSG, or replace one or several of them with proxies that might have been validated and found to be closely correlated in a general population. I would imagine that most apps would ask people questions or perhaps take measurements from external devices. Each of those would need to be tested and validated to check that the "diagnosis" or "conclusion" they produce from the app (whether that is to tell a consumer to go to the doctor or to help a patient monitor their diagnosed sleep apnea). Presumably the evidence base to support the app would come from peer reviewed research that compares the app against a PSG (for diagnosis/screening)?

Response: It’s really true as Reviewer suggest that the diagnosis or conclusion of these apps were not verified or tested. Because according to previous studies about sleep apps or smartphone technologies, some researchers compared the sensitivity, accuracy and specificity of some sleep apps with standard PSG. The correlation varies from apps to apps. Apps such as ApneaApp which designed by scientists was carefully verified, but ApneaAPP was not available in app store until now. Most of sleep apps were consume sleep apps which meant the main purpose of these apps were to earn money. These apps designed by companies and they were not verified by sleep specialists. Such as Sleep Cycle or Sleep Time, the sensitivity of these apps were high, but the accuracy and the specificity was quit low[9].

There are thousands of sleep apps in app store and these apps renewed frequently. Our sleep specialists should evaluate sleep apps to help more people, but there were no evaluated criteria to analyze these apps and it was almost impossible to verified every apps. Therefore, our study developed an evaluation scale which could evaluate sleep apps quickly. PSG was the golden standard to diagnose PSG, so, to analyze the quality of these sleep apps, we should compare them with PSG. Thence, we used the items of PSG and the diagnosis criteria of OSA (16 items in total) to evaluate these apps.

As the Reviewer suggested that, we are now comparing these selected apps with PSG. This study actually the first step of our plan to develop a practical app to help people screen OSA or monitor sleep. We firstly evaluated the sleep apps, found out the current situation of sleep apps in China, then we would analyze these apps further. Finally, we may try to develop an idea apps to screen OSA in China. Some apps had already been compared with PSG, we are now working on other apps, and as the suggestion given by Reviewers, we would publish our results in the future.

4)Response to comment “Evaluation Procedure: It appears that the scores were done by one person and then "checked" (but perhaps not independently). Because they weren't done independently there is no inter-rater reliability. Once the app ratings have been done independently, we would need to know the inter-rater reliability on the coding to determine whether the evaluation method was reliable and robust.”
Response: It's very true as the reviewer's suggestion that the reliability is very important for an evaluation scale. The inter-rater reliability was added in line 290-295.

5) Response to comment: “Discussion: I wonder if the apps are not intended to be used as screening but may have different aims (i.e. to make money through advertising by providing something people like to use on an ongoing basis). Apps that are designed for screening may only need to be used once, whereas apps used to support patients after diagnosis will be used in an ongoing fashion. Perhaps more could be said about this as "missed opportunities" or added value.”

Response: We have corrected this part according to reviewer's suggestion. We have rewritten this part. Line 373-379 was added.

We tried our best to improve the manuscript and made some changes in the manuscript. These changes will not influence the content and framework of the paper. We list the changes and marked in yellow in revised paper.

We appreciate for Editors/Reviewers’ warm work earnestly, and hope that the correction will meet with approval.

Once again, thank you very much for your comments and suggestions.

Your sincerely,

Yinyan Lai

Reference


