Author’s response to reviews

Title: Quick assessment with controlled attenuation parameter for hepatic steatosis in children based on MRI-PDFF as the gold standard

Authors:
Jaeseung Shin (drshinjs@yuhs.ac)
Myung-Joon Kim (mjkim@yuhs.ac)
Hyun Joo Shin (lamer-22@yuhs.ac)
Haesung Yoon (saydivine@yuhs.ac)
Seung Kim (pedks@yuhs.ac)
Hong Koh (khong@yuhs.ac)
Mi-Jung Lee (mjl1213@yuhs.ac)

Version: 1 Date: 06 Feb 2019

Author’s response to reviews:

Point-by-point response to the reviewer’s comments

Dear Editor,

We truly appreciate the Reviewers and the Editorial Board for their careful critique and helpful comments on our manuscript. We revised the manuscript in accordance with the suggestions and comments of the Reviewers. Detailed responses to each of the comments are as follows:

David Petroff (Reviewer 1)

Background

R1-1. Line 63: I do not believe that citation [1] supports the claim in the first sentence. Perhaps "NAFLD is the most prevalent liver disease in children" would be more appropriate.

We agree with this comment and changed the sentence as suggested.
R1-2. Lines 70-72: The authors mention that PDFF correlates "well" with histologic grades and cite [4, 5]. In citation [4] regarding adults, Fig. 1 shows strong overlap in PDFF signals between S1, S2 and S3. In the paediatric context, Fig. 2 in [5] looks somewhat better, but overlap is still common. The statement could be toned down.

\ We agree with the reviewer’s comment and modified the sentence as “MRI-estimated liver proton density fat fraction (PDFF) has shown a good correlation with histologic steatosis grade and the potential of clinical utility for the evaluation of NAFLD in both adults and children”. Thank you.

Method

R1-3. Line 95: It is not clear to me if all patients with attempted or with successful MRI-PDFF and CAP were included. It would be useful to know if/when one of the techniques was invalid. It would also be helpful to know what patient criteria led to use of both techniques in this retrospective context - there may be a selection bias.

\ For pediatric patients who are highly suspected of having fatty liver, we conduct a test of both abdominal MRI including PDFF and TE with CAP for diagnosis. Most of the intervals between these examinations are less than one month, but the results were excluded as they were more than one month depending on the patient's condition or the examination schedule. However, the decision of doing the exams was made by the pediatricians and there may be a selection bias in this process, so we described it as a limitation in the Discussion. Thank you.

R1-4. Line 101: Please add "…using the age and sex dependent 95th percentile…”

\ We added the words as suggested.

R1-5. Post-hoc tests were probably used in the analyses associated with Figure 2A (Lines 178-180). The authors could state which they were. On a related note, the ROC curve comparisons (Lines 189-191) should be corrected for multiple testing, e.g. with a Bonferroni-Holm procedure.

\ We agreed with your point on statistical analysis. According to the advice of statistician, the method section was modified. We also corrected the p-values for the ROC curve comparisons.
R1-6. The authors should add confidence intervals for the sensitivities and specificities at the Youden-optimized point and, ideally, for the value of the optimized point (241 dB/m) itself. This is an important point I return to below in point 10. There are a number of techniques for doing the latter (e.g. Bantis, Nakas, Reiser, 2018, Construction of confidence intervals for the maximum of the Youden index and the corresponding cut-off point of a continuous biomarker, Biometrical Journal) and the authors may wish to contact a statistician.

We added confidence intervals for the sensitivities and specificities at the Youden-optimized point. On top of your suggested method, statisticians at our institution recommended bootstrapping technique to obtain the confidence interval of the Youden-optimized point. Unfortunately, however, due to limited time for response, we could not get the confidence interval with either way.

R1-7. Some blood parameters, most notably ALT and AST, should be treated on a logarithmic scale. Large differences between mean and median values or large SD compared to the mean of a positive variable can indicate the need for this transformation (see also point 8).

We agree with your opinion. We performed t-test with logarithmic scale transformation and the logAST and logALT also showed significant difference. We remained the original values on the Table 1 because logarithmic transformed values are difficult to interpret clinical meaning.

Results

R1-8. Lines 162-164 and Table 1: These results may change qualitatively after considering a logarithmic scale. Moreover, age should be included as a covariate when comparing blood parameters between the groups in Table 1. This may also affect the Discussion (lines 267-269).

We modified the method as your suggestion and the advice of our statistician. When comparing blood parameters in Table 1, the analysis of covariance was applied with sex and age as covariate and the subsequent results were obtained.

R1-9. Lines 166-167: The null-hypothesis that the correlation is zero is not meaningful here, hence the p-value is not meaningful and could be removed. A confidence interval would be informative however.

We removed the p-value. Instead, a confidence interval for correlation coefficient was provided in the paragraph.
R1-10. Lines 181-189: Here is where the confidence intervals are essential. There are only 10 patients with S0, meaning that estimates cannot be accurate. The "optimal" value of 241 dB/m will be very uncertain as a confidence interval will show. It is essential to understand this point for the discussion.

\[ We agree with uncertain optimal value and necessity of confidence intervals regarding small number of patients in S0. We presented confidence intervals for the sensitivities and specificities at the Youden-optimized point in the manuscript and in Table 3. According to the statistician's opinion at our institution, this information is helpful to understand the uncertainty of optimal value and we additionally pointed this out in limitation. \]

R1-11. Lines 206-208: A formal statistical test with n=4 vs n=13 is extremely underpowered and should not be performed. Only median values should be provided. This then will change the statement completely. The difference between 326 dB/m (M probe) and 370 dB/m (XL probe) is not so small. It is incorrect to say they "were not different", even if the difference cannot be assessed so easily.

\[ We agree with reviewer's comment regarding underpowered statistical analysis. Only median values are provided in the Result section and the sentence about the difference between M and XL probes was removed. \]

Discussion

R1-12. Lines 239-240: Please tone down the statement "...demonstrated the ability...to...differentiate between histopathologic grades", since this was not very successful in citation [12], which had a small number of patients with steatosis and fairly large overlap (see Figure 1 in that paper).

\[ We agreed with reviewer’s comment regarding previous study using CAP in pediatric patient and we toned the statement down. \]

R1-13. Lines 270-277: The authors could add that histology measures percentage surface area covered by fat cells, MRI-PDFF measures a proportion of fat molecules and CAP measures physical properties of the liver. These are essentially different (see e.g. your citation [11]). Particularly in children, this could explain the expected correlation between CAP and AWT, but surprising lack of correlation between PDFF and AWT.
Thank you for your comment. We added the difference between the two modalities as suggested.

R1-14. Lines 284-294: The authors should point out explicitly that n=10 for S0 is a limitation and state that the estimate for the optimal cut-point is uncertain. They should also acknowledge more clearly the uncertainty in the gold-standard (lines 288-289). Again I refer to Fig. 1 in Permutt et al., i.e. reference [4].

We agreed with reviewer’s comment regarding the limitation of our study. We pointed out small number of normal steatosis group and lack of histopathologic grade clearly in the limitation section.

Minor language issues

R1-15. The level of English is excellent. Here are a few minor suggestions for improvement:

a. Lines 42-43: "...(NAFLD) who were assessed for PDFF and CAP…"

b. Line 49: Delete "For steatosis grades"

c. Line 77: add "s" to "clinics due to…"

d. Line 78: add "n" in "is an ultrasound-based…"

e. Line 81: "TE, shows good correlation…"

f. Lines 88-89: "based on PDFF with subgroup analyses based on body mass index (BMI)."

g. Line 224: "but is probably limited during…high BMI, though longitudinal data are lacking."

h. Line 251: "The portion examined with the XL probe…"

Thank you for your kind comment about minor language issues. All these issues are revised.
Oyekoya Taiwo Ayonrinde, MBBS, PhD (Reviewer 2):

R2-1. The introduction should include a summary of why steatosis assessment is important in children e.g. with rising obesity rates and consequently NAFLD rates in children and adolescent populations there is increased risk of NASH, cirrhosis and metabolic syndrome etc at young ages. What is the prevalence of obesity and NAFLD in the general population of children/adolescents in the region studied?

\ We agreed with reviewer’s comment regarding the importance of steatosis assessment in children. We pointed out rising rate of obesity and NAFLD worldwide, including South Korea.

\ In Korean children and adolescents, the prevalence of obesity increased from 6.8% in 1998 to 10.0% in 2013 [3]. The prevalence of NAFLD in Korean adolescents in the general population is unclear due to a lack of validated Korean-specific noninvasive diagnostic methods. In one study based on ALT level [4], prevalence of NAFLD among Korean adolescents were 5.9% in 2015 by an ALT level > 30 U/L for boys and > 19 U/L for girls, and 7.1% in 2015 by an ALT > 25.8 U/L for boys and > 22.1 U/L for girls. Additionally, the prevalence seemed to increase, especially for girls, since 2010 to 2015. However, since the upper normal values of ALT for Korean adolescents were unavailable, the study did not confirm whether the prevalence of NAFLD in Korean adolescents had increased or not.

R2-2. Page 3 line 46 - I'm not sure it is valid to use BMI30 rather than existing age-adjusted BMI.

\ We wanted to evaluate the effect of BMI for the measurements of hepatic steatosis using two different methods. However, as mentioned as a limitation in the Discussion, because there were no agreed criteria for severely obese patients in pediatric field, a BMI30 cutoff was applied. Nevertheless, the proportion of the BMI30 group was still low, so a further study to focus on severely obese patients is required. Thank you for your comment.

R2-3. Page 9 line 186 - the stated optimal CAP cutoff values of 299 and 303 are similar. Suggest adjusting sensitivity and specificity to reset the cutoff.

\ These results were the optimal cutoff values with the best performance. This result shows the limitation of CAP to discriminate low grade and high grade steatosis as presented in figure 2A.

R2-4. Please report positive predictive values and negative predictive values.
We added positive predictive values and negative predictive values in the Result section and Table 2.

R2-5. Page 13 - comment about utility of MR elastography and recent published cutoffs for elastography (Bazerbachi F, et al). Range of Normal Liver Stiffness and Factors Associated With Increased Stiffness Measurements in Apparently Healthy Individuals. Clin Gastroenterol Hepatol. 2019 Jan;17(1):54-64.e1.). Also, are you able to include liver stiffness measurement (LSM) results? Were there any children/adolescents with suspected NASH?

In relation to the LSM you mentioned, we added the content and the recommended reference in the Discussion.

In this study, we only wanted to check the hepatic steatosis and excluded the elasticity of the liver from the study as it may vary depending on various effects. The diagnosis of NASH requires pathological result but is not included in this study.

R2-6. A new table similar to Table 1 but comparing BMI, AWT, serum transaminases, lipids, insulin and glucose in those with raised CAP compared with those with normal range CAP, and also in those with different raised and normal PDFF categories would increase clinical appeal and interpretation of the significance of the results.

Table 2 for clinical characteristics according to steatosis group based on MR PDFF is added on manuscript. We agreed with your opinion that this table would increase clinical appeal. Because this research is aimed to find CAP optimal point based on PDFF, we presented comparison among groups based on PDFF rather than CAP.

R2-7. It may not be appropriate to analyze the whole cohort together, as adipose distribution in males and females is different, with girls having markedly higher subcutaneous thickness (including abdominal wall) than males do (see Ayonrinde OT, et al. Childhood adiposity trajectories and risk of nonalcoholic fatty liver disease in adolescents. J Gastroenterol Hepatol. 2015;30(1):163-171; and Ayonrinde OT, et al. Gender-specific differences in adipose distribution and adipocytokines influence adolescent nonalcoholic fatty liver disease. Hepatology. 2011;53(3):800-809).

In the present study, male (2.61 ± 0.63 cm) had significantly higher abdominal wall thickness (AWT) than female did (2.39 ± 0.69; p = 0.045), whereas previous studies showed markedly higher subcutaneous thickness (including abdominal wall) in female than in male childhood [reference 32 and 33 in manuscript]. Possible explanation for this discrepancy is that higher
prevalence of obesity and metabolic syndrome, including waist-to-height ratio, of boys than of girls in Korea [reference 3 and 4 in manuscript]. Likewise, this study showed a tendency for higher ratio of male in obesity group (77.4% in obese group vs. 64.6% in non-obese group, p = 0.168). The small number of female patients (n=24, 31.3%) in this study might also affect this result. We added the sentences about gender difference in manuscript.

R2-8. Figure 1 doesn't add much to known information and can be removed.
\ We deleted Figure 1.

R2-9. I suggest splitting Figure 2A and 2B by gender.

GENDER DIFFERENCES ARE ESSENTIAL TO INTERPRETATIONS IN THIS STUDY and would enrich it.
\ We agree with your comment that gender differences might affect the result of the present study. Unfortunately, separate analysis by gender is not performed due to small number of female patients (n=24, 31.3%). Clinically, the low proportion of female patients might be attributed to higher rate of obesity and metabolic syndrome in Korean boys than in girls. Instead, we included gender as covariate of our analysis. We added the sentences about gender difference and limitation of our analysis by gender in manuscript.

\ Figure 1A (Figure 2A in the previous manuscript) is split by gender on Figure 1B, but Figure 1C (Figure 2B in the previous manuscript) is not able to be split because separate ROC analysis is not appropriate with limited number of girls (n=24).

Again, we assure that this article has not been published elsewhere and is not being processed elsewhere.

Thank you for your consideration of our manuscript again.

Best regards,

Mi-Jung Lee