1. Was the referral pattern to the liver clinic in this study routine for all obese individuals, or was there referral bias (physician pre-selection)? Please specify.

R: All obese individuals referred to the liver clinic were considered eligible but only those with liver biopsy were enrolled in the study. The main aim of the study was to compare MR imaging techniques (MRS and MR-PDFF with histology

2. The spectroscopy methods are well described. Where were SAT and VAT obtained? (What lumbar level?) Were repeated studies done on the same subjects to assess reliability (ICC) in the same scan? Were studies of heterogeneity across the liver assessed for MRS versus PDFF?

R:SAT and VAT were calculated from 5 images extending from 5 cm below L4-L5 to 15 cm above L4-L5 (Quantitative Analysis last paragraph). Comparison between MRS and PDFF was conducted only in liver segment VI-VII firstly because it was the site of liver biopsy and because MRS suffers of motion artifacts and heart biting in the anterior segments (II, IV and VIII).

3. What is the advantage of using a multi-echo sequence versus a two-point Dixon sequence in the liver? (Henninger et al. Eur Radiol 2015; 25: 1356-65). The authors should justify this point.

R: As reported by the authors themselves, the method is more accurate.

4. Was breath holding employed for MRS?

R: Yes it was. It has been added in the text

5. What ROC optimal criterion was used to determine the ROC optimal cut point for diagnosis of NASH? The statistical methods requires more elaboration.

R: The ROC cut point is based on the best specificity and sensitivity achievable.

6. PDFF seems less well correlated with histology relative to MRS. Please explain more clearly what "correlation, concordance coefficient and accuracy" represent (in more colloquial terms for the reader). Why might this be? Does this impact the use of PDFF (arguably a much easier, widely accessible technique than STEAM MRS sequences). In the discussion, the authors suggest that "our data demonstrated optimal correlation between histology, with a faint concordance"-- this sentence does not make sense, and does not seem supported by the correlation data presented in Results. Please verify, reword, and confirm.

R: Statistical Analysis was simplified and clarified. The Lin test was changed with the Pearson test, and Bland -Alman plots were introduced.

7. The Discussion needs to be much more focused. Having performed these imaging analyses before, I am convinced that PDFF is easier (in fact some platforms, e.g., Siemens, have a WIP sequence that uses multi-echo acquisitions to calculate fat% in the liver in a single breath hold). However, the discussion needs to focus on how these results are novel or contributory nature - the last sentence of the limitations ("histology reveals...molecule") appears quite important, and needs to be expanded. What in the authors' opinion is the optimal measure of hepatic steatosis?

R: Conclusion of the manuscript was edited and the authors opinion was better explained.

8. Please simplify table 1 dramatically. The groups should be listed next to each other.

R: Table was modified and simplified according to reviewer's comment

9. With regard to Figure 4, was there bias with one technique versus other? Would include Bland-Altman plots here.

R: As reported in limits, it seems that MR-PDFF has some difficulties in the differentiation between moderate to high steatosis (seed discussion). Figure of Bland-Altman plot was introduced.