

**Title:** Comparison of four NAFLD detection scores in a Caucasian population

**Authors:** Lars Lind, Lars Johansson, Håkan Ahlström, Jan W Eriksson, Anders Larsson, Ulf Riserus, Joel Kullberg and Jan Oscarsson

**Reviewer #1**

**Comments:**

1. This study on two different groups from Sweden regarding metabolic risk for NAFLD was a sectional study that aimed to investigate the performance of 4 different scores for the diagnosis of NAFLD.

As expected the authors find a high prevalence of NAFLD in the high-risk population (74%) and 23% prevalence in the low risk population which is in accordance with the literature. The authors could have discussed these findings in the discussion session.

**Ans:** We thank the reviewer for pointing out this miss. We have now added the following sentence referring to other studies on prevalence of NAFLD in the Discussion section (page 11, paragraph 2, lines 5-6): "The prevalence of the NAFLD scores found in this study were similar to those found in other population-based studies (1-4, 9) and in high-risk groups, such as diabetes (15, 16)."

2. The original finding was that these scores had not been evaluated on a Nordic population with different risk factors regarding NAFLD. Although the study is well performed and organized, the results are not exciting since the authors found that the performance of the scores was at most of an AUROC of 0.80 (NAFLD LFS in the high-risk population, which it would be better to have a more accurate tool to identify NAFLD). The 95% confidence interval of the AUROCS of the scores were not presented and these should be included.

**Ans:** We thank the reviewer for raising this key point. We agree that it would have been more exciting to obtain higher AUCs, but the AUCs in the present study are in the same order as those reported in other populations, as we have mentioned in the discussion. We have now added the 95% CI limits for AUC to Table 2 (page 25) and Table 3 (page 26).

3. The strength of the study is to propose the use of a score that would be simple enough to be used as a first tool in the evaluation of NAFLD. This might be achieved in the low risk population, however, intake high risk population the NAFL-LFS is not simple enough and require many variables to be used in general clinical practice.

**Ans:** We thank the reviewer for the comment. It is true that LFS requires many variables to be calculated. That is why, in this study, we have tried to simplify LFS and showed that this was possible in the high-risk setting. We have added the following sentence to the Discussion section (pages 12-13, paragraph 4, lines 5-8) to expand on this information further:

“However, if this finding of simplified LFS score could be reproduced by others in a high-risk group, the use of this simplified LFS score could be an attractive tool in the clinical setting for screening of NAFLD in high-risk individuals.”

## **Reviewer #2**

### **Comments:**

1. In this study of "Comparison of four NAFLD detection scores in a Caucasian population" by Linds et al explored the extent to which the different scores can predict NAFLD in a high-risk individual versus a non-selected individual. This study was conducted to compare the ability of the above mentioned four scores [fatty liver index (FLI), hepatic steatosis index (HSI), lipid accumulation product (LAP), and NAFLD liver fat score (LFS)] to predict NAFLD in two sample sets, a population-based sample and a sample at high risk for NAFLD, using MRI-PDFF, which can accurately quantify liver fat values. In the EFFECT I study, screened patients were eligible for inclusion in the treatment

part of the study provided they were 40-75 years old and had a body mass index (BMI) of 25-40 kg/m<sup>2</sup>, serum triglyceride level of 1.7 mM (150 mg/dL) or higher. The EFFECT II study had similar inclusion and exclusion criteria to the EFFECT I study, with the exception that eligible patients must have had a prior history of type 2 diabetes, and serum triglyceride levels were not considered for inclusion. Data from the screening parts of the EFFECT I and II studies, including both patients who were randomized and screen failures, were used in the present study. The POEM study was a population-based study investigating individuals (all aged 50 years) from Uppsala. Of 502 individuals recruited (50% women), a successful MRI liver scan was performed in 310 individuals. Of the four evaluated scores, FLI was preferable in the population-based sample (NAFLD prevalence, 23%), whereas LFS performed best in the high-risk sample (NAFLD prevalence, 73%). As the POEM study included the sample of 50 years are not representative of the actual population. All the detection scoring system for NAFLD includes multiple variables those requires laboratory facilities. The utility of these are of limited clinical use where simple ultrasound can detect it. In this study severity was not assessed in any way. Prediction of severity is of global need. Inclusion of older group of patients made the study for limited implication. Rationale of the study is of limited value.

**Ans:** We thank the reviewer for raising this key issue. This is a study to see if simple biochemistry-based screening in primary care could be of value in two settings; the high-risk and the population-based setting. In primary care, no ultrasound is available and, therefore, other available methods have to be used. Since the primary care physician cannot refer all subjects with BMI>25 kg/m<sup>2</sup> to ultrasound, using available NAFLD scores could be a cost-effective way to select subjects for further investigations. We have now added the following sentence to the Introduction section (page 6, paragraph 1, lines 6-7) to better clarify the results of the present study: "Thus, the general physician should have simple tools available to use for screening, since not all obese subjects could be referred to imaging or biopsy."

**Reviewer #3:****Comments:**

1. The modern point of view is that there is no “gold standard” for the diagnosis of NAFLD. The diagnosis of this pathology is complex and, to a large extent, comes down to the exclusion of other etiological factors. In the modern literature, there is practically no information on the use of various diagnostic indices in populations with a high and low risk of pathology. In this regard, idea of articles and research design are undoubtedly relevant. The methodological level of work is modern, consistent with the purpose of the study. Mathematical processing was carried out correctly and allowed to obtain objective conclusions.  
It is advisable to indicate in the methods section whether other etiological factors causing liver diseases were diagnosed, as this is a prerequisite for generally accepted modern algorithms.

**Ans:** We thank the reviewer for the comment. We have now included a paragraph in the Discussion section (page 13, paragraph 2, lines 12-17) on this issue: “In the present study, we performed a very detailed history of previous diseases and alcohol intake to exclude other causes of liver steatosis than NAFLD. Thus, although it cannot be excluded that we missed some cases of liver disease other than NAFLD, the vast majority of individuals included in the present study are not likely to have any liver disease other than NAFLD.”

2. The results are certainly original and useful for practitioners and researchers. The authors first showed that the FLI score is a more correct diagnostic algorithm for diagnosing NAFLD in the general population, while the LFS index gives optimal results in populations with a high risk of detecting NAFLD. Certain restrictions on the use of the article are presented by the vague activity of applying these four diagnostic indices for the diagnosis of NAFLD in various countries. In the article, it would be useful to provide data on the frequency of application of the studied indices in practical health care in various countries.

**Ans:** We thank the reviewer for the suggestion. Although we have found research papers on NAFLD scores from across the globe, we could not find information on how frequently these tests are used in practical health care in various countries. We have searched in PubMed and other online resources on this topic without success. We would be grateful for any suggestions regarding how to find such information and happy to incorporate that in the manuscript, if found.

3. It would be useful for the reader of the article, who is a practical physician, to indicate a generally accepted algorithm for diagnosing NAFLD at the present stage and then explain what place the indexes studied in this work can occupy in known algorithms. In general, the article seems relevant, contains original data of interest to practitioners and researchers, and can be printed in the journal after a small adjustment.

**Ans:** We thank the reviewer for bringing this important point to our attention. However, we think that a meta-analysis and/or an expert opinion articulated in a review is a better way to conclude the best or generally accepted algorithm to find individuals with higher risk for NAFLD. We also think that the reply to Reviewer 1, question 3 is of relevance here. Thus, a simplified version of LFS may turn out to be the recommended algorithm, but more studies are needed before it is concluded.