Response to the Reviewers' comments on manuscript number 36496

In the following we address the specific points made by the editor and the referees and describe the changes we have made in the text (page, table and figure numbers refer to the current version). Changes that have been introduced within the revised manuscript are underlined and highlighted in blue color.

REVIEWER #1

The authors have performed an interesting study and results are worthy. However, several details of statistical analysis should be explained or amended. First, it is not clear which variables are included in the multivariable analysis and this is very important since several variables that seemed to have been included could have collinearity problems (it is clear with MELD and creatinine and bilirubin but it is also possible with Child score and MELD score). Indeed, authors should clearly indicate which variables and according which criteria are included in the multivariable analysis. Although this is controversial, I recommend that the authors present a table with a univariate analysis of the association of variables with MELD score improvement (not only of "clinically relevant" variables that have been selected by the authors to be included in the multivariable model). Although these variables are for sure relevant, we cannot know in advance which specific variables can be more relevant in a given sample.

Response:

All variables presented in Table 3 (age, Child-Pugh score, MELD score, creatinine, platelets, albumin and bilirubin) were used for univariable as well as multivariable analysis. To maintain the validity of the logistic regression analysis relative to the number of outcome events, a maximum of 7 variables was selected.[1] Variables were chosen due to clinical relevance based on the clinical experience of the authors, as well as based on the data of other publications.[2] Unfortunately, using this approach, we cannot entirely exclude to omit parameters, which might proof to be of relevance in the future. However, we see no other practical alternative to produce statistically valid data based on our patient sample.

The manuscript has been adjusted accordingly and the following sections have been inserted:

„Variables were chosen due to clinical relevance and are presented in table 3. Theses variables were selected based on the clinical experience of the authors, as well as based on the data of other publications.[2] These specified variables were used for univariable as well as for multivariable analysis." (page 8)

Of the baseline factors examined (age, Child-Pugh score, MELD score, creatinine, platelets, albumin and bilirubin), the Child-Pugh score, the MELD score, the number of platelets and the levels of albumin and bilirubin were significant factors for functional benefit in the univariable analysis. (page 11)

Authors should also specify if they have checked for confusion or interaction, if they have tested the presence of collinearity and which is the calibration of the model (e.g., Hosmer-Lemeshow goodness of fit test). Authors should also state in the Methods section if they have followed a stepwise or non-stepwise method for performing the multivariable logistic regression analysis.

Response:

We performed a non-stepwise method for the multivariable logistic regression analysis. MELD score and Child-Pugh score were assessed for interaction and showed none. We performed the Hosmer-Lemeshow test, which showed significance ($p=0.02$) and therefore demonstrated a low goodness of fit. The removal of creatinine from the multivariable assessments improved the goodness of fit considerably ($p=0.43$) and therefore was excluded from further analyses. Subsequent analyses still showed a trend for MELD ($p=0.082$) and for albumin ($p=0.057$) in the multivariable analyses, however significance at the level of 5% was not reached. Hereof unaffected albumin ($p<0.0001$) and MELD ($p<0.001$) remained significant in the univariable analyses.
The results in table 3 as well as the respective parts in the manuscript have been adapted accordingly.

It could be interesting to analyze differences in SVR and MELD score improvement according to HCV genotype.

Response:

We agree that an analysis of differences in SVR and MELD score improvement according to HCV genotype would be intriguing. However, a further analysis according to various genotypes was hampered by small subgroups of patients due to an uneven distribution of HCV genotypes in the patient sample and therefore was dispensed. In our patient group, in 78% of patients genotype 1 was detected, in 4% of patients genotype 2, in 15% of patients genotype 3 and in 3% of patients genotype 4. Our patient sample is a reflection of existing data for the distribution of genotypes in Central Europe: Approximately 70% for genotype 1, 3% for genotype 2, 21% for genotype 3 and 5% for genotype 4.[3, 4]

The manuscript has been changed accordingly and the following section has been inserted in the manuscript:
Our patient sample reflects existing data for the distribution of genotypes in Central Europe: Approximately 70% for genotype 1, followed by 21% for genotype 3, 3% for genotype 2 and 5% for genotype 4.[3, 4] (page 10)

Minor comments Please write multivariable instead of multivariate. See Hidalgo&Goodman. Am J Public Health. 2013 January; 103(1): 39–40.

Response:
The manuscript has been changed accordingly.

In table 1, authors should include average MELD score

Response:
Average MELD score has been included in table 1.

I find Figure 2A and figure 2B hard to read and I think this information could be conveyed without a figure

Response:
More recently, graphs called waterfall plots have begun to be used in the presentation of data to visually depict treatment responses in patients. In general, waterfall plots go from the worst value on one side of the plot, to the best value on the other side of the plot. The length of each vertical bar hanging below the horizontal axis increases as the plot moves to one side, thus resembling a waterfall and giving the graph its name. Thus, the data are not presented randomly, or in order of when a patient was first enrolled in a study, but are organized in order to provide a clear picture of the study population's results: from worst to best, based on the analysed parameters. The individual bars, besides representing a single subject, can also be used to represent other key patient characteristics using a different color, such as the type of response achieved by the subject (e.g. SVR 12 vs. relapse, as in our presentation). Consequently, a waterfall plot may provide two sets of data of a single subject and thereby visually prepare information in a condensed fashion.[2, 5] We believe, that the visual illustration of the data of our study contributes to a better understanding of key messages of our results and we therefore decided to keep the figures 2A and 2B in the revised version of the manuscript.

REVIEWER #2

Liver cirrhosis due to HCV is a major risk factor for the development of HCC. Interferon treatment has many adverse events therefore limits therapeutic indication, especially in patients with cirrhosis. As authors described, the introduction of DAA has dramatically expanded the indication of the therapy in liver cirrhosis and also has increased SVR. Ultimate goal of antiviral therapy in patients with HCV is to hinder development of HCC or falling the state of uncompensated liver failure. It takes long time to confirm the ultimate goal of the therapy, therefore simple and feasible clinical index is necessary to evaluate the effect of therapy. Improvement of MELD may be one of the surrogate evaluation system. Major 1. MELD score consists of PT, T-Bil and creatinine. Indicate these three factors before and after the treatment.

Response:

The MELD score consists of International Normalized Ratio (INR) for prothrombin time, serum bilirubin and creatinine levels.[6] Before treatment initiation in our study population, INR was 1.34±0.78, bilirubin was 1.4±0.9 mgdl-1 and creatinine 0.80±0.22 mgdl-1. At 12 weeks post-treatment INR was 1.13±0.23, bilirubin was 1.4±1.0 mgdl-1 and creatinine 0.80±0.21 mgdl-1.

The manuscript has been changed accordingly and the following sentences have been inserted:

The MELD score consists of International Normalized Ratio (INR), serum bilirubin and creatinine levels.[6] At treatment initiation, the mean MELD score in our cohort was 9±3 and the respective variables were as follows: INR was 1.34±0.78; bilirubin was 1.4±0.9 mgdl-1 and creatinine 0.80±0.22 mgdl-1. Of all patients, 130 (65%) had a MELD score <10; 59 (30%) had a MELD score in the range of 10-15; and 10 (5%) had a MELD score >15. At 12 weeks post-treatment, laboratory data were available for 179 patients. The average MELD score in the total number of our studied patients remained unchanged with 9±3 at 12 weeks post-treatment and the respective variables were as follows: INR was 1.13±0.23; bilirubin was 1.4±1.0 mgdl-1 and creatinine 0.80±0.21 mgdl-1. (page 11)

2. Do you think which is more important, improvement of MELD score or virus eradication?

Response:

The primary aim of antiviral therapy is the eradication of the hepatitis c virus and therefore the cure of the patient. However, the MELD score is an important parameter in the evaluation of the effect of the therapy as well, as it reflects the changes in mortality associated with viral clearance.

3. When ascertaining risk and benefit of DAA treatment, is MELD score more useful than Child-Pugh score?

Response:

The Child-Pugh score is a prognostic model for liver cirrhosis, which has been a useful clinical tool in day-to-day clinical practice for over 50 years. [7] However, the Child-Pugh score includes subjective criteria (ascites and encephalopathy) besides the laboratory values of INR, bilirubin and albumin. Furthermore, the Child-Pugh score is not as accurate in predicting mortality as another score for the prognosis of patients with liver cirrhosis, which is the Model for End-Stage Liver Disease score (MELD score). In contrast to the Child-Pugh score, the MELD score has been derived from prospectively collected data rather than empirically constructed data. Also, the MELD score increases as the three variables (INR, bilirubin and creatinine) deteriorate, whereas the constituent parameters in the Child-Pugh score remain fixed once a defined threshold has been reached.[8] Therefore, the MELD score has been established as a classification system to determine the urgency for liver transplantation. Consequently, we assume, the MELD score also might be a valuable tool in assessing the risk and benefit of DAA treatment and might be more reliable than the Child-Pugh score.

This issue has been introduced into the discussion of the revised manuscript:

Considering that not all patients with cirrhosis benefit from HCV therapy despite SVR, the key question is which patients profit. ~~We identified a higher pre-treatment MELD score as a predictor of a favourable outcome of patients with cirrhosis receiving DAA therapy in our patient group.~~ The Child-Pugh score is a prognostic model for liver cirrhosis, which has been a useful clinical tool in day-to-day clinical practice for over 50 years.[7] However, the Child-Pugh score includes subjective criteria (ascites and encephalopathy) besides the laboratory values of INR, bilirubin and albumin. Furthermore, the Child-Pugh score is not as accurate in predicting mortality of patients with liver cirrhosis as is the MELD score. In contrast to the Child-Pugh score, the MELD score has been derived from prospectively collected data rather than empirically constructed data. Also, the MELD score increases as the three variables (INR, bilirubin and creatinine) deteriorate, whereas the constituent parameters in the Child-Pugh score remain fixed once a defined threshold has been reached.[8] Therefore, the MELD score has been established as a classification system to determine the urgency for liver transplantation. Consequently, we suggest, that the MELD score also might be a valuable tool in assessing the risk and benefit of DAA treatment and might be more reliable than the Child-Pugh score. (page 14)

References

1 Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; **49**(12): 1373-1379 [PMID: 8970487]

2 Foster GR, Irving WL, Cheung MC, Walker AJ, Hudson BE, Verma S, McLauchlan J, Mutimer DJ, Brown A, Gelson WT, MacDonald DC, Agarwal K, Hcv Research UK. Impact of direct acting antiviral therapy in patients with chronic hepatitis C and decompensated cirrhosis. *J Hepatol* 2016; **64**(6): 1224-1231 [PMID: 26829205 DOI: 10.1016/j.jhep.2016.01.029]

3 Petruzziello A, Marigliano S, Loquercio G, Cacciapuoti C. Hepatitis C virus (HCV) genotypes distribution: an epidemiological up-date in Europe. *Infect Agent Cancer* 2016; **11**: 53 [PMID: 27752280 PMCID: PMC5062817 DOI: 10.1186/s13027-016-0099-0]

4 Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, Barnes E. Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* 2015; **61**(1): 77-87 [PMID: 25069599 PMCID: PMC4303918 DOI: 10.1002/hep.27259]

5 Gillespie TW. Understanding waterfall plots. *J Adv Pract Oncol* 2012; **3**(2): 106-111 [PMID: 25031936 PMCID: PMC4093310]

6 Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL, D'Amico G, Dickson ER, Kim WR. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001; **33**(2): 464-470 [PMID: 11172350 DOI: 10.1053/jhep.2001.22172]

7 Child CG, Turcotte JG. Surgery and portal hypertension. *Major Probl Clin Surg* 1964; **1**: 1-85 [PMID: 4950264]

8 Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 2000; **31**(4): 864-871 [PMID: 10733541 DOI: 10.1053/he.2000.5852]