Improving survival prediction models for liver transplantation candidates

Goudsmit, B.F.J.

Chapter 10

Appendix

10.1 Supplement Chapter: The Model for End-stage Liver Disease 3.0: an update without proven accuracy

Goudsmit BFJ, Putter H, van Hoek B. The Model for End-stage Liver Disease 3.0: an update without proven accuracy. *Gastroenterology*, 2021; doi: 10.1053/j.gastro.2021.09.047.

Letter

With great interest we read the study by Kim et al.[1] In this work, the authors showed that MELD-Na performance is improved by including serum albumin levels, LT candidate sex, a creatinine cap set to 3 mg/dL, and significant interactions. Most notably, the MELD 3.0 concordance statistic (c-index) was 0.869, versus a MELD-Na c-index of 0.862. However, we have some concerns regarding this study.

First, the authors report only discrimination (c-index) as model performance indicator. Indeed, high discrimination is important when ranking patients for LT, as it ensures that the model prioritizes the sickest patients. However, when basing treatment decisions on estimated mortality risks, it is vital to assess and report how accurate risks are estimated, i.e., model calibration. This is because a badly calibrated model can still have a high c-index, but treatment decisions should not be based on such a model.[2] Model calibration is typically reported with calibration plots, that give insight in possible over- or underestimation of risk. Previous work showed that MELD-Na overestimated risks for the sickest patients.[3,4] More importantly, recent study found that MELD predicted risks inaccurately.[5] Therefore, the authors cannot conclude that "MELD 3.0 affords more accurate mortality prediction," as calibration was not reported. It would be interesting to assess and report MELD 3.0 calibration, especially for male versus female LT candidate sex.

Second, the authors report net 8.8% reclassification of deceased patients from a lower MELD-Na stratum to a higher MELD 3.0 stratum, for women this number was 14.9%. The idea is that higher MELD 3.0 scores thus better reflect mortality risks. The first important concern with proving MELD 3.0 prediction improvement through reclassification methods is that a poorly calibrated model can show improved prediction performance, even when this is not possible.[6] These false effects can be found both in actual cohorts and simulated data. In

part, this is due to the fact that the actual waiting list population cannot be separated into the suggested MELD strata (6-9, 10-19, etc.). Instead, when evaluating added biomarkers, measures like the Brier score, that simultaneously assess discrimination and calibration, should be used in independent validation data.[6] A second concern is that reclassification allows for 'stage migration bias,'[7] i.e., assigning patients to new strata improves strata-specific survival, even though survival of individual patients has not changed. The sickest patients from a lower MELD-Na stratum are moved to a higher MELD 3.0 stratum and survival is better in both strata. Therefore, stating that MELD 3.0 will lower deaths on the waiting list based on reclassification tables must be done cautiously, as this can inflate within-strata survival rates.

Third, the authors keep the lower borders of bilirubin, creatinine, and INR set to 1. These borders were chosen 20 years ago, to prevent negative logarithm transformation in the linear MELD formula. The more pressing clinical fact is that a substantial number of patients on the waiting list had creatinine (55%) and bilirubin (24%) values below 1 mg/dL at first registration.[8] Including these lower measurements when predicting survival would be a better representation of the actual waiting list and would place the higher values in a more appropriate context, especially considering the lower creatinine values for women. Also, even though linear models are more easily understood and used, non-linear effects are clearly present (creatinine, sodium, and albumin). Therefore, flexible models could be considered to model more measurements and their non-linear effect on mortality.

In conclusion, MELD 3.0's accuracy must be proven before it can be considered as new allocation model, e.g., with calibration plots and Brier scores. Reclassification cannot be used alone to prove clinical improvement. We agree with the authors that efforts should be made to continuously improve MELD and liver graft allocation, but appropriate evidence must be presented.

References

1. Kim WR, Mannalithara A, Heimbach JK, et al. MELD 3.0: The Model for End-stage Liver Disease Updated for the Modern Era. Gastroenterology. Published online 2021. doi:10.1053/j.gastro.2021.08.050
2. Van Calster B, McLernon DJ, Van Smeden M, et al. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):1-7. doi:10.1186/s12916-019-1466-7
3. Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and Mortality among Patients on the Liver-Transplant Waiting List. N Engl J Med. 2008;359(10):1018-1026. doi:10.1007/s11250-017-1262-3
4. Goudsmit BFJ, Putter H, Tushuizen ME, et al. Validation of the Model for End-stage Liver Disease sodium (MELD-Na) score in the Eurotransplant region. Am J Transplant. Published online 2020. doi:10.1111/ajt.16142
5. D'Amico G, Maruzzelli L, Airoldi A, et al. Performance of the model for end-stage liver disease score for mortality prediction and the potential role of etiology. J Hepatol. Published online 2021. doi:10.1016/j.jhep.2021.07.018
6. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: Do not rely on integrated discrimination improvement and net reclassification index. Stat Med. 2014;33(19):3405-3414. doi:10.1002/sim.5804
7. Feinstein AR, Sosin DM, Wells CK. The New England Journal of Medicine Downloaded from nejm.org at BOSTON UNIVERSITY on September 19, 2013. For personal use only. No other uses without permission. From the NEJM Archive. Copyright © 2010 Massachusetts Medical Society. All rights reserved. N Engl J Med. 1985;312(12):1604-1608.
8. Goudsmit BFJ, Putter H, Tushuizen ME, et al. Refitting the Model for End-Stage Liver Disease for the Eurotransplant Region. Hepatology. 2021;74(1):351-363. doi:10.1002/hep.31677