# The Model for End-Stage Liver Disease 3.0: An Update Without Proven Accuracy

Goudsmit, B.J.; Putter, H.; Hoek, B. van

these stable patients would not receive undue priority compared with patients with decompensated cirrhosis.

Third, the increase in priority of bilirubin from a coefficient of 3.78 to 4.56 will improve the equity of MELD allocation for patients with primary biliary cholangitis.[8] Patients with cholestatic liver disease represent a minority of listed patients, and it would be educational to evaluate how the new MELD 3.0 compares with MELDNa for their allocation, which we hope and expect will better represent their true mortality risk.

In conclusion, MELD 3.0 is a definite step in the right direction of increasing equity and evolving the current MELDNa score to keep up with changing demographics. However, further information on important subgroups is needed. Ultimately, reconsideration of replacing the new objective but divisive and politically charged variable of sex for a neutral variable that objectively quantitates muscle mass will improve operationalization and elevate MELD 3.0 from one small step for womankind to one big step for everyone.

*JACQUELINE G. O'LEARY*
Department of Medicine
Dallas Veterans Medical Center and
University of Texas Southwestern Medical Center
Dallas, Texas

*JASMOHAN S. BAJAJ*
Division of Gastroenterology, Hepatology and Nutrition
Department of Medicine
Virginia Commonwealth University *and*
Richmond Veterans Affairs Medical Center
Richmond, Virginia

## References

1. Kim WR, et al. Gastroenterology 2021;161:1887–1895.
2. O'Leary JG, et al. Dig Dis Sci 2017;62:768–776.
3. Tandon P, et al. J Hepatol 2021;75(Suppl 1):S147–S162.
4. Engelmann C, et al. Aliment Pharmacol Ther 2018;48:1271–1281.
5. Kamath PS, et al. Hepatology 2001;33:464–470.
6. Martin-Llahi M, et al. Gastroenterology 2011;140:488–496.e4.
7. Wong F, et al. Liver Transpl 2019;25:870–880.
8. Singal AK, et al. Transpl Int 2017;30:454–462.

**Conflicts of interest**
The authors disclose no conflicts.

Most current article

https://doi.org/10.1053/j.gastro.2021.09.014

# The Model for End-Stage Liver Disease 3.0: An Update Without Proven Accuracy

Dear Editors:

With great interest we read the study by Kim et al.[1] In this work, the authors showed that Model for End-stage Liver Disease (MELD)-Na performance is improved by including serum albumin levels, liver transplantation (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 because 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 (ie, 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 overestimation or underestimation of risk. Previous work showed that MELD-Na overestimated risks for the sickest patients.[3,4] More importantly, a recent study found that MELD predicted risks inaccurately.[5] Therefore, the authors cannot conclude that "MELD 3.0 affords more accurate mortality prediction" because 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, which simultaneously assess discrimination and calibration, should be used in independent validation data.[6]

A second concern of reclassification is that reclassification allows for "stage migration bias,"[7] (ie, assigning patients to new strata improves strata-specific survival), although 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 because 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 $<1$ 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, although linear models are more

easily understood and used, nonlinear effects are clearly present (creatinine, sodium, and albumin). Therefore, flexible models could be considered to model more measurements and their nonlinear effect on mortality.

In conclusion, MELD 3.0's accuracy must be proven before it can be considered as a new allocation model (eg, 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.

*BEN F.J. GOUDSMIT*
Division of Gastroenterology and Hepatology
Leiden University Medical Center
Leiden, The Netherlands

*H. PUTTER*
Division of Biostatistics
Leiden University Medical Center
Leiden, The Netherlands

*B. VAN HOEK*
Division of Gastroenterology and Hepatology
Leiden University Medical Center
Leiden, The Netherlands

## References

1. Kim WR, et al. Gastroenterology 2021;161:1887–1895.
2. Van Calster B, et al. BMC Med 2019;17:1–7.
3. Kim WR, et al. N Engl J Med 2008;359:1018–1026.
4. Goudsmit BFJ, et al. Am J Transplant 2020;21:229–240.
5. D'Amico G, et al. J Hepatol 2021. in press.
6. Hilden J, et al. Stat Med 2014;33:3405–3414.
7. Feinstein AR, et al. N Engl J Med 1985;312:1604–1608.
8. Goudsmit BFJ, et al. Hepatology 2020;74:351–363.

**Conflicts of interest**
The authors disclose no conflicts.

**Most current article**

https://doi.org/10.1053/j.gastro.2021.09.047

**Reply.** We appreciate Dr O'Leary and Dr Bajaj's letter as a valuable contribution to the discussion regarding refinement of the Model for End-Stage Liver Disease (MELD).[1,2] The authors note relevant considerations concerning the proposed modifications in MELD 3.0, which include (1) the addition of sex and albumin, (2) updated coefficients and interaction terms, and (3) a lowered upper bound of serum creatinine from 4 to 3 mg/dL.

The addition of sex aims to correct the disparity that arises from the deficiency of creatinine to accurately reflect muscle mass, particularly in women with advanced liver disease. There are 2 interrelated issues: (1) women, in general, have lower muscle mass than men, and (2) sarcopenia is prevalent in cirrhosis. Sarcopenia is indeed an important prognostic factor; however, the most widely accepted parameter, the skeletal muscle index, may be difficult to implement and standardize for widespread application such as organ allocation. Moreover, the current criteria to define sarcopenia also include sex, which defeats the purpose if the goal is to remove sex as a variable.

With regard to gender alteration, we agree that the sex variable in the MELD 3.0 equation needs to be clearly defined, particularly if it is to be applied in organ allocation. It should be clear that the role that sex plays in the equation is physiologic (ie, a modification for the serum creatinine) rather than a matter of gender identity. This is most often and most accurately reflected by the sex at birth, although there may be some allowance for reclassification in specific situations, such as receipt of hormonal therapies of a certain duration, which could alter total body muscle mass.[3]

We appreciate the astute observations regarding the updated coefficients in MELD 3.0. Overall, the effect appears to be a greater weight given to liver rather than renal dysfunction. Indeed, the coefficient of creatinine has increased; yet overall, there are fewer potential creatinine-derived points available for those with elevated creatinine, due to the upper bound of 3 mg/dL and the negative interaction term of creatinine with albumin. In the hypothetical patient with compensated cirrhosis on dialysis given the maximum creatinine and albumin of 2 g/dL, the MELD 3.0 score would be 19, no better off than the current MELD-Na of 20 and well below the median MELD in nearly all regions. Patients with cholestatic liver disease, whose mortality risk is not well represented by MELD or MELD-Na, should benefit from the higher coefficient afforded to bilirubin.[4] We recognize the need for careful analysis to assess the impact of these various changes on such subgroups.

The implementation of MELD-based allocation for liver transplantation in 2002 improved outcomes but introduced a measurable sex disparity, underrepresenting medical urgency of women relative to men at the same mortality risk.[5] While a more accurate representation of renal dysfunction or muscle mass, or both, that is unbiased to sex or race would be ideal, it does not exist in a form that can be currently deployed. The addition of sex—as a biological, rather than "divisive or politically charged" variable—addresses this disparity and provides a more accurate representation of waiting list mortality, where women are at least not unequal to men. While the addition of sex is the headline of MELD 3.0, we appreciate the call to attention toward the other updates to MELD 3.0 that also improve its performance, including the addition of albumin, interaction terms, and updated coefficients. These changes are long overdue, given the rapidly shifting landscape of chronic liver disease, yet deserve close scrutiny to ensure that MELD is not just a small step for womankind, but also a big step for everyone.

We also appreciate the comments submitted by Goudsmit et al.[6] We agree that calibration is an important component of evaluating model performance beyond discrimination and net reclassification. In our preliminary work, we evaluated calibration via the Greenwood-Nam-