approaches in the era of CFTR modulator therapies. J Hepatol 2021;76(2):420–434.
[3] Wu M, Wu L, Jin J, Wang J, Li S, Zeng J, et al. Liver stiffness measured with two-dimensional shear-wave elastography is predictive of liver-related events in patients with chronic liver disease due to hepatitis B viral infection. Radiology 2020;295:353–360.
[4] Levitte S, Fuchs Y, Wise R, Sellers ZM. Effects of CFTR modulators on serum biomarkers of liver fibrosis in children with cystic fibrosis. Cold Spring Harbor Lab 2022;7(2):e0010.
[5] Drummond D, Dana J, Berteloot L, Schneider-Futschik EK, Chedevergne F, Bailly-Botuha C, et al. Lumacaftor-ivacaftor effects on cystic fibrosis-related liver involvement in adolescents with homozygous F508 del-CFTR. J Cyst Fibros 2021;21(2):212–219.
[6] Sugimoto K, Moriyasu F, Oshiro H, Takeuchi H, Abe M, Yoshimasu Y, et al. The role of multiparametric US of the liver for the evaluation of nonalcoholic steatohepatitis. Radiology 2020;296:532–540.
[7] Deffieux T, Gennisson JL, Bousquet L, Corouge M, Cosconea S, Amroun D, et al. Investigating liver stiffness and viscosity for fibrosis, steatosis and activity staging using shear wave elastography. J Hepatol 2015;62:317–324.
[8] Allen AM, Shah VH, Therneau TM, Venkatesh SK, Mounajjed T, Larson JJ, et al. The role of three-dimensional magnetic resonance elastography in the diagnosis of nonalcoholic steatohepatitis in obese patients undergoing bariatric surgery. Hepatology 2020;71:510–521.

Fairness metrics: Additional principles to consider for improving MELD

To the Editor:
I would like to congratulate all cocreators of the model for end-stage liver disease 3.0 (MELD 3.0) score[1] on the decision of the Organ Procurement and Transplant Network (OPTN) Board to replace the MELD-Na with the MELD 3.0 to determine organ allocation priorities in the US. In a letter published on November 16, 2022, in the *Journal of Hepatology*,[2] the authors highlighted several principles that were followed in the development of the new score. All variables included in the score must be (1) measurable in an objective fashion, (2) generalizable, (3) devoid of unnecessary volatility without biological significance, and (4) reportable to the OPTN without causing an undue burden.

Sex was incorporated into the MELD 3.0 for two reasons: 1) to mitigate against the sex disparity in access to transplantation, and 2) to improve predictive performance. The rationale behind this decision is valid for a variety of reasons. For example, sex differences were incorporated into the new score to correct for sex disparity caused by creatinine and differences in risk of death.[1,3] Although it was mentioned in the MELD 3.0-related manuscripts that the objective of adding the new sex variable, as well as revising the creatinine coefficient, was to improve fairness across the sexes, it was not thoroughly discussed. The changes that were applied to account for fairness are truly important; however, performance of the new score was only assessed by using standard metrics such as discrimination and calibration. As a more modern score that accounts for fairness, it is also crucial to assess the performance of the MELD 3.0 by using metrics specific to fairness such as statistical parity difference (*i.e.*, likelihood of being classified as high risk), true positive rate (sensitivity) difference, and true negative (specificity) rate difference.[4] Such measures can assess fairness within the context of patient characteristics such as sex, race, and age.

In research that I led that resulted from a collaboration between IBM Research and the Broad Institute of MIT and Harvard, we assessed performance of widely used risk scores in cardiology: the CHARGE-AF (cohorts for heart and aging in genomic epidemiology atrial fibrillation) score for atrial fibrillation and the pooled cohort equations (PCE) score for atherosclerotic cardiovascular disease (ASCVD).[5] We assessed performance by using standard metrics such as discrimination, calibration, and standard hazard ratios as well as fairness-related metrics considering sex, race, and age ranges. We observed evidence of potentially unfair performance, with significant differences in fairness metrics for sex and race in both scores, considering three large independent datasets: Explorys Life Sciences Dataset, Mass General Brigham, and the UK Biobank. For instance, the sensitivity difference of both scores was much lower for females than males in the intermediate-age subgroups, suggesting that current scores may miss more

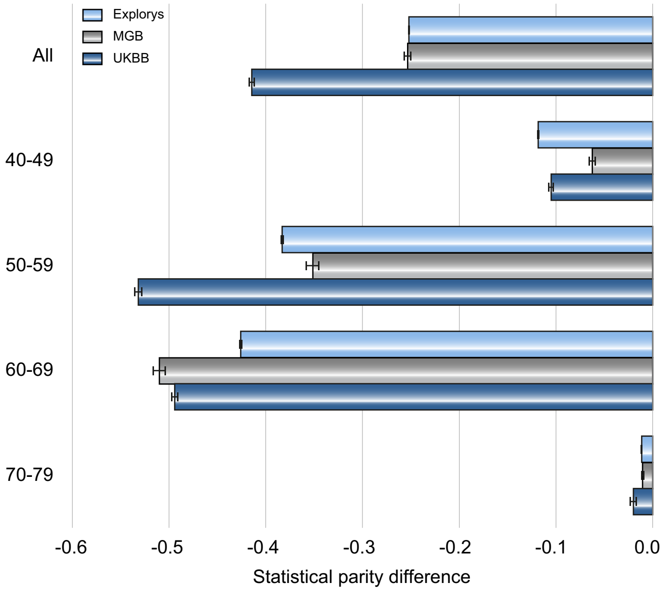

Fig. 1. Statistical parity difference for sex using the PCE with a stratification into 10-year age ranges.[5] Comparison includes the following three databases: Explorys Life Sciences Dataset (Explorys, n = 21,853,866), Mass General Brigham (MGB, n = 520,868), and the UK Biobank (UKBB, n = 502,521). PCE, pooled cohort equations.

females at high risk for events, potentially worsening existing sex-related treatment gaps.[6] Overall, our findings underscore the importance of evaluating prognostic models across the many specific subpopulations for which risk prediction is intended to better understand the accuracy and potential unfairness of the prognostic information used to drive clinical decisions at the point of care.

A highly fair score yields exact values for two individuals – for example, male and female if they have an exact medical profile (*e.g.*, identical age, laboratory values, and comorbidities). Furthermore, a highly fair score yields exact values at any age (*e.g.*, their scores may increase or decrease as a function of age but will stay identical). A highly unfair MELD-related score indicates a reduced risk for a female compared to a male to survive the next 90 days, while the actual level of severity of each is identical. Within the context of the PCE, US guidelines recommend determining whether individuals without established ASCVD should be considered for cholesterol-lowering therapy if their scores exceed the 7.5% threshold.[7,8] As illustrated in Fig. 1, ASCVD risk estimates were found to be lower for females compared to males across all age ranges in the three independent databases, and to a higher extent in intermediate ages (up to approximately 50% lower).[5] As an example, when a clinician sees two patients, such as a female and a male in the age range of 60 to 69 with identical medical profiles – if the male has a ASCVD risk score of 10% (thus the decision would be to initiate statins), the female will have a score of approximately 5% only. As a result of this underestimation of the risk of the female by the PCE, the decision would be to not initiate statins, while she is likely to benefit from them. In a highly fair scenario that requires a redesign of the PCE, the values in all bars in Fig. 1 will be as close as possible to zero, as achieved only for the 70- to 79-year-old individuals. Similarly, with current and future versions of MELD, a score used to prioritize livers must be associated with statistical parity difference values that indicate small to non-existent bias across all age ranges, considering characteristics such as sex and race as well as other characteristics suspected of contributing to a potential bias.

Future versions of the CHARGE-AF and PCE that also account for fairness across a range of protected variables are expected to help clinicians treat patients more equitably. Similarly, risk assessment tools are expected to more equitably rank patients on the liver allocation list, once configured properly to optimize fairness-related metrics. Thus, I suggest that the development of future versions of the MELD score should also include comprehensive reporting on fairness-related metrics in addition to the traditional, widely used metrics.

Uri Kartoun*
IBM Research, Cambridge, MA, USA
*Corresponding author. Address: Center for Computational Health, IBM Research, Cambridge, MA, USA.
E-mail address: uri.kartoun@ibm.com

Received 6 December 2022; Received in revised form 29 December 2022; Accepted 21 January 2023; Available online 7 February 2022
https://doi.org/10.1016/j.jhep.2023.01.028

Financial support

The author did not receive financial support.

Conflict of interest

The author has no conflicts of interest to disclose as described by the Journal of Hepatology.

Please refer to the accompanying ICMJE disclosure form for further details.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2023.01.028.

References

Author names in bold designate shared co-first authorship

[1] Kim WR, Mannalithara A, Heimbach JK, Kamath PS, Asrani SK, Biggins SW, et al. Meld 3.0: the model for end-stage liver disease updated for the modern era. Gastroenterology 2021 Dec;161(6):1887–1895.e4. https://doi.org/10.1053/j.gastro.2021.08.050. Epub 2021 Sep 3, 34481845. PMCID: PMC8608337.

[2] Ge J, Kim WR, Lai JC, Kwong AJ. Title: on building a better mousetrap response to: "towards optimally replacing the current version of MELD".. 3300-3301 J Hepatol 2022 Nov 16;(22):S0168–S8278. https://doi.org/10.1016/j.jhep.2022.11.008. Epub ahead of print, 36402449.

[3] Ge J, Kim WR, Lai JC, Kwong AJ. "Beyond MELD" – emerging strategies and technologies for improving mortality prediction, organ allocation and outcomes in liver transplantation. J Hepatol 2022;76(6):1318–1329.

[4] Bellamy RKE, Dey K, Hind M, Hoffman SC, Houde S, Kannan K, et al. AI fairness 360: an extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. 2018. https://doi.org/10.48550/arXiv.1810.01943. Published online October 3.

[5] **Kartoun U**, **Khurshid S**, Kwon BC, Patel AP, Batra P, Philippakis A, et al. Prediction performance and fairness heterogeneity in cardiovascular risk models. Sci Rep 2022;12(1):12542. https://doi.org/10.1038/s41598-022-16615-3.

[6] Mehran R, Vogel B, Ortega R, Cooney R, Horton R. The Lancet Commission on women and cardiovascular disease: time for a shift in women's health. The Lancet 2019;393(10175):967–968. https://doi.org/10.1016/S0140-6736(19)30315-0.

[7] Goff Jr DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, Robinson JG, Schwartz JS, Shero ST, Smith Jr SC, Sorlie P, Stone NJ, Wilson PW, Jordan HS, Nevo L, Wnek J, Anderson JL, Halperin JL, Albert NM, Bozkurt B, Brindis RG, Curtis LH, DeMets D, Hochman JS, Kovacs RJ, Ohman EM, Pressler SJ, Sellke FW, Shen WK, Smith Jr SC, Tomaselli GF. American college of cardiology/American heart association task force on practice guidelines. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American college of cardiology/American heart association task force on practice guidelines. Circulation 2014 Jun 24;129(25 Suppl 2):S49–S73. https://doi.org/10.1161/01.cir.0000437741.48606.98. Epub 2013 Nov 12. Erratum in: Circulation. 2014 Jun 24;129(25 Suppl 2):S74–5, 24222018.

[8] Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American college of cardiology/American heart association task force on clinical practice guidelines. Circulation 2019;140(11):e596–e646. https://doi.org/10.1161/CIR.0000000000000678.