

OPEN

MELD-GRAIL and MELD-GRAIL-Na Are Not Superior to MELD or MELD-Na in Predicting Liver Transplant Waiting List Mortality at a Single-center Level

John D. Chetwood, MBBS(Dist), DTM&H,^{1,2} Mark V. Wells, BSc,³ Tatiana Tsoutsman, PhD,^{1,2} Carlo Pulitano, MD, PhD, FRACS,^{2,3} Michael D. Crawford, MBBS, FRACS, M.Med,^{2,3} Ken Liu, BSc(Med), FRACP,^{1,2,3,4} Simone I. Strasser, MBBS, MD, FRACP, FAASLD,^{1,2,3} Geoffrey W. McCaughan, MBBS, PhD, FRACP, FAHMS,^{1,2,3,4} and Avik Majumdar, MPHTM, PhD, FRACP^{1,2,3}

Background. Controversy exists regarding the best predictive model of liver transplant waiting list (WL) mortality. Models for end-stage liver disease–glomerular filtration rate assessment in liver disease (MELD-GRAIL) and MELD-GRAIL-Na were recently described to provide better prognostication, particularly in females. We evaluated the performance of these scores compared to MELD and MELD-Na. **Methods.** Consecutive patients with cirrhosis waitlisted for liver transplant from 1998 to 2017 were examined in this single-center study. The primary outcome was 90-d WL mortality. MELD, MELD-Na, MELD-GRAIL, and MELD-GRAIL-Na at the time of WL registration were compared. Model discrimination was assessed with area under the receiver operating characteristic curves and Harrell's C-index after fitting Cox models. Model calibration was examined with Grønnesby and Borgan's modification of the Hosmer-Lemeshow formula and by comparing predicted/observed outcomes across model strata. **Results.** The study population comprised 1108 patients with a median age of 53.5 (interquartile range 48–59) y and male predominance (74.9%). All models had excellent areas under the receiver operating characteristic curves for the primary outcome (MELD 0.89, MELD-Na 0.91, MELD-GRAIL 0.89, MELD-GRAIL-Na 0.89; all comparisons $P > 0.05$). Youden index cutoffs for 90-d mortality were as follows: MELD, 19; MELD-Na, 22; MELD-GRAIL, 18; and MELD-GRAIL-Na, 17. Variables associated with 90-d mortality on multivariable Cox regression were sodium, bilirubin, creatinine, and international normalized ratio. There were no differences in model discrimination using Harrell's C-index. All models were well calibrated; however, divergence between observed and predicted mortality was noted with scores ≥ 25 . **Conclusion.** There were no demonstrable differences in discrimination or calibration of GRAIL-based models compared with MELD or MELD-Na in our cohort. This suggests that GRAIL-based models may not have meaningful improvements in discriminatory ability when applied to other settings.

(*Transplantation Direct* 2022;8: e1346; doi: 10.1097/TXD.0000000000001346).

INTRODUCTION

Accurate prediction of outcomes for those awaiting liver transplant (LT) is important, as it allows for equitable organ allocation based on the “sickest first” principle.¹ The majority of deceased donor programs rely on the model for end-stage liver disease

(MELD) score, which is calculated from serum bilirubin, creatinine, and international normalized ratio (INR), or variations of the formula such as MELD-Na (MELD corrected for serum sodium) to prioritize potential recipients at the highest risk of death.

Received 4 April 2022. Revision received 9 May 2022.

Accepted 10 May 2022.

¹ AW Morrow Gastroenterology and Liver Centre, Royal Prince Alfred Hospital, Sydney, Australia.

² Central Clinical School, The University of Sydney, Sydney, Australia.

³ Australian National Liver Transplant Unit, Royal Prince Alfred Hospital, Sydney, Australia.

⁴ Liver Injury and Cancer Program, Centenary Institute, Sydney, Australia.

The authors declare no funding or conflicts of interest.

J.D.C. was involved in data extraction, undertook statistical analysis, and composed the manuscript. A.M. was involved in study concept, design, statistical analysis, and critical review of the manuscript. M.G.W. was involved in data extraction and critical review of the manuscript. T.T., C.P., M.D.C., K.L., S.I.S., and G.W.M. were involved in study concept and critical review of the manuscript. All authors approved the final version of the manuscript.

Supplemental digital content (SDC) is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.transplantationdirect.com).

Correspondence: Avik Majumdar MBBS(Hons), MPHTM, FRACP, PhD, AW Morrow Gastroenterology and Liver Centre, Royal Prince Alfred Hospital, Missenden Rd, Camperdown, NSW 2050, Australia. (avik.majumdar@health.nsw.gov.au).

Copyright © 2022 The Author(s). *Transplantation Direct*. Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 2373-8731

DOI: 10.1097/TXD.0000000000001346

Although MELD-based models are the current standard of care for LT waiting list (WL) prioritization, inherent weaknesses are still present.^{1,2} Despite being associated with mortality, serum creatinine is a poor measure of renal function in patients with cirrhosis, particularly in females and in the presence of sarcopenia.³ Indeed, the difference between estimated glomerular filtration rate (eGFR) calculated with the Modification of Diet for Renal Disease (MDRD) formula and the true GFR measured by radionuclide studies has been shown to be $>20\text{ mL/min/1.73 m}^2$ in approximately 50% of patients with cirrhosis.³ As a result, the predictive capacity of MELD or MELD-Na is poorer at higher scores where correct organ allocation is of greatest consequence.² Furthermore, these scores were developed over a decade ago on populations not reflective of current LT WLs, which consist of patients with higher MELD scores, older age, more comorbidity, and different liver disease etiologies.⁴⁻⁷

Recently, the Glomerular Filtration Rate Assessment in Liver Disease (GRAIL) model was generated using serum creatinine, blood urea nitrogen, age, sex, race, and albumin to estimate GFR based on timing of measurement relative to LT and degree of renal dysfunction. It was found to have less bias and to be more accurate in predicting true GFR, compared with Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), MDRD-4, and MDRD-6 at time points before and after LT.⁸ The MELD-GRAIL and MELD-GRAIL-Na (MELD-GRAIL corrected for serum sodium) scores substituted GRAIL measurement for eGFR and were shown to improve prediction of WL mortality⁴; however, these scores have not been validated outside of the United States.

These scores were derived from a large national data set, and how these scores perform at an individual LT unit setting has not yet been quantified. We therefore aimed to evaluate the optimal model (MELD-GRAIL, MELD-GRAIL-Na, or MELD-Na) for predicting WL mortality in an Australian LT center.

MATERIALS AND METHODS

Patients

We performed a retrospective cohort study at a single state wide LT center. All adult patients (over 18 y of age) who were registered on the LT WL at our center for the indication of decompensated cirrhosis and/or hepatocellular carcinoma (HCC) from January 1, 1998, to January 1, 2018, were included. Patients were excluded if they did not have cirrhosis, were listed for acute or subacute liver failure or under MELD exception criteria, had a prior LT, or required additional kidney or another organ transplant. The study was conducted according to the Declaration of Helsinki and was approved by the Sydney Local Health District Human Ethics Research Committee with a waiver of informed consent (X19-0213 and 2019/ETH11680). No organs from executed prisoners were used.

Liver organ allocation in Australia is state based. The Australian National Liver Transplant Unit services the state of New South Wales and the Australian Capital Territory and performs approximately 80 to 90 adult LTs annually. WL priority is primarily determined by MELD score, with MELD-Na and the clinical consensus of the LT team used to discriminate patients with similar MELD scores. HCC does not attract MELD exception points. Instead, patients are prioritized on a

case-by-case basis at weekly consensus WL meetings based on the risk of delisting due to tumor progression or clinical deterioration. Since February 2016, a “Share 35” policy has been implemented, where patients with MELD score ≥ 35 are given similar priority as United Network for Organ Sharing status 2A patients and may access organs from any organ allocation jurisdiction across Australia and New Zealand.^{9,10}

Data

Data were extracted from a prospectively maintained LT database and corroborated with electronic and paper medical records. Demographic data, laboratory values, and clinical examination findings from the time of first registration on the LT WL were collected. Outcome data included reason and date for delisting, mortality, or transplantation. Patients were delisted if they had clinical improvement, clinical deterioration beyond medical suitability for LT (including HCC tumor progression), psychosocial factors deeming them unsuitable for LT, and/or transfer to another LT center.

Statistical Analysis

The primary outcome was death at 90 d after WL registration. The secondary outcomes were death at 30 d and at 1 y after waitlisting. Patients were censored if they underwent LT before 90 d for the primary outcome analysis or before 30 d or 1 y for the secondary outcome analyses. Death within 30 d of delisting due to clinical deterioration was classified as mortality at the time of delisting.

Data were expressed as numbers and percentages, means with SD, or medians with interquartile range (IQR) as appropriate. Proportions were expressed as counts with percentages. Baseline group comparisons between survivors and nonsurvivors were performed using Fisher exact and chi-square tests for proportions, Student *t* tests for parametric data, and Wilcoxon rank sum tests otherwise. The following scores were calculated at WL registration as previously described: MELD,¹¹ MELD-Na,¹² MELD-GRAIL, and MELD-GRAIL-Na.² Simultaneous calculations for MDRD-4, MDRD-6, and CKD-EPI equations were also made.^{13,14}

Model discrimination for the primary and secondary outcomes was assessed. Discrimination is the ability of a model to correctly identify patients with different outcomes, for example, death or survival. We initially assessed discrimination using areas under the receiver operating characteristic curves (AUROCs), which measure the performance of a binary outcome (death versus alive). Hence, patients who were transplanted before outcome of interest were excluded from this part of the analysis. Model AUROCs were then each compared to MELD using the DeLong method. The Youden index was calculated for each model to determine the optimal cutoff for the outcomes of interest. Using these cutoffs, the sensitivity, specificity, and positive and negative predictive values (NPV) for each score were calculated for primary and secondary outcomes. Survival analysis was then performed to include patients who had been transplanted before the outcome of interest (ie, the entire cohort including those who were excluded from the AUROC analysis). Variables associated with the primary outcome were identified with a multivariable Cox proportional hazards model that was constructed using stepwise forward selection and backward elimination of variables with $P < 0.10$ on univariable analysis. Cox models were then fitted for each of MELD, MELD-Na,

MELD-GRAIL, and MELD-GRAIL-Na and expressed as hazard ratios (HRs) with 95% confidence intervals (95% CIs). Model discrimination was then compared using Harrell's concordance statistic (C-index), which accounts for the survival time of those transplanted before 90 d.

Calibration, or goodness-of-fit, is a measure of how closely the predicted outcomes of a model compare with observed outcomes. We assessed calibration by comparing Kaplan-Meier mortality estimates (observed) with predicted Cox regression mortality across deciles of risk. We used a modification of the Hosmer-Lemeshow formula as described by Grønnesby and Borgan,¹⁵ with 9 *df* to assess differences between observed and predicted values (where $P > 0.05$ indicates a good fit). Additionally, we plotted observed and expected values over predefined MELD strata (6–14, 15–19, 20–24, 25–29, 30–34, 35–40) to visually assess calibration. Analyses were repeated in 3 subgroups: exclusively in females, patients without HCC, and those with a low eGFR (CKD-EPI <90 mL/min/1.73m²). A threshold of $P < 0.05$ was considered statistically significant. All statistical analyses were performed using Stata, version 16.1 (StataCorp, TX).

RESULTS

Baseline Characteristics

Over the study period, 1108 patients were waitlisted with baseline characteristics and outcomes summarized in Tables 1 and 2, respectively. Median age was 53.5 y (IQR 48–59), and there was a male predominance of 74.9%. The predominant primary etiologies for cirrhosis were hepatitis C virus (HCV,

42.9%), alcohol-associated (17.0%), and hepatitis B virus (10.6%). HCC was present in 410 patients (37.0%). Median days on the WL were 143 (IQR 49–315), during which 853 patients (77.0%) underwent LT, 176 (15.9%) died, and 79 (7.1%) were delisted (Table 2). Baseline comparisons between survivors and nonsurvivors at 90 d are presented in Table S1 ISDC, <http://links.lww.com/TXD/A432>). The number of transplant listings over the years of the study period is presented in Figure S1 (SDC, <http://links.lww.com/TXD/A432>).

Model Discrimination

AUROC

The AUROCs of all models in predicting the primary outcome approached or exceeded 0.90, indicating excellent performance (Table 3). Using the DeLong method, there was no difference between the AUROC of MELD when compared with any other model ($P > 0.05$ for all comparisons). The optimal cutoff score to predict death at 90 d was 19 for MELD, 22 for MELD-Na, 18 for MELD-GRAIL, and 17 for MELD-GRAIL-Na. Using these cutoffs, MELD-Na had the highest sensitivity (0.83), and MELD, MELD-GRAIL, and MELD-GRAIL-Na were the most specific (0.85) for death at 90 d. All models had an NPV of over 97% using the aforementioned cutoffs.

In terms of secondary outcomes, the performance of all models was excellent for predicting death at 30 d; however, performance was lower for death at 1 y with AUROCs ranging from 0.80 to 0.81 (Table 3). There was no difference in AUROC between MELD in comparison with other models for either secondary outcome. Optimal cutoff score values were higher for death at 30 d than at 90 d. Using these cutoffs, all scores had an NPV of over 99% for death at 30 d. For

TABLE 1.
Baseline characteristics of patients at registration on the liver transplant waiting list

Demographics	
Median age, y (IQR)	53.5 (48–59.0)
Female, (%)	278 (25.1)
Caucasian, (%)	818 (73.8)
Primary indication for liver transplant, n (%)	
HCV	475 (42.9)
ALD	188 (17.0)
HBV	118 (10.6)
NASH	88 (7.9)
PSC	87 (7.9)
PBC	49 (4.4)
Additionally with HCC	410 (37.0)
Laboratory values, mean (±SD)	
Albumin (g/L)	32.9 (±6.9)
Bilirubin (mg/dL)	6.2 (±8.5)
Creatinine (mg/dL)	0.9 (±0.5)
Urea (mmol/L)	7.2 (±6.1)
INR	1.7 (±0.6)
Sodium (mmol/L)	136.2 (±5.2)
Prognostic scores, median (IQR)	
MELD	16.4 (12.4–20.6)
MELD-Na	18.0 (12.7–23.4)
MELD-GRAIL	14.9 (11.4–19.3)
MELD-GRAIL-Na	13.8 (10.4–18.0)

ALD, alcohol-associated liver disease; GRAIL, GFR assessment in liver disease; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; INR, international normalized ratio; IQR, interquartile range; MELD, model for end-stage liver disease; NASH, nonalcoholic steatohepatitis; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis.

TABLE 2.
Outcomes of patients on the liver transplant waiting list

Primary outcome	
Outcome at 90 d, n (%)	
On transplant waitlist	694 (62.6)
Died	66 (6.0)
Transplanted	336 (30.3)
Delisted	12 (1.1)
Secondary outcomes	
Outcome at 30 d, n (%)	
On transplant waitlist	913 (82.4)
Died	41 (3.7)
Transplanted	153 (13.8)
Delisted	1 (0.1)
Outcome at 1 y, n (%)	
On transplant waitlist	228 (20.6)
Died	140 (12.6)
Transplanted	687 (62.0)
Delisted	53 (4.8)
Outcome at last follow-up	
Time on waitlist, median days (IQR)	143 (49–315)
Time to death, median days (IQR)	96.5 (25–220)
Time to death or disease progression, median days (IQR)	143 (42–287)
Died, n (%)	176 (15.9)
Delisted, n (%)	79 (7.1)
Transplanted, n (%)	853 (77.0)

IQR, interquartile range.

TABLE 3.
AUROC comparison of models to predict primary and secondary outcomes

Primary outcome								
Death at 90 d (n = 762)								
Score	AUROC	95% CI	P ^a	Optimal cutoff (Youden index)	Sensitivity	Specificity	PPV (%)	NPV (%)
MELD	0.89	0.85-0.93	—	19	0.80	0.85	34.0	97.8
MELD-Na	0.91	0.87-0.94	0.28	22	0.83	0.84	32.4	98.1
MELD-GRAIL	0.89	0.85-0.93	0.83	18	0.82	0.85	32.7	98.0
MELD-GRAIL-Na	0.89	0.85-0.93	0.81	17	0.82	0.85	32.9	98.0
Secondary outcomes								
Death at 30 d (n = 951)								
MELD	0.90	0.84-0.95	—	20	0.87	0.80	15.2	99.3
MELD-Na	0.92	0.88-0.96	0.22	25	0.84	0.89	24.6	99.3
MELD-GRAIL	0.91	0.85-0.96	0.28	21	0.82	0.88	20.7	99.1
MELD-GRAIL-Na	0.91	0.85-0.96	0.30	19	0.82	0.88	20.5	99.1
Death at 1 y (n = 361)								
MELD	0.80	0.75-0.85	—	17	0.69	0.77	63.2	81.6
MELD-Na	0.81	0.76-0.86	0.15	18	0.72	0.77	63.5	82.6
MELD-GRAIL	0.81	0.76-0.85	0.16	17	0.60	0.88	72.2	79.1
MELD-GRAIL-Na	0.81	0.76-0.85	0.16	14	0.69	0.79	64.3	81.5

^aComparison with MELD AUROC using the DeLong method.

AUROC, area under the receiver operating characteristic curve; CI, confidence interval; GRAIL, GFR assessment in liver disease; MELD, model for end-stage liver disease; NPV, negative predictive value; PPV, positive predictive value.

TABLE 4.
Cox regression for predictors of death at 90 d

Variable	Univariable			Multivariable		
	HR	95% CI	P	HR	95% CI	P
Age (per year increase)	1.00	0.97-1.03	0.90			
Etiology	0.93	0.81-1.05	0.23			
Male sex	0.86	0.48-1.53	0.69			
Creatinine (mg/dL)	3.01	2.31-3.92	<0.01	1.586	1.085-2.317	0.017
Bilirubin (mg/dL)	1.08	1.06-1.09	<0.01	1.060	1.043-1.077	<0.001
INR	2.48	2.04-3.02	<0.01	1.741	1.324-2.290	<0.001
Sodium (mmol/L)	0.88	0.84-0.92	<0.01	0.925	0.889-0.963	<0.001
Urea (mmol/L)	1.09	1.07-1.10	<0.01			
Albumin (g/L)	0.97	0.94-1.01	0.13			

CI, confidence interval; HR, hazard ratio; INR, international normalized ratio.

death at 1 y, optimal score cutoffs were lower than those for the primary outcome.

Table S2 (SDC, <http://links.lww.com/TXD/A432>), demonstrates model performance for the primary outcome in each of the subgroups. There was no difference in AUROC between MELD and any other score in each of the subgroup analyses. Notably, AUROCs for females were lower, although the optimal cutoffs to predict death at 90 d were similar compared with the overall cohort (20 for MELD, 22 for MELD-Na, 18 for MELD-GRAIL, and 17 for MELD-GRAIL-Na).

Survival Analysis

The results of univariable and multivariable Cox regression are summarized in Table 4. Baseline predictors of death at 90 d on univariable Cox regression were bilirubin, creatinine, INR, sodium and urea. Of note, albumin, sex, and age were not associated with the primary outcome. On multivariable analysis, a model (model 1) containing bilirubin (HR 1.06, 95% CI 1.04-1.08), creatinine (HR 1.59, 95% CI 1.09-2.32), INR (HR

1.74, 95% CI 1.32-2.29), and sodium (HR 0.93, 95% CI 0.89-0.96) best predicted the primary outcome (Table 4). Model discrimination using Harrell's C-index found no appreciable difference in the between existing models or model 1 (Table 5).

All calculated eGFR formulae were predictive of 90-d mortality on univariable Cox regression (Table S3, SDC, <http://links.lww.com/TXD/A432>). Substituting creatinine for MDRD-6 and GRAIL remained significant on multivariable analysis; however, other measures of eGFR were not (Table S4, SDC, <http://links.lww.com/TXD/A432>).

In the subgroups of patients with low eGFR (defined by CKD-EPI eGFR <90 mL/min/1.73 m²) and without HCC, there was no change in the variables associated with mortality up to 90 d on univariable Cox regressions. In females, sodium was the only additional variable not associated with mortality at 90 d (Table S5, SDC, <http://links.lww.com/TXD/A432>). Bilirubin, INR, and sodium were still associated with mortality for most subgroups on multivariable Cox regression; however, INR was not associated with mortality in the subgroup

TABLE 5.

Model discrimination

Model	HR (95% CI)	Harrell's C-index	95% CI	<i>P</i> ^a
MELD	1.19 (1.16-1.23)	0.86	0.82-0.90	—
MELD-Na	1.23 (1.18-1.27)	0.88	0.84-0.92	0.20
MELD-GRAIL	1.24 (1.20-1.28)	0.86	0.82-0.91	0.82
MELD-GRAIL-Na	1.24 (1.20-1.29)	0.86	0.82-0.91	0.79
Model 1	1.21 (1.15-1.28)	0.86	0.81-0.91	0.81

^aHarrell's C-index pairwise comparisons to MELD, *P* > 0.05 for all other pairwise comparisons between models.

CI, confidence interval; GRAIL, GFR assessment in liver disease; HR, hazard ratio; MELD, model for end-stage liver disease.

of a low eGFR, creatinine was not associated with mortality in those without HCC, and both creatinine and sodium were not associated with mortality in females (Table S5, SDC, <http://links.lww.com/TXD/A432>).

Model Calibration

All scores were well calibrated according to the Grønnesby and Borgan chi-square test (*P* > 0.05 for all), excluding model 1 (*P* = 0.002). Despite this, calibration risk deciles were different between scores (Figure S2, SDC, <http://links.lww.com/TXD/A432>).

Visual assessment of predicted versus observed survival (Figure 1) revealed that all models were accurate in predicting

mortality at lower scores. Divergence between predicted and observed mortality curves at scores above 25 was noted for MELD, MELD-GRAIL, and MELD-GRAIL-Na; however, this only occurred at scores above 30 for MELD-Na, suggesting better calibration (Figure 1). Expected and observed mortalities for all models were divergent at a score of 35 to 40 because of lower numbers of patients in this range and the high rate of transplantation within 90 d.

DISCUSSION

In our cohort of 1108 patients listed for LT, we found no significant difference between MELD-GRAIL, MELD-GRAIL-Na, and MELD-Na compared with MELD in predicting 90-d WL mortality. Moreover, there were no differences between models in predicting 30-d or 1-y mortality or in the subgroups of women, of low eGFR, or excluding HCC. All models performed well with excellent discrimination and calibration. Furthermore, we identified pragmatic cutoffs for each model to stratify those at high and low risk of death within 30 and 90 d. To our knowledge, this is the first external validation study of the discrimination and calibration of MELD-GRAIL and MELD-GRAIL-Na.

Our findings contrast with data from Asrani et al,² who used a large Scientific Registry of Transplant Recipients cohort of over 17 000 patients from 2014 to 2015 to demonstrate that MELD-GRAIL-Na had better discrimination than MELD

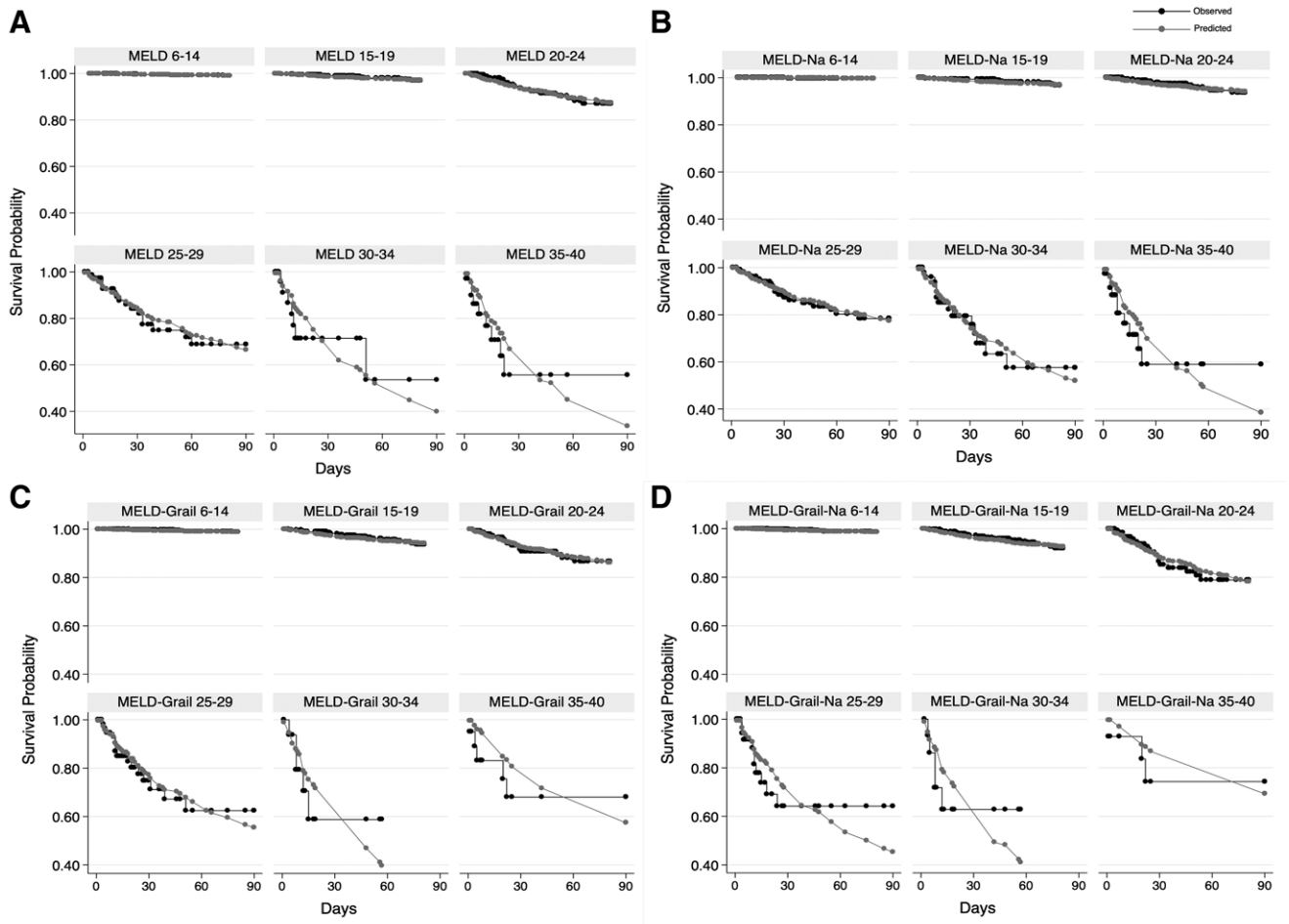


FIGURE 1. Predicted and observed survival for model (A) MELD, (B) MELD-Na, (C) MELD-GRAIL, and (D) MELD-GRAIL-Na. GRAIL, GFR assessment in liver disease; MELD, model for end-stage liver disease.

and MELD-Na for 90-d WL overall mortality (C-index 0.83 versus 0.82 versus 0.81, respectively, $P < 0.001$). This was also consistent for the subsets of sicker patients (defined as MELD ≥ 30) and females; however, important distinctions exist between study cohorts beyond the inherent differences of international registry data from another healthcare system. The Asrani et al cohort had a larger proportion of females (37.8% versus 25.1% in our cohort) and disparate proportions of etiologies of cirrhosis (eg, for HCV: 18.3%, versus 42.9% in our cohort). Despite this, we did not find any differences in the performance of any model in females and did not find etiology to be predictive of mortality on Cox regression in our cohort. Mortality in our cohort was higher than that of Asrani et al (6.8% versus 4.3%). Of note, the cohorts had similar ethnic diversity (73.8% versus 70.5% Caucasian), a similar median age (53.5 versus 57 y), and a similar median MELD score (16.4 versus 17). Additionally, Asrani et al used the primary outcome of WL mortality or delisting within 3 mo of listing, compared with our more robust primary endpoint of mortality that also included those who died within 30 d of delisting; however, when we used the primary outcome of Asrani et al in our cohort, there remained no difference between any models at 30 d, 90 d, or 1 y (data not presented). Recently, Woods et al assessed gender-based differences in mortality prediction scores including GRAIL in a larger cohort from the Scientific Registry of Transplant Recipients; however, they did not specifically address discrimination or calibration.¹⁶

Our study cohort spans from January 1998 to July 2019. During this period, there have been some significant changes in LT WL management in Australia. The cohort studied by Asrani et al coincides with the period when interferon-free direct-acting antiviral (DAA) regimens for HCV were emerging in the United States.¹⁷ Universal access to DAAs in Australia began on March 31, 2016, and before this, compassionate access began from 2014 for those with MELD ≥ 15 . Hence, our cohort includes WL registrants both before and after access to DAA therapy. Furthermore a “Share 35” policy was initiated in the United States in June 2013¹⁸ and in Australia in February 2016.^{10,19} The implementation of the “Share 35” policy in our center has enabled patients with MELD score ≥ 35 to be nationally prioritized for organs across Australia in New Zealand, which has reduced WL mortality^{11,20}; however, from February 2016 to the end of 2018, only 55 patients had received LT in Australia under this policy.¹⁹

We found that MELD, MELD-Na, MELD-GRAIL, and MELD-GRAIL-Na scores of 19, 22, 18, and 17, respectively, optimally stratified patients in to high- and low-risk groups of death within 90 d. Specifically, scores below these cutoffs had an NPV of 97% or greater for 90-d mortality. The cutoffs for MELD and MELD-Na scores are consistent with prior studies.^{20,21} Similarly, score cutoffs below MELD 20, MELD-Na 25, MELD-GRAIL 21, and MELD-GRAIL-Na 19 had a 99% NPV for death at 30 d. These cutoffs could be used to stratify patients for further prioritization on the WL, and those potentially could wait longer than 90 d, provided the appropriate cutoff is used for the correct model.

We demonstrated that bilirubin, creatinine, INR, and sodium are associated with WL mortality, which is consistent with prior studies.^{4,22-27} The additive effect of sodium with regard to mortality is consistent with other analyses, including a recent analysis of the Eurotransplant registry.²⁸ Indeed,

MELD-Na appeared to be better calibrated in our cohort at scores above 25 than in other models on visual inspection of observed and predicted survival. Age has been proposed as a risk factor^{21,23} but has not consistently been shown^{22,24} to be predictive of mortality. Similarly, female sex has been associated with an underestimation of mortality by MELD score, but the magnitude of this association is inconsistent.^{29,30} Albumin has not been shown to be consistently predictive of WL mortality.^{31,32} MELD-GRAIL and MELD-GRAIL-Na scores use urea and albumin in the formula. In our cohort, urea was significant only on univariable Cox regression but was not independently associated with mortality on multivariable analysis. Conversely, albumin was not associated with mortality on univariable or multivariable Cox regression. Therefore, the fact that MELD-GRAIL and MELD-GRAIL-Na scores did not improve the predictive capacity for mortality may be due to the limited predictive capacity of urea and albumin in our cohort.

We suspect that, as the GRAIL-based models were derived in a large registry, these models may indeed be overfitted or at least not have meaningful improvements in discriminatory ability when applied to other settings. This is suggested by the aforementioned numerically small but statistically significant differences in Harrell’s C-index in the Asrani et al study. Furthermore, although the overall calibration of all models was good, the GRAIL-based models appeared to align less with Kaplan-Meier observed survival at higher scores (≥ 25). We therefore suggest that the additional complexity of the GRAIL-based models does not add any additional specific prognostic value compared with that of MELD or MELD-Na in our setting. This has implications for the applicability of the GRAIL-based models in nonnational organ allocation systems or in smaller organ sharing networks than United Network for Organ Sharing. Indeed, we would suggest that prognostic models derived from large registries undergo external validation before being applied in other settings.

This study has several strengths. We used data from well-characterized patients listed over 20 y from a large volume single LT center. The cohort reflects contemporaneous trends in LT and WL management over this period, and therefore, our findings may be more generalizable to similar settings. Data were extracted from a robust database³³ that is subject to quality audits by dedicated data managers, with no missing data from the 1108 patients included in this study. Data from a single LT center also remove center-level variation in practices and waitlist mortality, which has been shown to influence outcomes in candidates with the same biological MELD.³⁴ Furthermore, we used thorough statistical methods to comparatively assess both the discrimination and calibration of models in the overall cohort and in specific subgroups. Additionally, we identified model cutoffs to stratify those at higher risk of death. We found that variables associated with mortality in our cohort did not deviate from the literature.

However, our study has some limitations. First, as a single-center study, our cohort was smaller than the initial MELD-GRAIL cohort, which poses the question of type II error, and may suggest the study was underpowered to detect a difference. Nonetheless, with >1108 patients over 20 y, this still indicates either that there is no difference between the scores in our context or that the difference is so small that the benefit

is negligible from the point of view of a single-center allocation. Furthermore, our multivariate Cox model did not suggest an independent risk associated with the additional factors included in MELD-GRAIL derivation (albumin and urea). As such, though it would be of interest to validate these findings in a multicenter study, we still feel this suggests that MELD-GRAIL and MELD-GRAIL-Na need further validation and possibly calibration before they are adopted internationally, as do further models such as MELD 3.0.³⁵ Second, because our cohort was predominantly male and of Caucasian ethnicity, this limits the findings to other centers with distinct demographics and particularly the analysis of mortality in females and non-Caucasian ethnicities. Third, we did not attempt to derive our own calibrated predictive model because of the lack of unique variables associated with mortality on Cox regression and the lack of difference in performance between the existing models. Furthermore, we did not analyze the performance of models over time, as all models had excellent discrimination and calibration in the overall cohort, and because of the low numbers of patients with MELD score ≥ 35 , we did not analyze this group as a subset. Of note, we did not attempt to account for informative censoring in our setting, which is in line with the MELD-GRAIL study and further risk-based scoring analyses.^{3,35}

We comprehensively investigated the performance of MELD-GRAIL, MELD-GRAIL-Na, MELD-Na, and MELD scores and found that all models had excellent discrimination and calibration for predicting 90-d WL mortality. We found that there was no difference in the performance of these scores compared with MELD. Furthermore, there were no differences seen in subgroup analyses of women, those without HCC, and patients with low eGFR. The calculation of GRAIL-based models did not provide any additional prognostic benefit over MELD or MELD-Na in our setting. We suggest that prognostic models derived from large registries should be externally validated to demonstrate superior discrimination and calibration to the standard of care before being routinely adopted in other settings.

REFERENCES

1. Tschuor C, Ferrarese A, Kuemmerli C, et al.; Liver Allocation Study Group. Allocation of liver grafts worldwide—is there a best system? *J Hepatol.* 2019;71:707–718.
2. Asrani SK, Jennings LW, Kim WR, et al. MELD-GRAIL-Na: glomerular filtration rate and mortality on liver-transplant waiting list. *Hepatology.* 2020;71:1766–1774.
3. Francoz C, Prié D, Abdelrazek W, et al. Inaccuracies of creatinine and creatinine-based equations in candidates for liver transplantation with low creatinine: impact on the model for end-stage liver disease score. *Liver Transpl.* 2010;16:1169–1177.
4. Kim WR, Lake JR, Smith JM, et al. OPTN/SRTR 2016 annual data report: liver. *Am J Transplant.* 2018;18 (Suppl 1):172–253.
5. Yang JD, Larson JJ, Watt KD, et al. Hepatocellular carcinoma is the most common indication for liver transplantation and placement on the waitlist in the United States. *Clin Gastroenterol Hepatol.* 2017;15:767–775.e3.
6. Flemming JA, Kim WR, Brosgart CL, et al. Reduction in liver transplant wait-listing in the era of direct-acting antiviral therapy. *Hepatology.* 2017;65:804–812.
7. Calzadilla-Bertot L, Jeffrey GP, Jacques B, et al. Increasing incidence of nonalcoholic steatohepatitis as an indication for liver transplantation in Australia and New Zealand. *Liver Transpl.* 2019;25:25–34.
8. Asrani SK, Jennings LW, Trotter JF, et al. A model for glomerular filtration rate assessment in liver disease (GRAIL) in the presence of renal dysfunction. *Hepatology.* 2019;69:1219–1230.
9. McCaughan GW, Munn SR. Liver transplantation in Australia and New Zealand. *Liver Transpl.* 2016;22:830–838.
10. The Transplantation Society of Australia and New Zealand. *Clinical guidelines for organ transplantation from deceased donors.* Version 1.6. TSANZ; 2021. Available at https://tsanz.com.au/storage/documents/TSANZ_Clinical_Guidelines_Version-16_11052021.pdf. Accessed May 2021.
11. Wiesner R, Edwards E, Freeman R, et al; United Network for Organ Sharing Liver Disease Severity Score Committee. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology.* 2003;124:91–96.
12. Policy and system changes effective January 11, 2016, adding serum sodium to MELD calculation UNOS. 2021. Available at <https://unos.org/news/policy-and-system-changes-effective-january-11-2016-adding-serum-sodium-to-meld-calculation/>. Accessed August 22, 2021.
13. Levey AS, Bosch JP, Lewis JB, et al. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med.* 1999;130:461–470.
14. Levey AS, Stevens LA, Schmid CH, et al; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150:604–612.
15. Grønnesby JK, Borgan O. A method for checking regression models in survival analysis based on the risk score. *Lifetime Data Anal.* 1996;2:315–328.
16. Wood NL, VanDerwerken D, Segev DL, et al. Correcting the sex disparity in MELD-Na. *Am J Transplant.* 2021;21:3296–3304.
17. Majumdar A, Kitson MT, Roberts SK. Systematic review: current concepts and challenges for the direct-acting antiviral era in hepatitis C cirrhosis. *Aliment Pharmacol Ther.* 2016;43:1276–1292.
18. Kwong AJ, Goel A, Mannalithara A, et al. Improved posttransplant mortality after share 35 for liver transplantation. *Hepatology.* 2018;67:273–281.
19. Fink M, Gow P, Balderson G, et al. Introduction of share 35 inter-regional allocation for high MELD liver transplant waiting list patients in Australia and New Zealand. *Transplantation.* 2018;102. doi:10.1097/01.tp.0000542964.77501.4a
20. Lee J, Lee JG, Jung I, et al; Advisory Committee on Improving Liver Allocation. Development of a Korean liver allocation system using model for end stage liver disease scores: a nationwide, multicenter study. *Sci Rep.* 2019;9:7495.
21. Biggins SW, Kim WR, Terrault NA, et al. Evidence-based incorporation of serum sodium concentration into MELD. *Gastroenterology.* 2006;130:1652–1660.
22. Ruf AE, Kremers WK, Chavez LL, et al. Addition of serum sodium into the MELD score predicts waiting list mortality better than MELD alone. *Liver Transpl.* 2005;11:336–343.
23. Barber K, Madden S, Allen J, et al; United Kingdom Liver Transplant Selection and Allocation Working Party. Elective liver transplant list mortality: development of a United Kingdom end-stage liver disease score. *Transplantation.* 2011;92:469–476.
24. Weismüller TJ, Negm A, Becker T, et al. The introduction of MELD-based organ allocation impacts 3-month survival after liver transplantation by influencing pretransplant patient characteristics. *Transpl Int.* 2009;22:970–978.
25. Gorgen A, Rosales R, Sadler E, et al. Patterns and predictors of mortality after waitlist dropout of patients with hepatocellular carcinoma awaiting liver transplantation. *Transplantation.* 2019;103:2136–2143.
26. Fink MA, Berry SR, Gow PJ, et al. Risk factors for liver transplantation waiting list mortality. *J Gastroenterol Hepatol.* 2007;22(1):119–124. Erratum in: *J Gastroenterol Hepatol.* 2007;22:457.
27. Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. *N Engl J Med.* 2008;359:1018–1026.
28. Goudsmit BFF, 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.* 2021;21:229–240.
29. Allen AM, Heimbach JK, Larson JJ, et al. Reduced access to liver transplantation in women: role of height, MELD exception scores, and renal function underestimation. *Transplantation.* 2018;102:1710–1716.
30. Melk A, Babitsch B, Borchert-Mörlins B, et al. Equally interchangeable? How sex and gender affect transplantation. *Transplantation.* 2019;103:1094–1110.
31. Myers RP, Shaheen AA, Faris P, et al. Revision of MELD to include serum albumin improves prediction of mortality on the liver transplant waiting list. *PLoS One.* 2013;8:e51926.

32. Porrett PM, Baranov E, ter Horst M. Serum hypoalbuminemia predicts late mortality on the liver transplant waiting list. *Transplantation*. 2015;99:158–163.
33. Pandya K, Sastry V, Panlilio MT, et al. Differential impact of extended criteria donors after brain death or circulatory death in adult liver transplantation. *Liver Transpl*. 2020;26:1603–1617.
34. Ishaque T, Kernodle AB, Motter JD, et al. MELD is MELD is MELD? Transplant center-level variation in waitlist mortality for candidates with the same biological MELD. *Am J Transplant*. 2021;21:3305–3311.
35. Kim WR, Mannalithara A, Heimbach JK, et al. MELD 3.0: the model for end-stage liver disease updated for the modern era. *Gastroenterology*. 2021;161:1887–1895.e4.