Development and validation of the Gender-Equity Model for Liver Allocation (GEMA) to prioritise candidates for liver transplantation: a cohort study

Manuel Luis Rodríguez-Perálvarez, Antonio Manuel Gómez-Orellana*, Avik Majumdar, Michael Bailey, Geoffrey W McCaughan, Paul Gow, Marta Guerrero, Rhiannon Taylor, David Guijo-Rubio, César Hervás-Martínez, Emmanuel A Tsochatzis*

Summary

Background The Model for End-stage Liver Disease (MELD) and its sodium-corrected variant (MELD-Na) have created gender disparities in accessing liver transplantation. We aimed to derive and validate the Gender-Equity Model for liver Allocation (GEMA) and its sodium-corrected variant (GEMA-Na) to amend such inequities.

Methods In this cohort study, the GEMA models were derived by replacing creatinine with the Royal Free Hospital glomerular filtration rate (RFH-GFR) within the MELD and MELD-Na formulas, with re-fitting and re-weighting of each component. The new models were trained and internally validated in adults listed for liver transplantation in the UK (2010–20; UK Transplant Registry) using generalised additive multivariable Cox regression, and externally validated in an Australian cohort (1998–2020; Royal Prince Alfred Hospital [Australian National Liver Transplant Unit] and Austin Hospital [Victorian Liver Transplant Unit]). The study comprised 9320 patients: 5762 patients for model training, 1920 patients for internal validation, and 1638 patients for external validation. The primary outcome was mortality or delisting due to clinical deterioration within the first 90 days from listing. Discrimination was assessed by Harrell's concordance statistic.

Findings 449 (5·8%) of 7682 patients in the UK cohort and 87 (5·3%) of 1638 patients in the Australian cohort died or were delisted because of clinical deterioration within 90 days. GEMA showed improved discrimination in predicting mortality or delisting due to clinical deterioration within the first 90 days after waiting list inclusion compared with MELD (Harrell's concordance statistic 0·752 [95% CI 0·700–0·804] *vs* 0·712 [0·656–0·769]; p=0·001 in the internal validation group and 0·761 [0·703–0·819] *vs* 0·739 [0·682–0·796]; p=0·036 in the external validation group), and GEMA-Na showed improved discrimination compared with MELD-Na (0·766 [0·715–0·818] *vs* 0·742 [0·686–0·797]; p=0·0058 in the internal validation group and 0·774 [0·720–0·827] *vs* 0·745 [0·690–0·800]; p=0·014 in the external validation group). The discrimination capacity of GEMA-Na was higher in women than in the overall population, both in the internal (0·802 [0·716–0·888]) and external validation cohorts (0·796 [0·698–0·895]). In the pooled validation cohorts, GEMA resulted in a score change of at least 2 points compared with MELD in 1878 (52·8%) of 3558 patients (25·0% upgraded and 27·8% downgraded). GEMA-Na resulted in a score change of at least 2 points compared with MELD-Na in 1836 (51·6%) of 3558 patients (32·3% upgraded and 19·3% downgraded). In the whole cohort, 3725 patients received a transplant within 90 days of being listed. Of these patients, 586 (15·7%) would have been differently prioritised by GEMA compared with MELD; 468 (12·6%) patients would have been differently prioritised by GEMA-Na compared with MELD-Na. One in 15 deaths could potentially be avoided by using GEMA instead of MELD and one in 21 deaths could potentially be avoided by using GEMA-Na instead of MELD-Na.

Interpretation GEMA and GEMA-Na showed improved discrimination and a significant re-classification benefit compared with existing scores, with consistent results in an external validation cohort. Their implementation could save a clinically meaningful number of lives, particularly among women, and could amend current gender inequities in accessing liver transplantation.

Funding Junta de Andalucía and EDRF.

Introduction

Liver organ allocation in most organ sharing networks is based on the principle of urgency, in which the sickest patients are granted highest priority to prevent mortality within the waiting list. The Model for End-stage Liver Disease (MELD) incorporating serum sodium (MELD-Na)[1] is used for deceased donor waiting list prioritisation worldwide.[2] The MELD-Na score comprises four parameters: serum creatinine, total serum bilirubin, international normalised ratio (INR), and serum sodium.

Lancet Gastroenterol Hepatol 2022

Published Online December 14, 2022 https://doi.org/10.1016/S2468-1253(22)00354-5

*Joint first authors

Department of Hepatology and Liver Transplantation, Hospital Universitario Reina Sofía, IMIBIC, Córdoba, Spain (M L Rodríguez-Perálvarez PhD, M Guerrero MD); **Department of Medicine** (M L Rodríguez-Perálvarez) **and Department of Computer Science and Numerical Analysis** (A M Gómez-Orellana MCS, D Guijo-Rubio PhD, Prof C Hervás-Martínez PhD), **University of Córdoba, Córdoba, Spain; Centro de investigación biomédica en red de enfermedades hepáticas y digestivas (CIBERehd), Madrid, Spain** (M L Rodríguez-Perálvarez, M Guerrero); **AW Morrow Gastroenterology and Liver Centre and Australian National Liver Transplant Unit, Royal Prince Alfred Hospital, Sydney, NSW, Australia** (A Majumdar PhD, Prof G W McCaughan PhD); **Central Clinical School, The University of Sydney, Sydney, NSW, Australia** (A Majumdar, Prof G W McCaughan); **Victorian Liver Transplant Unit, Austin Health, Melbourne, VIC, Australia** (A Majumdar, P Gow MD); **The University of Melbourne, Melbourne, VIC, Australia** (A Majumdar, P Gow); **Australian and New Zealand Intensive Care Research Centre (ANZIC RC), Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia** (Prof M Bailey PhD); **ANZICS Centre for Outcome and Resource Evaluation (CORE), Melbourne, VIC, Australia** (Prof M Bailey); **Liver Injury and Cancer Program, Centenary Institute, Sydney, NSW, Australia** (Prof G W McCaughan);

Department of Statistics and Clinical Studies, NHS Blood and Transplant, Stoke Gifford, Bristol, UK (R Taylor PhD); **School of Computing Sciences, University of East Anglia, Norwich, UK** (D Guijo-Rubio); **Sheila Sherlock Liver Unit and UCL Institute for Liver and Digestive Health, Royal Free Hospital, London, UK** (Prof E A Tsochatzis PhD)

Correspondence to: Prof Emmanuel A Tsochatzis, Sheila Sherlock Liver Unit and UCL Institute for Liver and Digestive Health, Royal Free Hospital, London NW3 2QG, UK
e.tsochatzis@ucl.ac.uk

Research in context

Evidence before this study

The Model for End-stage Liver Disease (MELD) was first adopted in the USA in 2002 and spread rapidly to other countries and organ allocation systems. In 2008, an update of the MELD score was implemented by correcting for serum sodium (MELD-Na). These models, based on the principle of urgency, allowed a reduction of waiting list mortality rates among liver transplant candidates. However, gender disparities became evident; the likelihood of death or delisting in women compared with men after the implementation of the MELD score were as high as 30%. We searched MEDLINE, Embase, and Science Citation Index databases for studies published from Jan 1, 2008, to Dec 31, 2021, evaluating recipient-based allocation systems specifically oriented to amend gender disparities for accessing liver transplantation. We used different combinations of the following keywords or equivalent free-text terms, without language restrictions: ("gender" OR "sex" OR "women" OR "disparities") AND ("waiting list") AND ("liver transplantation") AND ("MELD" OR "MELD-Na" OR "creatinine"). In addition, within the past 3 years, we hand-searched literature reviews for additional relevant papers. We identified two different models: MELD-GRAIL, which replaced serum creatinine by glomerular filtration rate in liver disease using data from the USA Scientific Registry of Transplant Recipients (2014–15); and MELD 3.0, which added female sex and serum albumin to MELD-Na, without replacement of serum creatinine, using data from the USA Organ Procurement and Transplant Network (2016–18). Both models reported an improved discriminative capacity over MELD-Na, but with a modest increase in the Harrell's concordance statistic (difference 0·01 for MELD-GRAIL and 0·007 for MELD 3.0 *vs* MELD-Na). Neither model has been externally validated and most patients retained the same score when comparing with MELD-Na: only 16·7% of patients were reported to change prioritisation score with MELD-GRAIL and 15·4% of patients were reported to change with MELD 3.0.

Added value of this study

In this study involving 9320 liver transplant candidates from two different countries (UK and Australia) and organ allocation systems, we derived and externally validated a new model that replaces serum creatinine with a validated formula for estimating cirrhosis-specific glomerular filtration rate in the MELD and MELD-Na equations, with re-fitting and re-weighting of the remaining model components. The Gender-Equity Model for liver Allocation (GEMA), and its sodium-corrected variant (GEMA-Na), showed improved discrimination compared with MELD, MELD-Na, and MELD 3.0 to predict mortality or delisting due to clinical deterioration within the first 90 days after waiting list inclusion, both in men and women, but with a more pronounced benefit for women. GEMA would change the prioritisation status (with two or more points) in half of liver transplant candidates, providing extra prioritisation to women and to patients with ascites, which is another group historically penalised by the MELD-based prioritisation system. In particular, the implementation of GEMA-Na instead of MELD-Na could potentially avoid one in 19 deaths overall, and one in eight deaths among women. Importantly, to our knowledge, GEMA is the first prioritisation score for 3-month survival in the transplant waiting list derived and validated outside the USA.

Implications of all the available evidence

GEMA showed improved discrimination and a significant re-classification benefit compared with existing scores, with consistent results in an external validation cohort. Its implementation could potentially save a clinically meaningful number of lives, particularly among women, and could amend the current gender inequities in accessing liver transplantation.

The weighting of serum creatinine within MELD-Na makes the model vulnerable to external factors, particularly body muscle mass, thus penalising patients with malnutrition and sarcopenia,[3] which are features strongly associated with worse outcomes among individuals with cirrhosis.[4] A report from the United Network for Organ Sharing database that included 44 388 patients consecutively enlisted for liver transplantation found increased mortality rates or delisting due to clinical worsening in women as compared with men.[5] Indeed, the odds of death or delisting were 30% higher in women compared with men after the implementation of the MELD score.[6] This gender imbalance could be explained by lower serum creatinine concentrations in women than men, which do not reflect their true renal function and result in a systematically lower calculated MELD-Na score.[5,7,8]

To correct the gender imbalance for accessing liver transplantation, serum creatinine should ideally be replaced in the MELD-Na formula by a more accurate measure of renal function. We have developed and validated a formula for estimating cirrhosis-specific glomerular filtration rate (the Royal Free Hospital [RFH]-GFR) in a cohort of patients with cirrhosis, which includes a gender correction, and is more accurate compared with existing glomerular filtration rate formulae.[9] However, the inclusion of RFH-GFR in liver transplantation allocation models has not yet been examined.

We aimed to design and externally validate a new score, the Gender-Equity Model for liver Allocation (GEMA), through the replacement of serum creatinine by RFH-GFR within the MELD and MELD-Na equations, and re-fitting their components, and to explore if the

implementation of GEMA could amend gender disparities for accessing liver transplantation.

Methods

Study design and participants

This cohort study was done following TRIPOD guidelines.[10] The derivation and internal validation cohort comprised adult patients consecutively registered for first elective liver transplantation on the UK Transplant Registry, held by the National Health Service (NHS) Blood and Transplant authority, between April 1, 2010, and March 31, 2020. The external validation cohort included patients enlisted for elective liver transplantation from the two largest Australian transplant units, namely Royal Prince Alfred Hospital (Australian National Liver Transplant Unit) and Austin Hospital (Victorian Liver Transplant Unit), from January 1, 1998, to December 31, 2020.

Paediatric patients aged younger than 17 years at registration, patients with acute liver failure, patients without cirrhosis listed with MELD exceptions, and candidates for re-transplantation or combined organ transplantation were excluded. The sample size calculation is in the appendix (p 1). This study was conducted according to the principles contained in the Declaration of Helsinki. The study protocol and data sharing were evaluated and approved by the NHS Blood and Transplant authority, and the Sydney Local Health District Human Research Ethics Committee.

Outcomes

Patients were followed until transplantation, death, or withdrawal from the waiting list, whichever occurred first. The primary outcome of the study was mortality within the waiting list or delisting due to clinical worsening within the first 90 days from listing as a time dependent outcome. Patients were censored if they remained alive and active on the waiting list at 90 days or if they underwent transplantation or were excluded for reasons other than worsening before that timepoint. Censoring data are presented in the appendix (p 4).

Development of the GEMA models

To construct the GEMA models, the derivation cohort was randomly assigned (3:1) to training and internal validation datasets with a computer-generated number sequence, further stratified by the primary outcome risk. The GEMA models were designed within the training dataset after replacing serum creatinine by RFH-GFR[9] in the original MELD formula (GEMA) and in the MELD-Na formula (GEMA-Na). The RFH-GFR was calculated at inclusion in the waiting list according to the original formula: $45 \cdot 9 \times (\text{creatinine}^{-0 \cdot 836}) \times (\text{urea}^{-0 \cdot 229}) \times (\text{INR}^{-0 \cdot 113}) \times (\text{age}^{-0 \cdot 129}) \times (\text{sodium}^{0 \cdot 972}) \times 0 \cdot 809$ (if female) $\times 0 \cdot 92$ (if moderate or severe ascites).

Ascites was considered moderate or severe if evident in physical examination according to established guidelines.[11] Smoothing splines were derived for each one of the variables of the GEMA models, using the function generalised additive models with integrated smoothness estimation ("gam"), from the "mgcv" package from R. As the expected relationship between continuous parameters and the primary outcome of the study was not linear in their upper and lower values, variables were capped where appropriate through visual inspection of the splines. The GEMA model was constructed by combining the three capped variables (bilirubin, INR, and RFH-GFR) into a generalised additive multivariable Cox regression model. This method was preferred over competing risk analysis, aligning with previous studies,[12,13] as Cox regression is more suited for estimating survival in the absence of liver transplantation, whereas a competing risk method better estimates survival in the presence of a transplantation. The assumptions underlying Cox regression including proportional hazards are presented in the appendix (pp 2–3). All potential interactions among the covariates of GEMA were tested and kept only if they had a meaningful effect on discrimination. The 25th and 75th quartiles of GEMA were matched to fit the same range as the MELD score (ie, 6–40). The GEMA-Na model resulted after the addition of capped serum sodium as a correction variable of GEMA, similarly, as previously described for MELD-Na.[1] All models were rounded to the nearest integer.

See Online for appendix

Performance of the GEMA models

The GEMA models were compared with the original MELD, MELD-Na, and MELD 3.0 scores[1,12] in terms of discrimination, calibration, and re-classification[14] for the primary outcome. Discrimination refers to the ability of the model to rank patients according to their risk of developing the primary outcome and was assessed by the Harrell's concordance statistic. Harrell's concordance statistics of the different models evaluated were compared using a one-shot non-parametric approach, which does not require resampling as described by Kang and colleagues.[15] The Brier score (ie, mean squared error of predicted probabilities) was used to assess the overall accuracy of the predictions. Calibration and goodness-of-fit refer to the capacity of the model to predict absolute risks in the whole spectrum of disease severity. Calibration was assessed by the goodness-of-fit after stratification of the population in deciles of risk as proposed by D'Agostino and Nam,[16] merging groups to ensure a minimum of two events per group.

Linear and bar calibration plots were constructed to visually represent the agreement between predictions and observations in each group of risk. The re-classification rate was defined as the proportion of patients with a score change of at least 2 when comparing MELD versus GEMA, MELD-Na versus GEMA-Na, or MELD 3.0 versus GEMA-Na. Patients granted at least 2 extra points with the GEMA models were considered upgraded on the list, whereas patients with the same score reduction were considered to be downgraded.

To calculate the number of potential lives saved, we first calculated the number of transplantations performed within 90 days in the whole cohort and considered this number equal to the number of available organs within this period. Subsequently, patients were ranked according to each prioritisation score evaluated and the cohort was stratified in patients who would have been prioritised by both models, and patients who would have been differently prioritised by either of them. The number of potential lives saved resulted from subtracting the number of patients reaching the primary outcome in the GEMA-prioritised group from the number of patients reaching the primary outcome in the MELD-prioritised group, divided by the total number of patients included.[17] The threshold of p less than 0·05 was considered statistically significant. Analyses were performed by using R (version 4.1.2) and SPSS (version 27.0).

Role of the funding source

The funder of the study had no role in the study design, data collection, data analysis, manuscript preparation, or in the decision to publish the study.

Results

9320 patients were included—7682 patients from the UK Transplant Registry for model training and internal validation (derivation cohort), and 1638 patients from Australia (external validation cohort; table 1). Compared with the derivation cohort, the external validation cohort included a higher proportion of patients with hepatitis C, and lower prevalence of alcohol-related liver disease, non-alcoholic fatty liver disease or cryptogenic cirrhosis, and primary sclerosing cholangitis (table 1). Patients in the external validation cohort showed features of more advanced cirrhosis, including higher prevalence of moderate-to-severe ascites and increased MELD,

	Derivation Cohort (n=7682)	External validation cohort (n=1638)	p value
Age, years	53·22 (11·55)	53·52 (9·28)	0·27
Sex	..	..	<0·0001
Men	5104 (66·4%)	1206 (73·6%)	..
Women	2578 (33·6%)	432 (26·4%)	..
Cause of liver disease			
Alcohol	2783 (36·2%)	474 (28·9%)	<0·0001
Hepatitis C	1242 (16·2%)	681 (41·6%)	<0·0001
NAFLD or cryptogenic	1374 (17·9%)	203 (12·4%)	<0·0001
Primary sclerosing cholangitis	808 (10·5%)	87 (5·3%)	<0·0001
Primary biliary cholangitis	633 (8·2%)	142 (8·7%)	0·57
Ascites	..	..	0·0003
No	3285 (42·8%)	616 (37·6%)	..
Mild	1986 (26·0%)	440 (26·9%)	..
Moderate-to-severe	2411 (31·4%)	582 (35·5%)	..
Urea (mmol/L)	5·10 (3·9–7·1)	6 (4–8)	0·0901
Creatinine (µmol/L)	80·13 (35·04)	84·13 (39·14)	<0·0001
RFH-GFR (ml/min)	69·81 (25·04)	66·51 (24·99)	<0·0001
International normalised ratio	1·45 (0·44)	1·63 (0·59)	<0·0001
Bilirubin (µmol/L)	43·95 (24–87)	53 (27–116)	<0·0001
Sodium (mmol/L)	136·24 (4·65)	136·17 (5·04)	0·63
Albumin (g/L)*	31·86 (6·61)	32·26 (6·91)	0·034
MELD	15·03 (5·71)	16·94 (6·94)	<0·0001
MELD-Na	17·25 (6·44)	18·90 (7·65)	<0·0001
MELD 3·0*	17·15 (6·29)	18·80 (7·59)	<0·0001
Primary outcome†	449 (5·8%)	87 (5·3%)	0·40

Data are n (%), median (IQR), or mean (SD). p values were obtained from the following hypothesis contrast tests: χ^2 for frequencies, Student's *t*-test for continuous variables with normal distribution, and Mann-Whitney U test for continuous variables with asymmetric distribution. MELD=Model for End-stage Liver Disease. MELD-Na=MELD corrected by serum sodium. NAFLD=non-alcoholic fatty liver disease. RFH-GFR=Royal Free Hospital glomerular filtration rate. *Albumin and MELD 3.0 were not available for 549 (7·1%) patients from the derivation cohort. †Mortality or delisting due to clinical deterioration within the first 90 days after inclusion in the waiting list.

Table 1: **Descriptive analysis of patients enlisted for liver transplantation stratified into a derivation set from the UK Transplant Registry (2010–20) and an external validation set enlisted in two Australian institutions (1998–2020)**

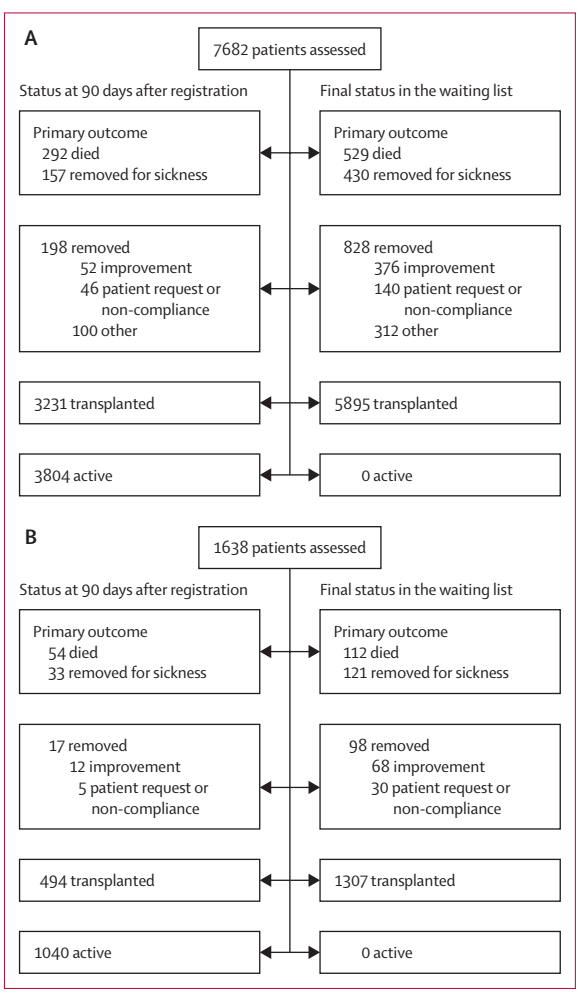


Figure 1: **Study profile**

Outcomes within the waiting list in 7682 patients enlisted for liver transplantation in the UK Transplant Registry (derivation cohort; A), and in 1638 patients from two Australian institutions (external validation cohort; B). Outcome statuses are shown at 90 days after registration and at database closure.

MELD-Na, and MELD 3.0 scores compared with the derivation cohort. Outcomes within the waiting list are presented in figure 1. Mortality or delisting due to clinical deterioration within the first 90 days after waiting list inclusion occurred in 449 (5·8%) of 7682 patients from the derivation cohort and in 87 (5·3%) of 1638 patients from the external validation cohort (p=0·40).

Baseline differences between men and women were evident when the entire study population was analysed by sex (appendix p 5). Women had lower serum creatinine compared with men (72·89 µmol/L [SD 32·14] *vs* 84·62 µmol/L [36·86]; p<0·0001), despite having worse renal function as estimated by RFH-GFR (67·20 mL/min [25·45] *vs* 70·19 mL/min [24·82]; p<0·0001). 1165 (38·7%) of 3010 women and 2560 (40·6%) of 6310 men had undergone transplantation by 90 days (p=0·08). Death or delisting due to clinical worsening at 90 days occurred in 178 (5·9%) of 3010 women and in 358 (5·7%) of 6310 men (p=0·64).

The training (n=5762) and internal validation (n=1920) cohorts were comparable in terms of clinical characteristics and outcomes (appendix p 6). In the training cohort, all four variables of the GEMA models were independent predictors of mortality or delisting due to clinical worsening in the multivariable Cox regression analysis: serum bilirubin (hazard ratio [HR] 1·004, 95% CI 1·004–1·005; p<0·0001), INR (1·619, 1·405–1·867; p<0·0001), RFH-GFR (HR 0·981, 0·976–0·986; p<0·0001), and serum sodium (0·925, 95% CI 0·905–0·945; p<0·0001). Lower and upper bounds of each parameter were defined according to the splines: serum bilirubin (20–550 µmol/L), INR (1–3), RFH-GFR (20–100 mL/min), and serum sodium (122–138 mmol/L; appendix p 12). Interactions between variables in the GEMA model were discarded as they did not provide a meaningful discrimination improvement. The equation of the GEMA model according to generalised additive multivariable Cox regression was expressed as:

$$GEMA = 3{\cdot}777 \times \ln(\text{Bilirubin}) + 7{\cdot}883 \times \ln(\text{INR}) - 8{\cdot}306 \times \ln(\text{RFH-GFR}) + 31{\cdot}932$$

The equation of GEMA-Na after including the correction by serum sodium was expressed as:

$$\text{GEMA-Na} = \text{GEMA} - \text{Na} - [0{\cdot}025 \times \text{GEMA} \times (140 - \text{Na})] + 140$$

Table 2 show Harrell's concordance statistics for the different models tested. In the training cohort, the GEMA model and the GEMA-Na model were better predictors of the mortality or delisting due to clinical deterioration by 90 days than the MELD and the MELD-Na scores, which persisted in the internal validation cohort (table 2). The discriminative capacity of GEMA-Na was also superior to that obtained from MELD 3.0, both in the training cohort and in the internal validation cohort (table 2). These results were consistent in the external validation cohort. The Harrell's concordance statistics for GEMA and GEMA-Na were higher in the subgroup of women than in the overall population in all cohorts, with more pronounced discriminative benefit compared to the MELD equations (table 2). The performance of GEMA and GEMA-Na in patients included in the waiting list with hepatic insufficiency and according to the cause of liver disease is presented in the appendix (p 7). The Brier scores obtained from each model across different study cohorts are presented in the appendix (p 8). The highest accuracies were consistently observed in the GEMA models, both in the whole cohort and in women.

The GEMA models were well calibrated, both in the internal and external validation cohorts (GEMA: χ^2 6·62, degrees of freedom (df) 7; p=0·47 for the internal

	MELD	MELD-Na	MELD 3.0	GEMA		GEMA-Na		
				Hc statistic	p value*	Hc statistic	p value†	p value‡
Training (overall; n=5762)	0·753 (0·723–0·783)	0·783 (0·755–0·810)	0·770 (0·740–0·800)	0·780 (0·751–0·808)	0·0003	0·796 (0·769–0·823)	0·022	p<0·0001
Training (women; n=1955)	0·743 (0·690–0·795)	0·784 (0·739–0·829)	0·766 (0·718–0·815)	0·795 (0·748–0·842)	<0·0001	0·821 (0·781–0·860)	0·0007	p<0·0001
Internal validation (overall; n=1920)	0·712 (0·656–0·769)	0·742 (0·686–0·797)	0·720 (0·657–0·784)	0·752 (0·700–0·804)	0·0010	0·766 (0·715–0·818)	0·0058	0·0014
Internal validation (women; n=623)	0·751 (0·658–0·844)	0·779 (0·688–0·871)	0·763 (0·660–0·867)	0·786 (0·698–0·874)	0·0854	0·802 (0·716–0·888)	0·0866	0·16
External validation (overall; n=1638)	0·739 (0·682–0·796)	0·745 (0·690–0·800)	0·749 (0·696–0·802)	0·761 (0·703–0·819)	0·036	0·774 (0·720–0·827)	0·014	0·008
External validation (women; n=432)	0·736 (0·628–0·844)	0·714 (0·592–0·835)	0·732 (0·625–0·839)	0·789 (0·686–0·892)	0·0044	0·796 (0·698–0·895)	0·0086	0·0068

Data are Harrell's concordance index (95% CI), unless otherwise stated. In each dataset, the subgroup of women was analysed separately. In the training and internal validation cohorts, 549 patients (7·14%) of 7682 did not have albumin data and were excluded from comparisons between MELD 3.0 and GEMA-Na.*p values for comparing discrimination are shown for GEMA versus MELD. †p values for comparing discrimination are shown for GEMA-Na versus MELD-Na. ‡p values for comparing discrimination are shown for GEMA-Na versus MELD 3.0. GEMA=Gender-Equity Model for liver Allocation. GEMA-Na=GEMA corrected by serum sodium. MELD=Model for End-stage Liver Disease. MELD-Na=MELD corrected by serum sodium.

Table 2: **Harrell's concordance statistic for each model in the different cohorts of the study**

validation cohort and χ^2 8·58, df 9; p=0·48 for the external validation cohort; GEMA-Na: χ^2 14·33, df 9; p=0·11 and χ^2 12·59, df 9; p=0·18). For MELD, MELD-Na, and MELD 3.0, good calibration was observed in the external validation cohort (MELD χ^2 6·34, df 8; p=0·61, MELD-Na χ^2 8·96, df 8; p=0·35, and MELD 3.0, χ^2 6·72, df 6; p=0·35), but there was a weaker calibration in the internal validation cohort (MELD χ^2 15·56, df 8; p=0·049, MELD-Na χ^2 25·41, df 9; p=0·003, and MELD 3.0 χ^2 19·05, df 8; p=0·015; figure 2; appendix pp 13–15). As there were insufficient numbers of women in either validation cohorts to enable meaningful calibration, these were combined. GEMA and GEMA-Na and MELD-Na and MELD 3.0 were well calibrated, but MELD showed poorer calibration (χ^2 15·01, df 7; p=0·036; appendix pp 16–17).

Re-classification plots in the pooled validation cohort are shown in figure 3. The GEMA model resulted in a score change of at least 2 points compared with MELD in 1878 (52·8%) of 3558 patients (25·0% upgraded and 27·8% downgraded). The GEMA-Na model resulted in a score change of at least 2 points compared with MELD-Na in 1836 (51·6%) of 3558 patients (32·3% upgraded and 19·3% downgraded). The proportion of patients upgraded by the GEMA models compared with MELD and MELD-Na was higher among women than in the overall population (36·2% for GEMA and 41·3% for GEMA-Na; figure 3). Kaplan-Meier curves showing the cumulative incidence of mortality within the waiting list or delisting for sickness according to different thresholds of GEMA and GEMA-Na are shown in the appendix (p 18).

In the whole cohort, 3725 patients received a liver graft within 90 days of listing. When MELD and GEMA were compared, 586 (15·7%) patients would have been differently prioritised. When comparing MELD-Na and GEMA-Na, 468 (12·6%) patients would have been differently prioritised. Similarly, after excluding 198 transplanted patients with missing albumin data, different prioritisation was observed in 509 (14·4%) of 3527 patients when MELD 3.0 was compared with GEMA-Na. Clinical characteristics of patients differently prioritised are shown in the appendix (pp 9–11). Patients allocated only by GEMA and GEMA-Na had approximately triple the risk of death or delisting for sickness within the first 90 days compared with patients allocated only by MELD, MELD-Na, and MELD 3.0: GEMA *vs* MELD (OR 3·63, 95% CI 2·02–6·53; p<0·0001); GEMA-Na

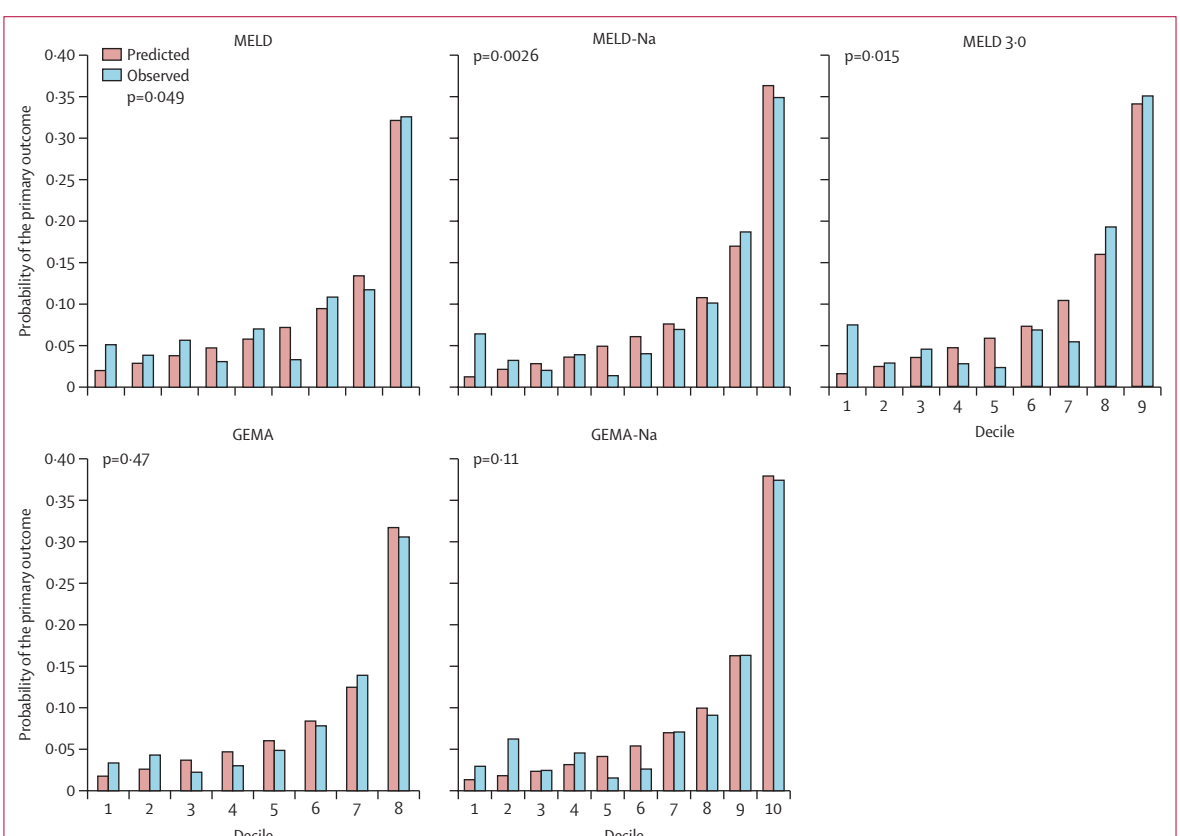
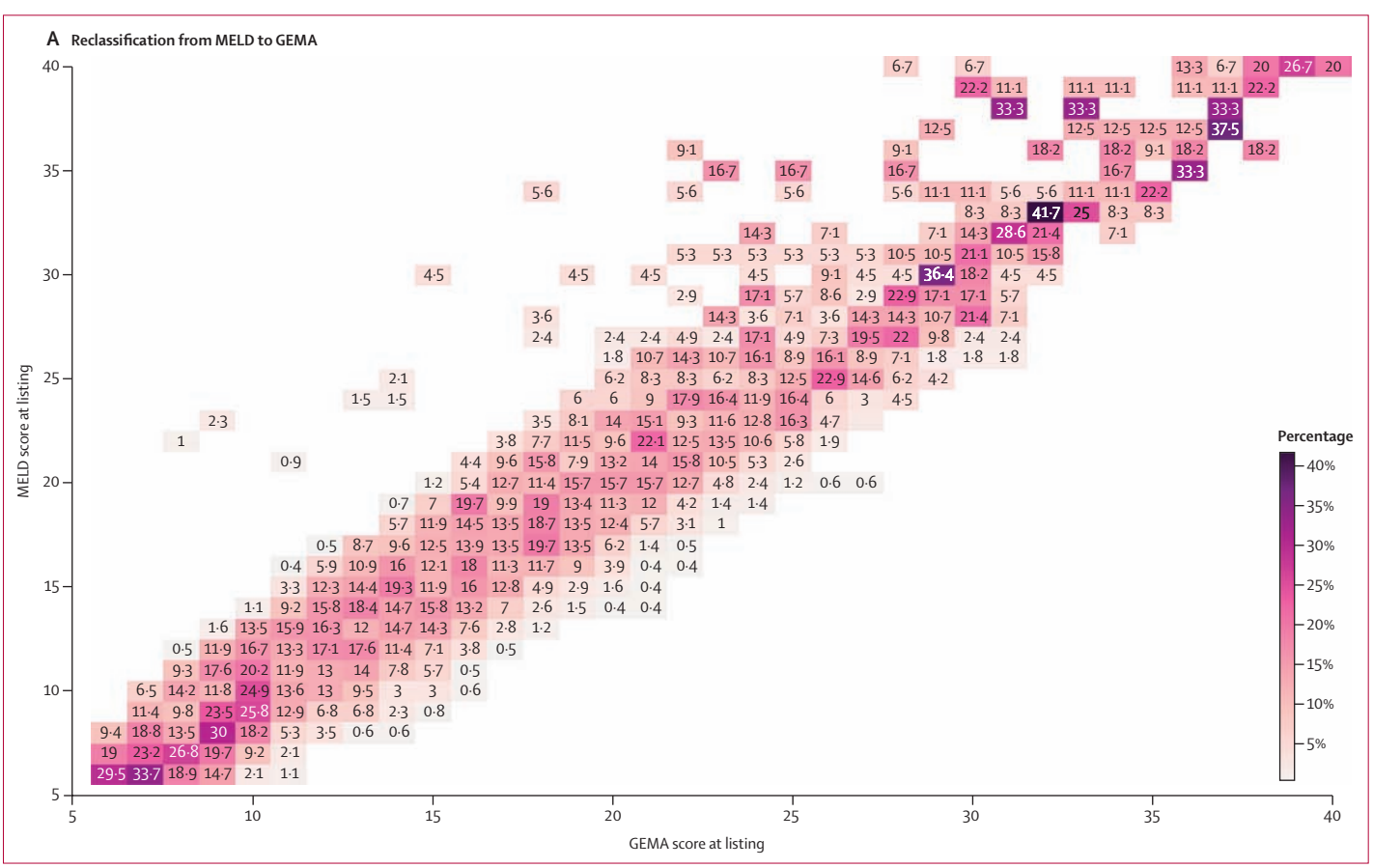


Figure 2: **Bar calibration plots of the GEMA, GEMA-Na, MELD, MELD-Na, and MELD 3.0 in the internal validation cohort**
The predicted and observed probabilities for the primary outcome are presented. For each model, the cohort was stratified into deciles of risk, which were merged to ensure a minimum of two events per group of risk. p values correspond to the Greenwood-Nam-D'Agostino goodness-of-fit test. GEMA=Gender-Equity Model for liver Allocation. GEMA-Na=GEMA corrected by serum sodium. MELD=Model for End-stage Liver Disease. MELD-Na=MELD corrected by serum sodium.

versus MELD-Na (3·26, 1·68–6·34; p=0·0005); and GEMA-Na versus MELD 3.0 (3·08, 1·62–5·85; p=0·001). Within the first 90 days after inclusion, one in 15 deaths could be potentially saved and one life per 259 patients included could be saved by using GEMA instead of MELD. Using GEMA-Na instead of MELD-Na would potentially save one in 21 deaths and could save one life per 373 patients included. Using GEMA-Na instead of MELD 3.0 would potentially save one in 19 deaths and could save one life per 351 patients included. Compared with MELD and MELD-Na (appendix pp 9–11), patients prioritised only by GEMA and GEMA-Na were more often women (293 [50%] of 586 patients *vs* 175 [29·9%] of 586 patients, and 239 [51·1%] of 468 patients *vs* 135 [28·8%] of 468 patients; p<0·0001 for both comparisons) and had increased prevalence of moderate-to-severe ascites (347 [59·2%] of 586 patients *vs* 117 [20%] of 586 patients, and 248 [53%] of 468 patients *vs* 101 [21·6%] of 468 patients; p<0·0001 for both comparisons). There were more women among patients prioritised by MELD 3.0 than among patients prioritised by GEMA-Na (246 [48·3%] of 509 patients *vs* 179 [35·2%] of 509 patients; p<0·0001), but the risk of death or delisting due to clinical deterioration was higher in the GEMA-Na group (38 [7·5%] of 509 patients *vs* 13 [2·6%] of 509 patients; p=0·0003). In the subgroup of women, the effect of using the GEMA models was increased. The prevalence of the primary outcome in women prioritised only by MELD was 1·1% (two of 175 patients) compared with 8·5% (25 of 293 patients) among women prioritised only by GEMA, resulting in one in eight deaths potentially avoided and one life potentially spared per 131 women included. Regarding the models containing serum sodium, the prevalence of the primary outcome in women prioritised only by MELD-Na was 0% compared with 9·6% (23 in 239 patients) among women prioritised only by GEMA-Na, resulting in one in eight deaths potentially avoided and one life potentially spared per 131 women included. Finally, the primary outcome occurred in two (0·8%) out of 246 women prioritised only by MELD 3.0 versus 13 (7·3%) of 179 women prioritised only by GEMA-Na, resulting in one in 14 deaths potentially avoided and one life potentially spared per 253 women included.

Discussion

In this study, we derived and validated the GEMA models to rectify sex-based disparities in liver transplantation candidates. Our models showed improved discrimination and re-classification compared



(Figure 3 continues on next page)

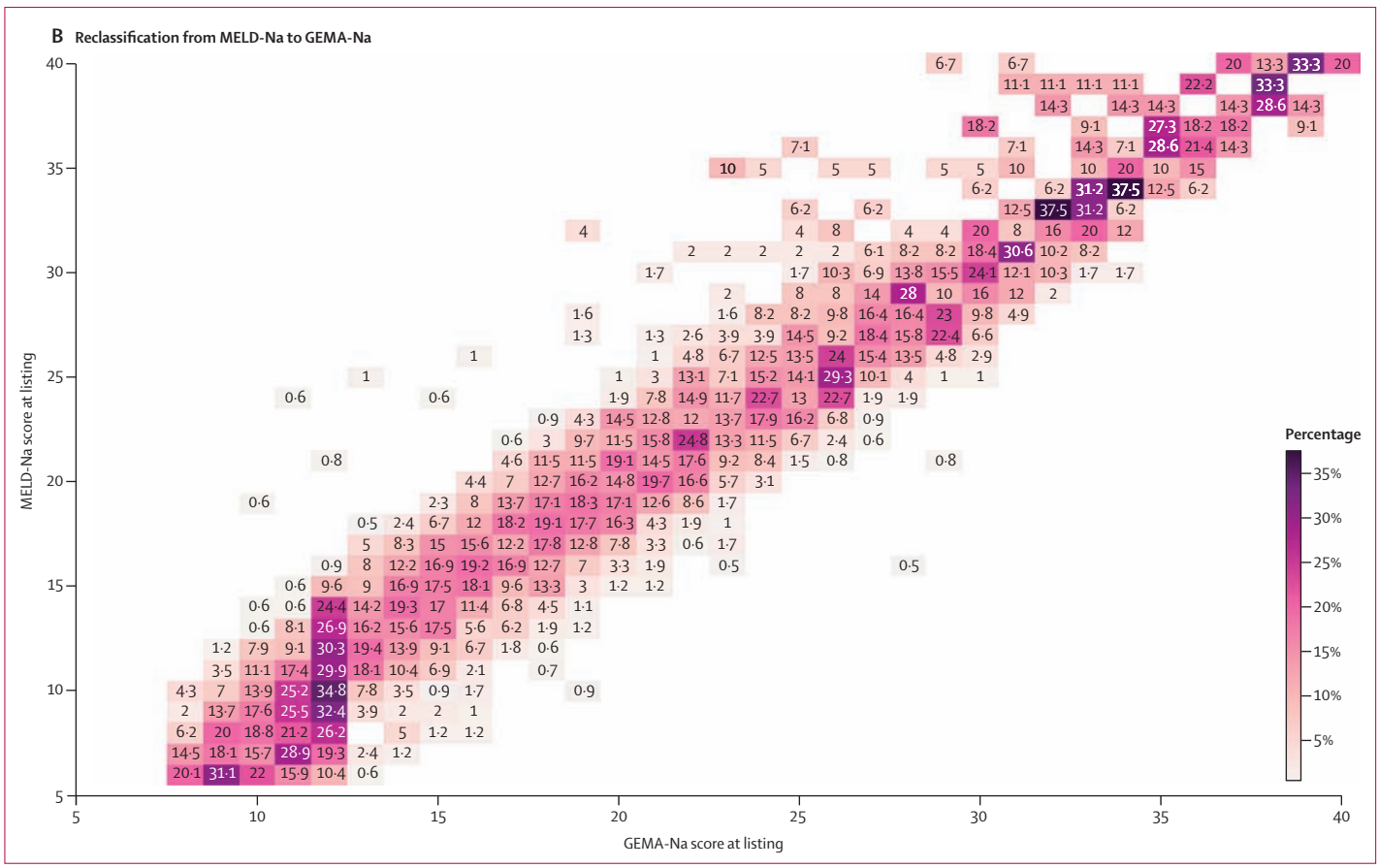


Figure 3: **Re-classification plots showing the concordance of GEMA versus MELD (A) and of GEMA-Na versus MELD-Na (B) in the pooled validation cohort (n=3558)**
The diagonal line shows coinciding values of compared models. Deviations to the left indicate lower GEMA and GEMA-Na scores compared with MELD and MELD-Na, and deviations to the right indicate higher GEMA and GEMA-Na scores compared with MELD and MELD-Na. Numbers in each box are percentages of GEMA or GEMA-Na for a certain (fixed) value of MELD and MELD-Na, respectively. GEMA=Gender-Equity Model for liver Allocation. GEMA-Na=GEMA corrected by serum sodium. MELD=Model for End-stage Liver Disease. MELD-Na=MELD corrected by serum sodium.

with the MELD equations with adequate calibration, and these effects were more pronounced in women. We used two large cohorts of liver transplantation candidates from different countries and organ allocation systems to ensure generalisability of our findings. To the best of our knowledge, GEMA is the first prioritisation score for 3-month survival in the transplant waiting list derived and validated outside the USA and offers the highest improvement in discrimination since the introduction of MELD.

The use of the urgency principle for organ allocation relies on objective and accurate predictors of mortality within the waiting list. The MELD and MELD-Na scores are the standard of care for liver allocation and their implementation has improved waiting list outcomes,[2,18] but has also created sex-based disparities in accessing liver transplantation.[8,19,20] The introduction of MELD has resulted in a 30% increase in the odds of death of women in the transplant waiting list.[6] Assuming that women had equal access to liver transplantation compared with men, the deaths of more than 800 women could have been prevented in the past decade in the USA.[21] Eliminating gender bias in patient selection for liver transplantation is therefore a priority.

Renal impairment is an important prognostic factor among patients with decompensated cirrhosis, justifying its inclusion and weighting within the initial MELD formula,[22] along with its variants MELD-Na[1] and UKELD.[23] However, serum creatinine is also influenced by muscle mass and it is systematically lower in women compared with men with similar renal function.[19,24] It has been estimated that women accrue between one to three fewer MELD points from serum creatinine from the same glomerular filtration rate compared with men.[7] The implementation of estimated glomerular filtration rate did not improve outcome predictions of MELD in the waiting list,[25] but two cirrhosis-specific formulas have been developed and validated in the past 5 years: the glomerular filtration rate assessment in liver disease (GRAIL)[26] and the RFH-GFR.[9] Compared with GRAIL, the RFH-GFR did not incorporate race, but considered INR, sodium, and presence of moderate-to-severe ascites, which are widely recognised markers of more advanced liver disease. The replacement of serum creatinine by

GRAIL within the MELD-Na formula (MELD-GRAIL-Na) was evaluated within the US Scientific Registry of Transplant Recipients.[27] The MELD-GRAIL-Na improved the predictive ability of MELD-Na (Harrell's concordance statistic 0·83 vs 0·82) and re-classified 16·7% of patients, but unfortunately these results were not reproduced outside the USA.[28] Of note, training and validation cohorts were defined according to a different date of inclusion within the waiting list (2014 for the training cohort and 2015 for the validation cohort), in a period strongly influenced by the introduction of direct-acting antivirals against hepatitis C, which might have introduced bias due to improvements in liver function after sustained virological response, potentially resulting in delisting.[29] Although we could not compare head-to-head with MELD-GRAIL, our results suggest that GEMA could make a greater impact on organ allocation and, more importantly, could amend the historical gender disparities for accessing liver transplantation.[8]

Another approach to correct gender disparities for accessing liver transplantation would be including sex as part of the score. In a study from the US Organ Procurement and Transplant Network, female sex and serum albumin were added to MELD-Na, without replacement of serum creatinine, resulting in the MELD 3.0 score.[12] The discriminative improvement obtained with MELD 3.0 was modest (difference in Harrell's concordance statistic 0·007 *vs* MELD-Na) and most patients retained the same score when comparing with MELD-Na. Importantly, the calibration of MELD 3.0 was not reported and external validation is lacking. In this study, the GEMA-Na model had significantly increased discrimination capacity compared with MELD 3.0 in both cohorts. We also tested the addition of albumin within the GEMA models, but it did not result in a significant discrimination benefit.

There were more women among patients prioritised by MELD 3.0 than among those prioritised by GEMA-Na, but the risk of mortality or delisting due to clinical worsening was tripled in the women prioritised by GEMA-Na. This finding highlights a major limitation of MELD 3.0 and other models that add an arbitrary number of points for women without replacing serum creatinine, which could result in overcorrection of sex differences without eliminating creatinine-derived bias.[13] Importantly, MELD 3.0 was not superior to MELD-Na in either the UK or the Australian transplant cohort. These results shed important doubts on the superiority of MELD 3.0 over MELD-Na, particularly as it is under consideration for organ allocation in the USA.

The original MELD formula underestimates the risk of 90-day mortality in some patients with end-stage liver disease.[30] To overcome this limitation, MELD-Na included serum sodium in the equation but its predictive capacity in patients with ascites and other MELD exceptions is still suboptimal. Our study enrolled patients with MELD exceptions, including patients with hepatocellular carcinoma, to mirror the composition of the liver transplantation waiting list as accurately as possible. A sensitivity analysis showed equivalent performance of the models when such patients were excluded (appendix p 9). The presence of ascites adds relevant information to the MELD-Na formula to predict 1-year mortality, particularly in patients with lower MELD-Na scores.[31] The inclusion of ascites as part of RFH-GFR is a potential advantage for GEMA as it could better capture the interplay between ascites, serum sodium, and renal function, which is paramount for the prognosis of patients with end-stage liver disease.

GEMA-Na had the highest discrimination benefit over the current standard. Indeed, when GEMA-Na is compared with MELD-Na, the difference in Harrell's concordance statistics was 0·024 in the UK cohort and 0·029 in the Australian cohort. To contextualise these data, MELD 3.0 had a difference in Harrell's concordance statistic of 0·007 compared with MELD-Na,[12] and the original validation study of MELD-Na showed a difference in Harrell's concordance statistic of 0·018 compared with MELD.[1] The implementation of GEMA in different countries and transplant systems will require local evaluation and validation. It will also require the recording of readily available parameters, which might not be routinely captured to date, such as urea blood levels and the presence of moderate-to-severe ascites. To ease this process, an online calculator of GEMA has been made available. The calculator allows the user to input either urea or blood urea nitrogen as per local practice, making transformations automatically. A potential barrier for the implementation of GEMA could be the need of recording ascites and the perception of subjectivity in the evaluation, which could be addressed by combining physical examination with imaging techniques such as ultrasound or CT, as recommended every 3 months for portal vein thrombosis screening while the patient is on the waiting list.[32]

For the **online GEMA calculator** see http://gema-transplant.com

This study has inherent limitations. The inclusion of patients enlisted for hepatocellular carcinoma could have negatively affected the accuracy of predictions in all models, particularly in the low-score strata of patients. Unlike previous models, the calculation of the GEMA models requires evaluation of the presence of ascites by physical examination, which could be influenced by using diuretics. The combination of physical examination with imaging techniques routinely performed in patients waiting for liver transplantation would minimise subjectivity in evaluating ascites. Height information was not available to determine if it could have added to the GEMA models. The reclassification analysis does not mirror the true complexity of organ allocation and potential lives saved are only indicative. The external validation cohort had a limited sample size and a long recruitment period, leaving the cohort open to the effects of changes in management. Finally, it is unclear how variations of the GEMA scores over time would affect outcomes.

In conclusion, the replacement of serum creatinine by RFH-GFR in the original MELD-Na formula with a subsequent re-fitting and re-weighting of the model components, results in more accurate predictions of mortality or delisting due to clinical deterioration within the first 90 days in patients awaiting liver transplantation. The accuracy of the GEMA models was highest among women, suggesting that their implementation in clinical practice could obviate current gender inequities for accessing liver transplantation.

Contributors

MLR-P designed the study and participated in data analysis and drafting of the article. AMG-O did the mathematical modelling and participated in drafting the article. AM participated in data collection and critical revision of the article. MB critically revised the article. GWM, PG, and RT participated in data collection and critical revision of the article. MG and DG-R participated in data curation and analysis. CH-M participated in data curation, analysis, and critical revision of the article. EAT conceived they study and participated in its design, critically revised the article and is the study guarantor. MLR-P and AMG-O had full access to the study data and verified the results. All authors revised and approved the final version of the manuscript before submission. MLR-P and EAT were responsible for the decision to submit the manuscript.

Declaration of interests

MLR-P has received lecture fees from Astellas Pharma, Chiesi, and Eisai, outside the present work. EAT has attended advisory boards for Boehringer, Intercept, Pfizer, NovoNordisk, Alexion, Orphalan, outside the present work. All other authors declare no competing interests.

Data sharing

The data used for model training and internal validation was extracted from the UK Transplant Registry, held by NHS Blood and Transplant. Deidentified participant data could be shared with an external investigator only after approval by NHS Blood and Transplant. For this purpose, proposals must be referred to the representative of NHS Blood and Transplant in this study, Rhiannon Taylor, by e-mail at rhiannon.taylor@nhsbt.nhs.uk. A signed confidentiality agreement would be required.

Acknowledgments

This study was funded by Junta de Andalucía (grants PI-0164–2019, PS-2020–780, UCO-1261651, and PY20_00074), Agencia Española de Investigación (grant PID2020–115454GB-C22/AEI/10.13039/501100011033), and co-funded by the European Regional Development Fund. The funding agents covered the contract of an expert in mathematical modelling, the creation of the GEMA online calculator, and publication expenses, but they had no role in the data analysis, manuscript preparation, or in the decision to publish the study. AMG-O was supported by Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía (grant ref: PREDOC-00489). DG-R was supported by the requalification of the Spanish university programme, financed by the European Union -NextGenerationEU (grant ref: UCOR01MS). This study is dedicated to the memory of the late Andrew K Burroughs (who died in March, 2014), whose mentorship and early observations of the overestimation of kidney function in cirrhosis by creatinine and GFR provided the basis for this work.

References

1 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:** 1018–26.

2 Tschuor C, Ferrarese A, Kuemmerli C, et al. Allocation of liver grafts worldwide—is there a best system? *J Hepatol* 2019; **71:** 707–18.

3 Yoo JJ, Kim SG, Kim YS, et al. Estimation of renal function in patients with liver cirrhosis: impact of muscle mass and sex. *J Hepatol* 2019; **70:** 847–54.

4 Bhanji RA, Montano-Loza AJ, Watt KD. Sarcopenia in cirrhosis: looking beyond the skeletal muscle loss to see the systemic disease. *Hepatology* 2019; **70:** 2193–203.

5 Cullaro G, Sarkar M, Lai JC. Sex-based disparities in delisting for being "too sick" for liver transplantation. *Am J Transplant* 2018; **18:** 1214–19.

6 Moylan CA, Brady CW, Johnson JL, Smith AD, Tuttle-Newhall JE, Muir AJ. Disparities in liver transplantation before and after introduction of the MELD score. *JAMA* 2008; **300:** 2371–78.

7 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:** 1710–16.

8 Melk A, Babitsch B, Borchert-Mörlins B, et al. Equally interchangeable? How sex and gender affect transplantation. *Transplantation* 2019; **103:** 1094–110.

9 Kalafateli M, Wickham F, Burniston M, et al. Development and validation of a mathematical equation to estimate glomerular filtration rate in cirrhosis: the royal free hospital cirrhosis glomerular filtration rate. *Hepatology* 2017; **65:** 582–91.

10 Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162:** W1-73.

11 Angeli P, Bernardi M, Villanueva C, et al. EASL clinical practice guidelines for the management of patients with decompensated cirrhosis. *J Hepatol* 2018; **69:** 406–60.

12 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:** 1887–95.

13 Wood NL, VanDerwerken D, Segev DL, Gentry SE. Correcting the sex disparity in MELD-Na. *Am J Transplant* 2021; **21:** 3296–304.

14 Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 2017; **318:** 1377–84.

15 Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med* 2015; **34:** 685–703.

16 D'Agostino R, Nam B. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In: Handbook of Statistics. Amsterdam: Elsevier, 2003; **23:** 1–25.

17 Goudsmit BFJ, Putter H, Tushuizen ME, et al. Refitting the model for end-stage liver disease for the eurotransplant region. *Hepatology* 2021; **74:** 351–63.

18 Sacleux SC, Samuel D. A critical review of MELD as a reliable tool for transplant prioritization. *Semin Liver Dis* 2019; **39:** 403–13.

19 Cholongitas E, Marelli L, Kerry A, et al. Female liver transplant recipients with the same GFR as male recipients have lower MELD scores—a systematic bias. *Am J Transplant* 2007; **7:** 685–92.

20 Sarkar M, Watt KD, Terrault N, Berenguer M. Outcomes in liver transplantation: does sex matter? *J Hepatol* 2015; **62:** 946–55.

21 Verna EC, Lai JC. Time for action to address the persistent sex-based disparity in liver transplant access. *JAMA Surg* 2020; **155:** 545–47.

22 Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001; **33:** 464–70.

23 Barber K, Madden S, Allen J, Collett D, Neuberger J, Gimson A. Elective liver transplant list mortality: development of a United Kingdom end-stage liver disease score. *Transplantation* 2011; **92:** 469–76.

24 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:** 1169–77.

25 Myers RP, Shaheen AA, Aspinall AI, Quinn RR, Burak KW. Gender, renal function, and outcomes on the liver transplant waiting list: assessment of revised MELD including estimated glomerular filtration rate. *J Hepatol* 2011; **54:** 462–70.

26 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:** 1219–30.

27 Asrani SK, Jennings LW, Kim WR, et al. MELD-GRAIL-Na: glomerular filtration rate and mortality on liver-transplant waiting list. *Hepatology* 2020; **71:** 1766–74.

28 Chetwood JD, Wells MG, Tsoutsman T, et al. 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. *Transplant Direct* 2022; **8:** e1346.

29 Pascasio JM, Vinaixa C, Ferrer MT, et al. Clinical outcomes of patients undergoing antiviral therapy while awaiting liver transplantation. *J Hepatol* 2017; **67:** 1168–76.

30 Heuman DM, Abou-Assi SG, Habib A, et al. Persistent ascites and low serum sodium identify patients with cirrhosis and low MELD scores who are at high risk for early death. *Hepatology* 2004; **40:** 802–10.

31 Somsouk M, Guy J, Biggins SW, Vittinghoff E, Kohn MA, Inadomi JM. Ascites improves upon [corrected] serum sodium plus [corrected] model for end-stage liver disease (MELD) for predicting mortality in patients with advanced liver disease. *Aliment Pharmacol Ther* 2009; **30:** 741–48.

32 Montalvá E, Rodríguez-Perálvarez M, Blasi A, et al. Consensus statement on hemostatic management, anticoagulation, and antiplatelet therapy in liver transplantation. *Transplantation* 2022; **106:** 1123–31.