

ORIGINAL ARTICLE

Superiority of the new sex-adjusted models to remove the female disadvantage restoring equity in liver transplant allocation

Giuseppe Marrone^{1,2}  | Valerio Giannelli³ | Salvatore Agnes^{1,2} |
 Alfonso Wolfango Avolio^{1,2}  | Leonardo Baiocchi⁴ | Giammauro Berardi³ |
 Giuseppe Maria Ettorre³ | Flaminia Ferri⁵  | Stefano Ginanni Corradini⁵ |
 Antonio Grieco^{1,2}  | Nicola Guglielmo³ | Ilaria Lenci⁴ | Raffaella Lionetti⁶  |
 Gianluca Mennini⁵ | Martina Milana⁴ | Massimo Rossi⁵ | Gabriele Spoletini^{1,2} |
 Giuseppe Tisone⁴ | Tommaso Maria Manzia⁴ | Quirino Lai⁵

¹Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy

²Università Cattolica del Sacro Cuore, Rome, Italy

³San Camillo Hospital, Rome, Italy

⁴Tor Vergata University, Rome, Italy

⁵Sapienza University, Rome, Italy

⁶Istituto Nazionale Malattie Infettive Lazzaro Spallanzani-IRCCS, Rome, Italy

Correspondence

Gabriele Spoletini, Fondazione Policlinico Universitario Agostino Gemelli IRCCS—Università Cattolica del Sacro Cuore, Largo Agostino Gemelli 8, Rome, 00168, Italy
 Email: gabriele.spoletini@policlinicogemelli.it

Handling Editor: Alejandro Forner

Abstract

Background and Aims: Model for End-stage Liver Disease (MELD) and MELDNa are used worldwide to guide graft allocation in liver transplantation (LT). Evidence exists that females are penalized in the present allocation systems. Recently, new sex-adjusted scores have been proposed with improved performance respect to MELD and MELDNa. GEMA-Na, MELD 3.0, and sex-adjusted MELDNa were developed to improve the 90-day dropout prediction from the list. The present study aimed at evaluating the accuracy and calibration of these scores in an Italian setting.

Methods: The primary outcome of the present study was the dropout from the list up to 90 days because of death or clinical deterioration. We retrospectively analysed data from 855 adults enlisted for liver transplantation in the Lazio region (Italy) (2012–2018). Ninety-day prediction of GEMA-Na, MELD 3.0 and sex-adjusted MELDNa with respect to MELD and MELDNa was analysed. Brier score and Brier Skill score were used for accuracy, and the Greenwood-Nam-D'Agostino test was used to evaluate the calibration of the models.

Results: GEMA-Na (concordance = .82, 95% CI = .75–.89), MELD 3.0 (concordance = .81, 95% CI = .74–.87) and sex-adjusted MELDNa (concordance = .81, 95% CI = .74–.88) showed the best 90-day dropout prediction. GEMA-Na showed a higher increase in accuracy with respect to MELD ($p = .03$). No superiority was shown with respect to MELDNa. All the tested scores showed a good calibration of the models. Using GEMA-Na instead of MELD would potentially save one in nine dropouts and could save one dropout per 285 patients listed.

Abbreviations: 95% CI, 95% confidence intervals; CLD, chronic liver disease; DAAs, direct-acting antivirals; GEMA-Na, Gender-Equity Model for liver Allocation sodium; HCC, hepatocellular carcinoma; INR, international normalized ratio; IQRs, interquartile ranges; LT, liver transplantation; MELD, Model for End-stage Liver Disease; MELDNa, Model for End-stage Liver Disease Sodium; NNT, number needed to treat; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis; SRTR, Scientific Registry of Transplant Recipients; STROBE, The Strengthening the Reporting of Observational Studies in Epidemiology.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Liver International* published by John Wiley & Sons Ltd.

Conclusions: Validation and reclassification of the sex-adjusted score GEMA-Na confirm its superiority in predicting short-term dropout also in an Italian setting when compared with MELD.

KEYWORDS

allocation, cirrhosis, equity, GEMA, liver transplantation, MELD Na, MELD, MELD 3.0, Na, sex

1 | INTRODUCTION

Since 2002, the Model for End-stage Liver Disease (MELD) has been adopted worldwide to stratify the prognosis of patients with chronic liver disease (CLD) and to rank patients on the liver transplantation (LT) waiting list.¹⁻⁴ Starting in 2016, MELD Sodium (MELDNa) has substituted MELD score, thanks to its higher ability to predict survival and stratify patients on the waiting list in risk classes.^{5,6}

Over time, a sex-based disparity effect has become manifest, with the female sex being penalized by the current allocation systems.⁷⁻¹² This phenomenon led to various efforts to conceive a more equitable system. Attempts have been made at sex-balancing creatinine among the MELDNa variables, although this proved ineffective due to inherent limitations related to muscle loss and sarcopenia in most patients with CLD.¹³

Recently, MELD 3.0 was developed using a simulated allocation system based on data from all the candidates registered on the liver transplant wait list in the 2016-2018 US national registry. Female sex and serum albumin were added to the MELDNa variables, allowing the reclassification of 8.8% of decedents to higher scores and, therefore, better chances of receiving LT, particularly in female patients with lower creatinine values due to smaller muscle mass.¹⁴ Furthermore, a single-centre experience from the United States also developed a sex-adjusted MELDNa score, where female LT recipients had more decompensated traits despite having lower median MELDNa scores.¹⁵ Finally, a large population ($n = 7682$) from the UK Transplant Registry consented to create the Gender-Equity Model for liver Allocation sodium (GEMA-Na): this score was externally validated in an Australian cohort ($n = 1638$), showing an improvement in the discrimination and a significant reclassification benefit compared with existing scores.¹⁶

Italian LT patients compose a peculiar population due to a specific donor- (i.e. up to 50% prevalence of extended criteria donors) and recipient-related characteristics (i.e. elderly age).¹⁷ The Italian allocation system is based on MELDNa, offering national priority for patients with MELDNa ≥ 30 .^{18,19}

How the novel sex-adjusted scores¹⁴⁻¹⁶ fit into such a particular setting has not been investigated yet. Moreover, discrimination ability and calibration of these scores have not been validated externally on a multicentre population.

We hypothesized that the recently proposed scores have a more accurate dropout prediction than the unadjusted versions in CLD patients awaiting LT.

Key points

Waitlist priority for liver transplantation is defined through prognostic scores (namely, MELD and MELDNa). In our population, new scores considering sex (GEMA-Na, MELD 3.0 and sex-adjusted MELDNa) better predict the risk of death and delisting, thus improving equity in organ allocation.

Using the data of four LT centres in Rome, Italy,²⁰ we aimed to evaluate the predictive capacity of MELD 3.0, sex-adjusted MELDNa and GEMA-Na scores in an independent Italian cohort.

2 | MATERIALS AND METHODS

2.1 | Study design

This is a retrospective multicentre observational study investigating the data of cirrhotic patients enlisted for receiving a first LT. Data were prospectively recorded for allocation needs.

The Institutional Review Board approved the present study of the coordinator centre. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines were followed to create the study.

2.2 | Setting

The involved centres were four LT centres in Rome, namely Sapienza University, Tor Vergata University, Catholic University and San Camillo Hospital. The four collaborative centres shared the listing criteria, the donor pool and the allocation system.

2.3 | Population

We performed a retrospective analysis of the data of 855 adult (≥ 18 years) cirrhotic patients listed for LT from 1 January 2012 to 31 December 2018. Non-inclusion criteria were (a) age < 18 years, (b) non-cirrhotic disease or acute liver failure as the main indication for waiting list inscription, (c) combined transplantations

and (d) re-transplantation. Patients with hepatocellular carcinoma (HCC) on cirrhosis were included in the study. Although the MELD value was corrected in the HCC patients for the allocation process by adding exception points,¹⁹ this study considered HCC patients with their 'biochemical' MELD and not with HCC exception points.

The study was performed following the ethical standards of the Declaration of Helsinki.

2.4 | Outcomes

The primary outcome of the study was dropout up to 90 days from the time of listing. Dropout was defined as mortality within the waiting list or delisting due to clinical deterioration. 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. The last follow-up date was 31 March 2019.

2.5 | Definitions

For the calculation of MELD, MELDNa, MELD 3.0 with or without albumin, sex-adjusted MELD and GEMA-Na, the following standard formulas were used, as previously described:

$MELD = 9.57 \times \log e(\text{creatinine}) + 3.78 \times \log e(\text{bilirubin}) + 11.20 \times \log e(\text{INR}) + 6.43$, where creatinine (mg/dL), bilirubin (mg/dL) and INR values below 1.0 were set to 1.0, and creatinine values were set to 4.0 mg/dL if serum creatinine was >4 mg/dL or the patient received two or more dialysis treatments within the prior week.¹

$MELDNa = MELD + [1.32 \times (137 - Na)] - [.033 \times MELD \times (137 - Na)]$, where the serum sodium concentration (Na) is between 125 and 137 mmol/L.⁶

$$MELD\ 3.0\ with\ albumin = 1.33\ (if\ female) + [4.56 \times \log e(\text{bilirubin})] + [.82 \times (137 - Na)] - [.24 \times (137 - Na) \times \log e(\text{bilirubin})] + [9.09 \times \log e(\text{INR})] + [11.14 \times \log e(\text{creatinine})] + [1.85 \times (3.5 - \text{albumin})] - [1.83 \times (3.5 - \text{albumin}) \times \log e(\text{creatinine})] + 6$$

$$MELD\ 3.0\ no\ albumin = 1.40\ (if\ female) + [4.85 \times \log e(\text{bilirubin})] + [.88 \times (137 - Na)] - [.25 \times (137 - Na) \times \log e(\text{bilirubin})] + [9.66 \times \log e(\text{INR})] + [10.47 \times \log e(\text{creatinine})] + 6$$

Sex-adjusted MELDNa was calculated according to the score map proposed by Sealock et al.¹⁵

$$GEMA - Na = GEMA - Na - [.025 \times GEMA \times (140 - Na)] + 140$$

GEMA was calculated according to the formula:

$$3.777 \times \ln(\text{Bilirubin}) + 7.883 \times \ln(\text{INR}) - 8.306 \times \ln(\text{RFH} - \text{GFR}) + 31.932$$

The Royal Free Hospital Glomerular Filtration Rate (RFH-GFR) was calculated at inclusion in the waiting list according to the original formula: $45.9 \times (\text{creatinine}^{-.836}) \times (\text{urea}^{-.229}) \times (\text{INR}^{-.113}) \times (\text{age}^{-.129}) \times (\text{sodium}^{.972}) \times .809$ (if female) $\times .92$ (if moderate or severe ascites).

All the models were calculated at waiting list inclusion. The resulting scores were rounded to the nearest whole number to yield the corresponding scores. In the present analysis, in the case of patients receiving exception points (i.e. HCC), we calculated the laboratory scores without adding extra points.

2.6 | Statistical analysis

Continuous variables were reported as medians and interquartile ranges (IQRs). Categorical variables were reported as numbers and percentages. Missing data relative to study covariates always involved less than 10% of patients. In all the cases, missing data were handled with a single imputation method. In detail, a median of imputation in nearby points was adopted. The median instead of the mean was adopted due to the potentially skewed distribution of the managed variables.

The discriminatory ability of the GEMA-Na, sex-adjusted MELDNa and MELD 3.0 with and without albumin was tested against MELD and MELDNa. Model ability to rank enlisted patients for LT according to their risk of dropout within 90 days was evaluated with the c-statistic using the method by Harrell et al. The c-statistics and 95% CIs were reported.²¹ The method described by Kang et al. was adopted for comparing the c-statistics results.²²

The Brier score and the Brier Skill score were reported with the intent to explore the additive accuracy in the prediction of the risk of 90-day dropout provided by the new proposed scores.²³ A positive value of the Brier Skill score corresponds to a superior performance of the tested model with respect to the MELD considered as the reference score.

The models' reliability was tested by comparing the expected versus observed survival rates. The Greenwood-Nam-D'Agostino test was used for assessing the calibration between observed and expected event rates.²⁴ Bar calibration plots were constructed to

visually represent the agreement between predictions and observations in each group of risk.

The reclassification rate was defined as the proportion of patients with a score change of at least 2 when comparing MELD versus the other scores. Patients granted at least 2 extra points with the new 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 prioritization score evaluated and the cohort was stratified in patients who would have been prioritized by both models, and patients who would have been differently prioritized by either of them. The number of potential lives saved resulted from subtracting the number of patients reaching the primary outcome in the new score-prioritized group from the number of patients reaching the primary outcome in the MELD-prioritized group, divided by the total number of patients included.¹⁶

A *p*-value of $<.05$ was considered significant for all the analyses, and all tests were two-tailed. We used the SPSS statistical package version 27.0 (SPSS Inc), R (version 4.1.2) and the STATA package version 16.0 (StataCorp LLC) for the statistical analyses.

3 | RESULTS

3.1 | General patient characteristics and stratification according to sex

Patient characteristics are reported in [Table 1](#). The median follow-up period from listing was 24 (IQR = 8–46) months. The median age of the investigated population was 57 years (IQR = 50–62), with 19.1% of women. The median MELD and MELDNa were 15 (IQR = 11–19) and 17 (IQR = 13–22), respectively.

The median sex-adjusted MELDNa was 17 (IQR = 13–23). The median MELD 3.0 with and without albumin included in the formula were 15 (IQR = 10–20) and 15 (IQR = 10–21), respectively. The median GEMA-Na was 14 (IQR = 9–19).

At 90 days from listing, 26 (3.0%) cases dropped out due to death ($n = 22$) or severe clinical worsening ($n = 4$).

Stratifying the population according to sex, several relevant differences were observed between the male and female groups regarding the underlying liver disease, HCC, anthropomorphic aspects and creatinine ([Table 1](#)). In detail, HCV and alcohol were more commonly observed in males as the cause of cirrhosis, and biliary cirrhosis was more commonly reported in females. Notably, no differences were reported in terms of median MELD ($p = .22$) and MELDNa ($p = .08$) between the two groups. On the contrary, when the sex-adjusted scores were calculated, the median value of the different scores was significantly higher in females compared to males: in detail, 20 versus 17 for sex-adjusted MELDNa ($p < .0001$), 17 versus 15 for MELD 3.0 with albumin ($p < .0001$), 18 versus 15 for MELD 3.0 without albumin ($p < .0001$) and 19 versus 13 for GEMA-Na ($p < .0001$).

No differences were observed in dropout when males and females were compared.

3.2 | Concordance results

Investigating the entire population, GEMA-Na (.82, 95% CI = .75–.89), MELD 3.0 with albumin (.81, 95% CI = .74–.87), sex-adjusted

MELDNa (.81, 95% CI = .74–.88) and MELD 3.0 without albumin (.80, 95% CI = .74–.87) showed the highest *c*-statistics for the 90-day dropout.

MELD score had an inferior prognostic ability (*c*-statistics = .76, 95% CI = .68–.85). Difference was statistically significant when MELD was compared with GEMA-Na ($p = .03$). Although the numerical difference was modest, also MELD 3.0 with albumin showed a statistically relevant difference ($p = .049$). Sex-adjusted MELDNa very closely neared statistical relevance compared to MELD ($p = .06$). No relevant statistical differences were reported when MELDNa was compared with GEMA-Na ($p = .17$), MELD 3.0 with albumin ($p = .17$) or sex-adjusted MELDNa ($p = .15$) ([Table 2](#)).

When a sub-analysis focussed on females only was performed, the *c*-statistics of all the explored scores declined, with the GEMA-Na showing the best concordance (.79; 95% CI = .70–.88; $p = .002$). No statistical differences were observed comparing the different scores with the MELD used as the reference.

3.3 | Dropout prediction improvement with the proposed models

With the intent to clarify the magnitude of prediction improvement with the sex-adjusted scores, the Brier score and the Brier Skill score were calculated. Considering the best score as a Brier score = 0, the GEMA-Na reported the lowest value (.188) among the different scores, followed by the MELD 3.0 with albumin (.233) and the sex-adjusted MELDNa (.254).

As for the Brier Skill score results, MELD always presented a reduction in the prediction ability. In detail, the GEMA-Na had the best prediction improvement (4.4%), followed by MELD 3.0 with albumin (3.0%) and sex-adjusted MELDNa (2.4%) ([Table 3](#)).

As reported in [Figure 1](#), all the tested scores presented a good calibration when the predicted dropout events were compared with the observed ones.

In the whole cohort, 296/855 (34.6%) patients received a liver graft within 90 days of listing. When MELD and GEMA-Na were compared, 36 (12.2%) patients would have been differently prioritized. In detail, 18 (6.1%) patients would have been downgraded, while 18 (6.1%) would have been upgraded ([Table 3](#) and [Figure 2](#)). When comparing MELD and MELD 3.0 with albumin, 13 (4.4%) patients would have been differently prioritized. Comparing MELD and sex-adjusted MELDNa, 19 (6.4%) patients would have been differently prioritized.

Within the first 90 days after inclusion, one in nine dropouts could be potentially saved and one dropout per 285 patients included could be saved by using GEMA-Na instead of MELD. Using MELD 3.0 with albumin instead of MELD would potentially save one in 13 dropouts and could save one dropout per 428 patients included. Similarly, using sex-adjusted MELDNa instead of MELD would potentially save one in 13 dropouts and could save one dropout per 428 patients included ([Table 3](#)).

TABLE 1 General patient characteristics and stratification according to sex.

Variables	N = 855 Median (IQR) or n (%)	Males 692 (80.9%)	Females 163 (19.1%)	p-value
Age, years	57 (50–62)	57 (51–61)	58 (48–63)	.65
Male/female sex	692/163 (80.9/19.1)	–	–	–
Waiting time, days	133 (41–293)	131 (43–281)	144 (31–323)	.91
Liver disease ^a				
HCV-related cirrhosis	329 (38.5)	295 (42.6)	34 (20.9)	<.0001
HBV-related cirrhosis	142 (16.6)	114 (16.5)	28 (17.2)	.82
Alcohol-related cirrhosis	293 (34.3)	262 (37.9)	31 (19.0)	<.0001
NASH-related cirrhosis	81 (9.5)	70 (10.1)	11 (6.7)	.23
Biliary cirrhosis	53 (6.2)	27 (3.9)	26 (16.0)	<.0001
Other	114 (13.4)	67 (9.7)	47 (28.8)	<.0001
HCC as co-indication	424 (49.6)	368 (53.2)	56 (34.4)	<.0001
Height, cm	171 (165–177)	173 (169–178)	160 (158–166)	<.0001
Weight, kg	75 (66–85)	77 (70–87)	63 (56–72)	<.0001
BMI	25 (23–28)	26 (24–28)	24 (22–27)	<.0001
BSA	1.9 (1.8–2.0)	1.9 (1.8–2.1)	1.7 (1.6–1.8)	<.0001
Albumin, mg/dL	3.2 (2.8–3.7)	3.2 (2.8–3.7)	3.2 (2.8–3.7)	.54
Creatinine, mg/dL	.9 (.7–1.1)	1.0 (.8–1.1)	.8 (.7–1.0)	<.0001
RFH-GFR, mL/min	93 (70–121)	96 (76–126)	76 (59–102)	<.0001
INR	1.46 (1.25–1.70)	1.46 (1.25–1.70)	1.46 (1.22–1.71)	.75
Total bilirubin, mg/dL	2.5 (1.3–4.7)	2.4 (1.3–4.4)	2.9 (1.4–5.8)	.14
Sodium, mEq/L	138 (135–140)	138 (135–140)	137 (134–140)	.045
MELD	15 (11–19)	15 (11–19)	16 (11–20)	.22
MELDNa	17 (13–22)	17 (13–21)	18 (13–23)	.08
Sex-adjusted MELDNa	17 (13–23)	17 (13–21)	20 (15–25)	<.0001
MELD 3.0				
With albumin	15 (10–20)	15 (9–20)	17 (12–23)	<.0001
Without albumin	15 (10–21)	15 (10–20)	18 (12–23)	<.0001
GEMA-Na	14 (9–19)	13 (8–18)	19 (12–22)	<.0001
Dropout caused by				
Death	62 (7.3)	49 (7.1)	13 (8.0)	.74
Clinical worsening	14 (1.6)	10 (1.4)	4 (2.5)	.36
Delisting caused by				
HCC progression	30 (3.5)	27 (3.9)	3 (1.8)	.24
Clinical improvement	19 (2.2)	14 (2.0)	5 (3.1)	.50
Other causes	19 (2.2)	16 (2.3)	3 (1.8)	1.00
Liver transplantation	646 (75.6)	528 (76.3)	118 (72.4)	.31

Abbreviations: BMI, body mass index; BSA, body surface area; GEMA-Na, Gender-Equity Model for liver Allocation Sodium; 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; MELDNa, Model for End-stage Liver Disease Sodium; NASH, non-alcoholic steatohepatitis; RFH-GFR, Royal Free Hospital Glomerular Filtration Rate.

^aIn 149 cases, more diseases contemporaneously present. Three diseases = 8 cases (HCV + HBV + alcohol = 4; HCV + HBV + NASH = 1; HCV + HBV + biliary = 1; HCV + HBV + other = 1; HCV + NASH + other = 1). Two diseases = 141 cases (HCV + HBV = 13; HCV + alcohol = 7; HCV + NASH = 5; HCV + other = 4; HBV + alcohol = 16; HBV + NASH = 6; HBV + biliary = 1; HBV + other = 7; alcohol + NASH = 8; alcohol + other = 5; NASH + other = 1; biliary + other = 4).

In a sub-analysis performed only including women, GEMA-Na showed the best results. Using GEMA-Na instead of MELD would potentially save one in three dropouts and could save one dropout per 54 women included. All the other scores showed smaller number of patients saved.

4 | DISCUSSION

This study evaluated the reliability and accuracy of the recently developed sex-adjusted models in an Italian multicentric homogeneous cohort for reclassifying patients with cirrhosis awaiting LT. Our

Scores	SE	C-statistics	95.0% CI		p ^a
			Lower	Upper	
Entire population (N = 855)					
MELD	.04	.76	.68	.85	Ref.
MELDNa	.04	.80	.73	.87	.17
MELD 3.0 no albumin	.03	.80	.74	.87	.09
Sex-adjusted MELDNa	.04	.81	.74	.88	.06
MELD 3.0 albumin	.03	.81	.74	.87	.049
GEMA-Na	.03	.82	.75	.89	.03

TABLE 2 Concordance for MELD, MELDNa, sex-adjusted MELDNa, MELD 3.0 with and without albumin, and GEMA-Na. MELD score used as reference.

Abbreviations: GEMA-Na, Gender-Equity Model for liver Allocation Sodium; MELD, Model for End-stage Liver Disease; MELDNa, Model for End-stage Liver Disease Sodium.

^aEvaluated with the method described by Kang et al.²²

TABLE 3 Accuracy of sex-adjusted scores compared to currently widely used scores for the prediction of 90-day dropout. MELD is the reference.

Scores	Brier skill score	Brier skill score (%)	Reclassification of transplanted cases, n (%)	Cases saved on the total number of dropouts	Dropout saved per patient listed
MELD	.333	Ref.	Ref.	Ref.	Ref.
MELDNa	.280	+1.6	17/296 (5.7)	1 in 26	1 in 855
MELD 3.0 no albumin	.276	+1.7	17/296 (5.7)	1 in 26	1 in 855
Sex-adjusted MELDNa	.254	+2.4	19/296 (6.4)	1 in 13	1 in 428
MELD 3.0 albumin	.233	+3.0	13/296 (4.4)	1 in 13	1 in 428
GEMA-Na	.188	+4.4	36/296 (12.2)	1 in 9	1 in 285

Abbreviations: GEMA-Na, Gender-Equity Model for liver Allocation Sodium; MELD, Model for End-stage Liver Disease; MELDNa, Model for End-stage Liver Disease Sodium.

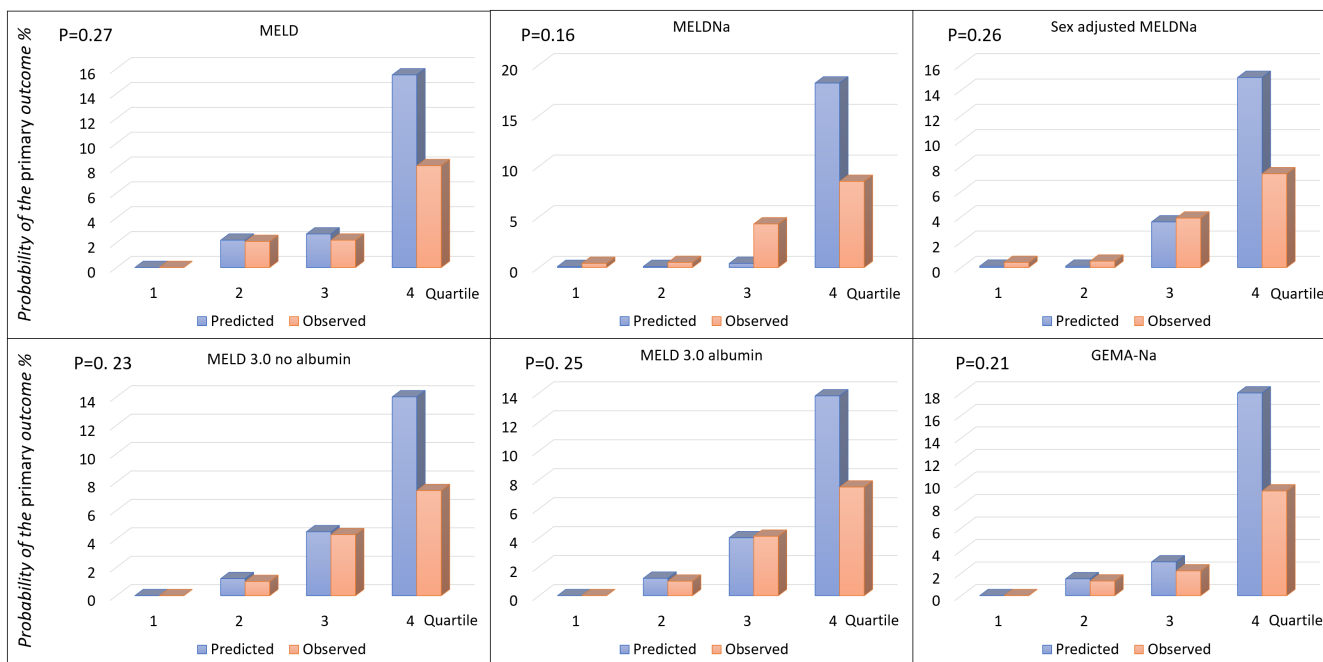


FIGURE 1 Concordance between observed and predicted curves in the investigated scores.

results showed a higher discriminatory ability of the sex-adjusted scores compared to MELD, also in the peculiar Italian setting, characterized by elderly donors and recipients, confirming the higher discriminatory ability of the new proposed scores.

In addition, we provided a detailed dropout prediction improvement of the sex-adjusted models using the Brier score values and the calibration of their observed/predicted cases using the Greenwood-Nam-D'Agostino test. The evaluation of these aspects in a new

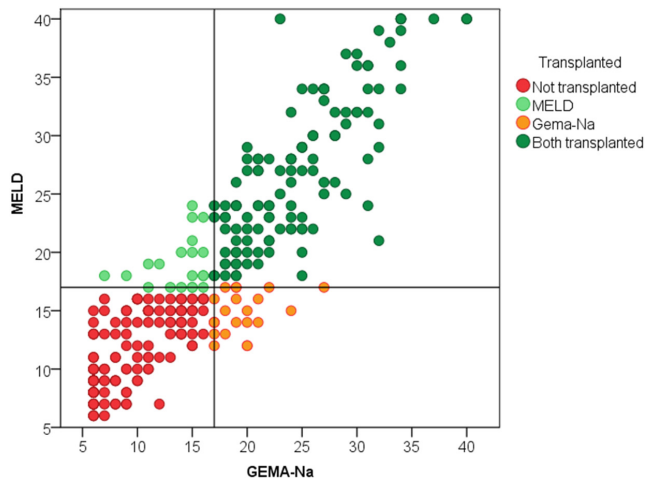


FIGURE 2 Correlation plot of MELD and GEMA-Na. Based on the number of transplanted patients after the first 90 days ($n = 296$), the highest-ranked patients according to both scores separately were assigned a liver graft, as represented by the horizontal (graft granted by MELD) and vertical (by GEMA-Na) lines.

proposed model represents the critical point to forward its large-scale application, as suggested in the MELD 3.0 criticism following the study of Kim et al.^{25,26} This is the first study investigating the calibration of the MELD 3.0 and sex-adjusted MELDNa. On the contrary, the GEMA-Na score calibration and reclassification abilities have been already investigated in the original study proposing the new score.¹⁶

Liver disease epidemiology is changing with the ongoing reduction of HCV-related cirrhosis, the increase in alcoholic liver disease and the predicted outbreak of non-alcoholic steatohepatitis (NASH)-related cirrhosis.²⁷⁻³¹ In this perspective, an evolving transplant candidate population needs a continuous refinement of organ allocation tools to ensure and maintain fairness in LT.

Furthermore, the enlargement of the donor pool and the availability of novel algorithms capable of stratifying the early allograft failure risk pushed us to ameliorate the scores to predict recipient outcomes.^{17,32}

There has been growing awareness of sex-related inequities in recent years. As yet, transplant programmes, like other health programmes, have suffered from sex disparities, as shown by several reports in the literature.^{9,31,33-35} Nevertheless, to date, none of the proposed tools to mitigate these inequities in LT has obtained a multicentre external validation, and neither reached widespread use in clinical practice.^{11,36}

The recent proposal of replacing MELDNa with sex-balanced MELD 3.0 rejuvenated the debate around disparities in access to LT.¹⁴ MELD 3.0 allowed a net reclassification of 8.8% of deceased patients (14.9% for women) who would have had better chances of survival if the system were in place.¹⁴ In parallel, Sealock et al. demonstrated sex-related differences in all the variables of the MELDNa score. The logarithmic nature of the score magnifies these differences, thus resulting in higher MELDNa values in males compared to females, despite higher decompensation counts in female LT candidates.¹⁵ A score map of sex-adjusted MELDNa values was calculated based on the lab values of a single-centre mixed population of healthy controls

and CLD and LT population. This provided a less pronounced effect on the reclassification of patients on the waiting list (i.e. female individuals had a 1% higher transplant rate compared with a .7% higher rate for male individuals).¹⁵

Rodríguez-Perálvarez et al. created the GEMA-Na score using a very rigorous statistical methodology.¹⁶ A European population was adopted for creating and internally validating the score, while an Australian cohort was used for external validation. GEMA-Na showed improved discrimination in predicting mortality or delisting due to clinical deterioration within the first 90 days after waiting list inclusion compared with MELDNa (Internal validation: .77 vs. .74, $p = .006$; external validation: .77 vs. .75, $p = .014$). Considering the patients that would have been differently prioritized for transplantation, one in 21 deaths could potentially be avoided by using GEMA-Na instead of MELDNa, showing an improved discrimination and a significant reclassification benefit.¹⁶

The composition of the CLD population in Italy has its peculiarities, mainly with a higher donor and recipient age and prevalence of HCC. Our work is the first to test the fit of the United States-derived MELD 3.0 and sex-adjusted MELDNa in a LT population outside the United States, and to explore the performances of GEMA-Na in the setting of a South European population.

Female LT recipients were 19.1% of our population, while the female-to-male ratio of Italian CLD patients reported in the literature ranges between 1:3 and 1:7.^{28,31,37,38} However, historical data do not reflect the evolution in the epidemiology of decompensated liver cirrhosis, due to the introduction of HCV direct-acting antivirals (DAAs). If we look at the sex differences by aetiology, male sex is clearly prevalent in the case of alcoholic cirrhosis, which is actually the first indication for LT, with a 2.3 male-to-female ratio. In a recent European study analysing the ELTR database regarding post-transplant outcomes, male sex was clearly prevalent over female (70.5% vs. 29.5%). Nevertheless, only few data about sex differences are reported in the literature with an intention-to-treat perspective, which is also a limitation of our work.³⁹ The American series¹⁴ reported almost twice higher values (36.8%). As well documented in numerous studies, we confirm the imbalance between female patients potentially needing a LT and those who effectively receive it.

Analysing different MELD-based scoring systems, we compared the performance of each in our study population and confirmed the improved dropout prediction ability of the novel scores compared with MELD. However, the new sex-adjusted scores do not appear to add significant discriminatory power concerning MELDNa, which is currently used in most allocation systems.

With the intent to unshell the magnitude of the prediction improvement with the novel scores, we applied the Brier Skill score.²³ Interestingly, having set MELD as the reference, we observed increased performance with the sex-adjusted scores compared to currently widely used scores for the prediction of 90-day dropout, with GEMA-Na resulting in the best score with a 4.4% prediction improvement.

Focussing on the number of avoidable dropouts after reclassifying the patients with the new scores, it is noteworthy that one

in nine dropout cases could be theoretically avoided using the GE-MA-Na, with one dropout avoided every 285 patients listed for transplantation. Even better, when only women were considered, using GEMA-Na instead of MELD would potentially save one in three dropouts and could save one dropout per 54 women included.

This datum potentially expands the role of the novel scores to a yet untested population that is epidemiologically different, if not unique.

As previously reported, in the work by Sealock et al,¹⁵ female LT candidates presented lower values of MELDNa components and the score itself, despite a higher number of registered disease decompensation events before transplant. It is plausible that lower MELDNa scores observed in women affect the timing of listing. Female patients may reach the threshold for listing (MELDNa 15 in many transplant programmes^{40,41}), probably after a longer disease course and increased CLD-related complications. Therefore, many female patients are supposedly deprived of the opportunity of LT due to overcoming clinical deterioration and fatal complications. In this context, using sex-corrected scores may allow timely access to the list, thus expanding the chance of receiving life-saving LT.

Our study suggests the potential of the novel sex-adjusted scores in minimizing and possibly resolving sex-related inequality in access to LT, as testified by the higher median value of these scores in the female group.

In detail, in our series, 6.2% of cases have biliary cirrhosis as an indication of LT, with a 16.0 female versus 3.9% male distribution, respectively. Biliary cirrhosis mainly results from primary biliary cholangitis (PBC) and primary sclerosing cholangitis (PSC). PBC has an overwhelming female sex prevalence (92%), while PSC has a 2:1 male-to-female ratio. Some pieces of evidence exist of higher waitlist mortality and dropout of patients with PBC, especially in the higher MELDNa category,⁴² while PSC subjects show a lower mortality rate.^{42,43} MELD and MELDNa scores show lower predictive ability in subjects suffering from PBC if compared to other aetiologies and prove to be suboptimal tools in this category of patients.^{44,45} It is possible to hypothesize, at least in part, a sex-driven effect to explain this and speculate that the new sex-based scores can correct or mitigate this effect. New and specifically designed studies are needed to address this issue.

Recent evidence in cirrhotic patients with ascites revealed that long-term albumin administration improves survival,⁴⁶ prevents complications⁴⁷ and reduces ascitic recurrence.⁴⁸ Obviously, such an administration can affect the result of the MELD 3.0 score with albumin. No data regarding albumin supplementation were available in our database. How albumin supplementation can affect the ability of the score to predict outcomes in cirrhotic subjects with ascites enlisted for liver transplant is unknown. Future studies are needed to balance the potential beneficial effects of long-term albumin supplementation with the reduction in the chance of LT resulting from a lower MELD 3.0 score.

Our work has some limitations. First, the study is retrospective, but this characteristic is shared with most studies testing new prognostic scores in LT. The proposal of such new scores in clinical practice will inevitably require a prospective testing phase in real life to

verify the effects on liver graft allocation and effects on the waiting list, as pointed out in some comments on Kim and Sealock papers.^{25,27,49} Moreover, unlike Kim et al., we decided to include HCC subjects in the analysed population. As we decided not to consider dropout due to tumour progression among the study outcomes, we consider this aspect unable to affect our results significantly, also considering that dropout from the list because of death from tumour progression is an unlikely event.

Another relevant issue to underline is that the reduced sample size of the present study and the reduced number of events occurring at 90 days hampered our possibility to perform specific analyses of model accuracy in the group of women. In addition, the study was not powered to detect differences in waiting list outcomes between men and women.

In conclusion, sex adjustment of scores utilized for allocating organs should represent a useful tool in removing the penalization against female candidates. The performance of the novel scores in our work showed a superiority when compared with MELD. However, no statistically significant differences in discrimination ability were observed when the new scores were compared to MELDNa. Yet, several issues remain unsolved, such as muscle mass weighing calculation and height of CLD patients (e.g. tall women with heavier muscle masses should not be prioritized over males with opposite characteristics) and the type of kidney dysfunction.

Recently, a height-adjusted MELD has been proposed in the literature using a simulation algorithm on data from the US Scientific Registry of Transplant Recipients (SRTR). This approach could provide an additional tool to counterbalance woman penalization in the allocation system without overcorrecting for taller women.⁵⁰

A balance between further refinements to MELD-based allocation systems and avoiding over-sophisticated scores should be the objective of future research to improve equity without hampering usability and the widespread of novel systems.

AUTHOR CONTRIBUTIONS

Giuseppe Marrone, Valerio Giannelli, Alfonso Wolfango Avolio, Gabriele Spoletini, Quirino Lai are responsible for the conception, design, analysis and writing of the study; Giuseppe Marrone, Valerio Giannelli, Flaminia Ferri, Ilaria Lenci, Raffaella Lionetti, Martina Milana, Tommaso Maria Manzia, Quirino Lai are involved with the collection and interpretation of data; Giuseppe Marrone, Salvatore Agnes, Alfonso Wolfango Avolio, Leonardo Baiocchi, Giammauro Berardi, Giuseppe Maria Ettore, Stefano Ginanni Corradini, Antonio Grieco, Nicola Guglielmo, Massimo Rossi, Giuseppe Maria Ettore, Gabriele Spoletini, Giuseppe Tisone and Quirino Lai participated in data management, review and editing of the manuscript.

FUNDING INFORMATION

None about the here presented work.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare about the present study.

DATA AVAILABILITY STATEMENT

If requested by the editors, data and information on which the results of this manuscript are based will be provided without any restriction. This is a retrospective multicentre observational study investigating the data of cirrhotic patients enlisted for receiving a first LT. Data were prospectively recorded for allocation needs.

ETHICS STATEMENT

The Institutional Review Board of the coordinator centre approved the present study.

ORCID

Giuseppe Marrone  <https://orcid.org/0000-0002-9475-3948>

Alfonso Wolfango Avolio  <https://orcid.org/0000-0003-2491-7625>

Flaminia Ferri  <https://orcid.org/0000-0001-8553-5589>

Antonio Grieco  <https://orcid.org/0000-0002-0544-8993>

Raffaella Lionetti  <https://orcid.org/0000-0003-2827-4591>

REFERENCES

- Kamath P, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology*. 2001;33:464-470.
- Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124:91-96.
- De La Mata M, Cuende N, Huet J, et al. Model for end-stage liver disease score-based allocation of donors for liver transplantation: a spanish multicenter experience. *Transplantation*. 2006;82:1429-1435.
- Ravaioli M, Grazi GL, Ballardini G, et al. Liver transplantation with the Meld system: a prospective study from a single European center. *Am J Transplant*. 2006;6:1572-1577. doi:10.1111/J.1600-6143.2006.01354.X
- Biggins SW, Kim WR, Terrault NA, et al. Evidence-based incorporation of serum sodium concentration into MELD. *Gastroenterology*. 2006;130:1652-1660.
- 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.
- 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;300:2371-2378.
- Nephew LD, Goldberg DS, Lewis JD, Abt P, Bryan M, Forde KA. Exception points and body size contribute to gender disparity in liver transplantation. *Clin Gastroenterol Hepatol*. 2017;15:1286-1293.e2.
- Locke JE, Shelton BA, Olthoff KM, et al. Quantifying sex-based disparities in liver allocation. *JAMA Surg*. 2020;155:e201129. doi:10.1001/JAMASURG.2020.1129
- Cullaro G, Sarkar M, Lai JC. Sex-based disparities in delisting for being "too sick" for liver transplantation. *Am J Transplant*. 2018;18:1214-1219.
- Lai JC, Terrault NA, Vittinghoff E, Biggins SW. Height contributes to the gender difference in wait-list mortality under the MELD-based liver allocation system. *Am J Transplant*. 2010;10:2658-2664.
- 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.
- 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;7:685-692.
- 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.
- Sealock JM, Ziogas IA, Zhao Z, et al. Proposing a sex-adjusted sodium-adjusted MELD score for liver transplant allocation. *JAMA Surg*. 2022;157:618-626. doi:10.1001/JAMASURG.2022.1548
- Rodríguez-Perálvarez ML, Gómez-Orellana AM, Majumdar A, et al. Development and validation of the Gender-Equity Model for Liver Allocation (GEMA) to prioritise candidates for liver transplantation: a cohort study. *Lancet Gastroenterol Hepatol*. 2023;8:242-252.
- Avolio AW, Franco A, Schlegel A, et al. Development and validation of a comprehensive model to estimate early allograft failure among patients requiring early liver retransplant. *JAMA Surg*. 2020;155:e204095. doi:10.1001/JAMASURG.2020.4095
- Ravaioli M, Lai Q, Sessa M, et al. Impact of MELD 30-allocation policy on liver transplant outcomes in Italy. *J Hepatol*. 2022;76:619-627.
- Cillo U, Burra P, Mazzaferro V, et al. A multistep, consensus-based approach to organ allocation in liver transplantation: toward a "blended principle model.". *Am J Transplant*. 2015;15:2552-2561.
- Lai Q, Magistri P, Lionetti R, et al. Sarco-Model: a score to predict the dropout risk in the perspective of organ allocation in patients awaiting liver transplantation. *Liver Int*. 2021;41:1629-1640.
- Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *Jama*. 1982;247:2543-2546.
- Le Kang L, Chen W, Petrick NA, et al. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med*. 2015;34:685-703.
- Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78:1-3.
- Demler OV, Paynter NP, Cook NR (2015) tests of calibration and goodness-of-fit in the survival setting. *Stat Med*. 2015;34:1659-1680. doi:10.1002/sim.6428
- Goudsmit B, Putter H, Van Hoek B. The model for end-stage liver disease 3.0: an update without proven accuracy. *Gastroenterology*. 2022;162:1781-1782.
- O'leary JG, Bajaj JS. MELD 3.0: one small step for womankind or one big step for everyone? *Gastroenterology*. 2022;162:1780-1781.
- Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*. 2018;67:123-133.
- Stroffolini T, Sagnelli E, Gaeta GB, et al. Characteristics of liver cirrhosis in Italy: evidence for a decreasing role of HCV aetiology. *Eur J Intern Med*. 2017;38:68-72.
- Fattovich G, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology*. 2004;127:S35-S50.
- Annicchiarico BE, Siciliano M, Avolio AW, et al. Treatment of chronic hepatitis C virus infection with pegylated interferon and ribavirin in cirrhotic patients awaiting liver transplantation. *Transplant Proc*. 2008;40:1918-1920.
- Rodríguez-Castro KI, De Martin E, Gambato M, Lazzaro S, Villa E, Burra P. Female gender in the setting of liver transplantation. *World J Transplant*. 2014;4:229-242.
- Avolio AW, Lai Q, Cillo U, Romagnoli R, De Simone P. L-GRAFT and EASE scores in liver transplantation: need for reciprocal external validation and comparison with other scores. *J Hepatol*. 2021;75:729-731.
- Breathett K, Yee E, Pool N, et al. Association of gender and race with allocation of advanced heart failure therapies. *JAMA Netw Open*. 2020;3:e2011044.
- Melk A, Schmidt BMW, Geyer S, Epping J. Sex disparities in dialysis initiation, access to waitlist, transplantation and transplant outcome in German patients with renal disease—a population based analysis. *PLoS One*. 2020;15:e0241556. doi:10.1371/JOURNAL.PONE.0241556

35. Melk A, Babitsch B, Borchert-Mörlins B, et al. Equally interchangeable? How sex and gender affect transplantation. *Transplantation*. 2019;103:1094-1110.
36. Huo SC, Huo TI, Lin HC, et al. Is the corrected-creatinine model for end-stage liver disease a feasible strategy to adjust gender difference in organ allocation for liver transplantation? *Transplantation*. 2007;84:1406-1412.
37. Stroffolini T, Sagnelli E, Almasio P, Ferrigno L, Crax A, Mele A. Characteristics of liver cirrhosis in Italy: results from a multicenter national study. *Dig Liver Dis*. 2004;36:56-60.
38. Sagnelli E, Stroffolini T, Sagnelli C, et al. Gender differences in chronic liver diseases in two cohorts of 2001 and 2014 in Italy. *Infection*. 2018;46:93-101.
39. Germani G, Zeni N, Zanetto A, et al. Influence of donor and recipient gender on liver transplantation outcomes in Europe. *Liver Int*. 2020;40:1961-1971.
40. Merion RM, Schaubel DE, Dykstra DM, Freeman RB, Port FK, Wolfe RA. The survival benefit of liver transplantation. *Am J Transplant*. 2005;5:307-313.
41. Burra P, Burroughs A, Graziadei I, et al. EASL clinical practice guidelines: liver transplantation. *J Hepatol*. 2016;64:433-485.
42. Nagai S, Safwan M, Kitajima T, Yeddula S, Abouljoud M, Moonka D. Disease-specific waitlist outcomes in liver transplantation - a retrospective study. *Transpl Int*. 2021;34:499-513.
43. Goldberg D, French B, Thomasson A, Reddy KR, Halpern SD. Waitlist survival of patients with primary sclerosing cholangitis in the post-MELD era. *Liver Transpl*. 2011;17:1355-1363.
44. Zhou K, Dodge JL, Xu E, Emamullee J, Kahn JA. Excess liver transplant waitlist mortality for patients with primary biliary cholangitis under MELD-Na allocation. *Clin Transplant*. 2022;36:e14527. doi:10.1111/CTR.14527
45. Stauffer K, Kivaranovic D, Rasoul-Rockenschaub S, Soliman T, Trauner M, Berlakovich G. Waitlist mortality and post-transplant survival in patients with cholestatic liver disease—impact of changes in allocation policy. *HPB*. 2018;20:916-924.
46. Caraceni P, Riggio O, Angeli P, et al. Long-term albumin administration in decompensated cirrhosis (ANSWER): an open-label randomised trial. *Lancet*. 2018;391:2417-2429. doi:10.1016/S0140-6736(18)30840-7
47. Fernández J, Clària J, Amorós A, et al. Effects of albumin treatment on systemic and portal hemodynamics and systemic inflammation in patients with decompensated cirrhosis. *Gastroenterology*. 2019;157:149-162.
48. Sandi BB, Leão GS, de Mattos AA, de Mattos ÂZ. Long-term albumin administration in patients with cirrhosis and ascites: a meta-analysis of randomized controlled trials. *J Gastroenterol Hepatol*. 2021;36:609-617.
49. Wood NL, Segev DL, Gentry SE. Sex-adjusted model for end-stage liver disease scores for liver transplant allocation. *JAMA Surg*. 2022;21:3296-3304.
50. Bernards S, Lee E, Leung N, et al. Awarding additional MELD points to the shortest waitlist candidates improves sex disparity in access to liver transplant in the United States. *Am J Transplant*. 2022;22:2912-2920. doi:10.1111/ajt.17159

How to cite this article: Marrone G, Giannelli V, Agnes S, et al. Superiority of the new sex-adjusted models to remove the female disadvantage restoring equity in liver transplant allocation. *Liver Int*. 2023;00:1-10. doi:10.1111/liv.15735