“Model for Early Allograft Function” outperforms “Early Allograft Dysfunction” as a predictor of transplant survival

Authors

Ina Jochmans, MD, PhD^{1, 2}

Steffen Fieuws, PhD³

Diethard Monbaliu, MD, PhD^{1, 2}

Jacques Pirenne, MD, PhD^{1, 2}

Affiliations

¹ Lab of Abdominal Transplant Surgery, Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium

² Abdominal Transplant Surgery, University Hospitals Leuven, Leuven, Belgium

³ Interuniversity Centre for Biostatistics and Statistical Bioinformatics, Department of Public Health, KU Leuven, Leuven, Belgium

Correspondence

Ina Jochmans

Abdominal Transplant Surgery

University Hospitals Leuven

Herestraat 49

3000 Leuven, Belgium

ina.jochmans@uzleuven.be

Authorship page

Contributions to work.

IJ designed the study; IJ, DM, and JP were responsible for data collection; SF together with

IJ analysed and interpreted the data; IJ wrote the manuscript.

All authors reviewed the manuscript and approved the final manuscript, and all agree to be accountable for all aspects of the work.

Disclosures

IJ, DM, and JP are holders of a named chair at the KU Leuven from the Centrale Afdeling voor Fractionering van plasma (CAF-DCF) Belgium.

JP is holder of a named chair at the KU Leuven from the IGL group

Funding

No funding was received for this work.

Abbreviations

BMI, body mass index

DBD, donation after brain death

DCD, donation after circulatory death

CVA, cerebrovascular accident

EAD, Early Allograft Dysfunction

HBV, hepatitis B virus

HCV, hepatitis C virus

HTK, histidine-tryptophan-ketoglutarate

IDI, integrated discrimination improvement

NASH, non-alcoholic steatohepatitis

NRI, net reclassification improvement

MEAF, Model for Early Allograft Function

MELD, model for end stage liver disease

UW, University of Wisconsin solution

Abstract

Background: The Model of Early Allograft Function (MEAF) grades the severity of liver graft dysfunction. Unlike the categorical Early Allograft Dysfunction (EAD) classification, MEAF is a continuous score, based on bilirubin, international normalised ratio, and alanine aminotransferase within 3 days posttransplant.

Methods: Multivariable regression models were used to validate the MEAF score in 660 liver-only transplants performed between 2000-2014. MEAF performance for prediction of transplant survival was compared with that of EAD in univariable and multivariable models by means of Harrell's *c*-indices, integrated discrimination improvement, and net reclassification improvement.

Results: Median donor and recipient age was 52y (IQR 41-62) and 58y (50-64), respectively. MELD-score was 15 (11-21), cold ischemia time 8.0h (6.4-9.7), MEAF 4 (3-6). EAD occurred in 182 (27.6%) cases. Transplant survival was 93%, 90%, and 88% at 3, 6, and 12 months. Both MEAF and EAD were independent predictors of transplant survival within 3, 6, and 12 months. MEAF outperformed EAD as predictor of transplant survival, either when used as a standalone parameter or when corrected for additional independent predictors of transplant survival.

Conclusion: MEAF is a more accurate predictor of transplant loss than the commonly used EAD classification. As a continuous score grading graft dysfunction, MEAF provides additional, granulated information that could be used both clinically and as a surrogate endpoint of transplant survival in clinical trials.

Introduction

Initial dysfunction of the liver graft in the immediate posttransplant period influences both graft and patient survival.¹⁻³ The parameters used to define liver dysfunction indicate hepatocellular damage and synthetic impairment, thought to be caused by ischemia-reperfusion injury. Ischemia-reperfusion injury is associated with oxidative damage and an inflammatory response leading to cellular injury.⁴⁻⁶

The condition is most often defined by Olthoff's criteria of early allograft dysfunction (EAD).² EAD is a dichotomous variable – available 7 days posttransplant – that does not allow grading of severity of the dysfunction. EAD is either present or it is not. Nevertheless, the transplant community intuitively understands there is a continuum between immediate graft function and primary non-function. A score grading the severity of initial dysfunction after liver transplantation would allow us to objectively assess graft dysfunction and any risks associated with that grade of dysfunction. It might also serve as an outcome measure in clinical trials and allow more granulated comparison between studies, centers, and reported outcomes.

The Model of Early Allograft Function (MEAF) has recently been suggested as an alternative way to classify early graft dysfunction after liver transplantation in such a continuous score.⁷ Based on bilirubin, international normalized ratio, and alanine aminotransferase within the first 3 postoperative days, a score between 0 and 10 is given. Pareja et al, have shown that MEAF is an independent predictor of graft and patient survival at 3, 6, and 12 months after liver transplantation.⁷

We wanted to validate these findings in our population of liver transplant recipients. Additionally, we set out to determine whether MEAF was a better predictor of outcome than the commonly used EAD classification.

Methods

Patient population

This retrospective study analysed prospectively collected data of 778 adult whole liver-only transplants performed at our centre between 2000-2014, acute liver failure (n=105) and early hepatic artery thrombosis (n=8) were excluded to match the Pareja et al, exclusion criteria. A further 5 cases were excluded because MEAF could not be calculated due to missing lab values. Donor and recipient demographics as well as procedural data and outcome data were collected prospectively. Recipients consented to the use of their data for research purposes and the protocol was approved by our Medical Ethical Committee (s59634).

MEAF was calculated as follows: $MEAF = (\text{"score ALT"} + \text{"score INR"} + \text{"score bilirubin"})$, where $\text{"score ALT"} = 3.29 / (1 + e^{-1.9132(\ln(ALT_{\max.3\text{days}}) - 6.1723)})$, $\text{"score INR"} = 3.29 / (1 + e^{-6.8204(\ln(INR_{\max.3\text{days}}) - 0.6658)})$, $\text{"score bilirubin"} = 3.4 / (1 + e^{-1.8005(\ln(bilirubin_{\text{day3}}) - 1.0607)})$.⁷ EAD was defined by the Olthoff et al, criteria where EAD is present when bilirubin >10mg/dL on day 7 or international normalized ratio >1.6 on day 7 or alanine or aspartate aminotransferases >2000 IU/L within the first 7 days.²

Transplant survival was calculated from the date of transplantation to the date of last follow up or graft loss (defined as need for relisting or patient death).

Statistical methods

Continuous variables are presented as median (interquartile range), categorical variables as number and percentage. Kaplan-Meier curves were constructed and the log-rank test was used to compare survival. The follow-up time was calculated using a Kaplan-Meier curve, censoring deceased patients at the moment of their death.⁸ Median follow-up time is given as the Kaplan-Meier estimate for potential follow-up, obtained by censoring the deceased patients.

First, logistic and Cox regression models were used to validate the MEAF score. Logistic regression was used to evaluate transplant survival within 3, 6, and 12 months. All recipients had a minimal follow up of 18 months so the logistic regression models were not influenced by differences in length of follow-up.

Second, we compared MEAF performance for prediction of transplant survival with that of the EAD score. Initially the value of MEAF and EAD as sole predictors of transplant survival was checked in 2 univariable models. Subsequently we assessed whether the initial findings would remain valid when other independent predictors were added to the models. Backward selection was used to construct the multivariable models retaining predictors with $p < 0.157$. This value corresponds to using the Aikake Information Criterion for model selection that requires that the increase in model Chi^2 has to be larger than 2 times the degrees of freedom.⁹ Initial variables to be considered were based on those identified in the literature as predictors of transplant survival. Restricted cubic splines were used to test for nonlinearity of the MEAF score.¹⁰ The performance of the prediction models was quantified using Harrell's c -index¹⁰ and compared between models using an approach presented by

DeLong et al,¹¹ and Kang et al,¹² in the logistic and Cox regression models, respectively. Harrell's *c*-index quantifies the predictive capacity of a predictor where a perfect discrimination between events and non-events would yield a *c*-index of 1. A predictor with a *c*-index of 0.5 is not better than chance.

As Harrell's *c*-index may not be sensitive enough to capture the added value of a new marker, 2 recently introduced alternatives that quantify the improvement of the new versus the old marker were also determined: the net reclassification improvement (NRI) and the integrated discrimination improvement (IDI).¹³ The continuous NRI quantifies the average improvement in discrimination with the new versus the old marker (ie, MEAF vs. EAD). For recipients that experience transplant loss, the NRI is the probability of having a higher risk score with the new prediction model minus the probability of having a lower risk score with the new prediction model ("event NRI"). For recipients who do not experience transplant loss, the NRI is the probability of having a lower risk score with the new prediction model minus the probability of having a higher risk score with the new prediction model ("non-event NRI"). The NRI is reported as the average of the "event NRI" and the "non-event NRI". The IDI is the difference in the discrimination slope of 2 prediction models.^{13,14} This discrimination slope is the difference in mean predicted probability of transplant loss between patients with an event (transplant loss) and patients without an event (transplant survival). Within the multivariable survival models for MEAF and EAD, NRI and IDI values have been determined using the predicted probabilities of transplant survival at 3, 6, and 12 months.

All analyses were performed using SAS software (version 9.2 of the SAS System for Windows) and R software (R version 3.2.3, The R foundation for Statistical Computing).

Specifically, the packages survIDINRI¹⁵ and compareC¹² were used for evaluation of the survival models. SAS-macros developed by N. Cook were used for the NRI and IDI of logistic regression models.

Results

Median follow-up of the 660 liver-only transplants was 8.6 years with a minimal follow-up of 18 months. Donor, transplant, and recipient characteristics and are shown in Table 1. Median intensive care and hospital stay were 3 (2-7) and 19 (14-29) days, respectively. EAD developed in 182 recipients (28%), median MEAF was 4 (3-6; min 0, max 10). Five recipients developed primary non-function (0.8%). Transplant and patient survival were 93% and 95% at 3 months, 90% and 93% at 6 months, and 88% and 91% at 12 months, respectively. Transplant survival until 1 year after transplantation, split by MEAF and EAD, are depicted in Fig. 1, which visualises that both MEAF and EAD are associated with transplant survival at the univariable level. Table S1 presents the univariable analysis for the dataset and shows that both MEAF and EAD are associated with transplant survival. As shown in Fig. S1, the effect of MEAF on transplant survival is most pronounced in the first year after transplantation.

MEAF outperforms EAD as a predictor of transplant survival

Both MEAF and EAD were independent predictors of transplant survival within 3, 6, and 12 months (Fig. 2). The decreasing odd ratio again points towards the fact that the effect MEAF and EAD on transplant survival are most pronounced early after transplantation. Other independent predictors of transplant survival were recipient age and MELD, and

additionally female donor (6 months), alcoholic cirrhosis (12 months), and hepatitis C cirrhosis (12 months) (Tables S2-S4).

MEAF *per se* -so at the univariable level- was a better predictor of transplant survival compared to EAD within 3, 6, and 12 months as is shown by Harrell's *c*-indices (Table 2). Although the difference in *c*-indices was only significant for transplant survival within 3 months, it is known that *c*-indices aren't always the most sensitive measure to capture the added value of a new predictor. We therefore also calculated the NRI and IDI. These are subtler ways to ascertain whether 1 predictor is more accurate than another. Positive NRIs or IDIs signify better predictive capacity of MEAF over EAD. NRIs were positive and significant for transplant survival within 3 and 6 but not 12 months (Fig. 3, Panel A). All IDIs -although relatively small in magnitude- were positive and significant for transplant survival within 3, 6, and 12 months (Fig. 3, Panel B). This confirms that MEAF was a better predictor of transplant survival compared to EAD when no other predictors were considered.

When other independent predictors were added to the models, the significance of the difference in *c*-index disappeared (Table 2). However, NRI and IDI remained positive and significant at 3 and 6 but not 12 months (Table 3), confirming MEAF's superiority as predictor of transplant survival.

MEAF allows the chance of transplant loss to be estimated

As there was no evidence for nonlinearity of the MEAF score (data not shown), the impact of a 1-unit difference in MEAF score on the log odds of transplant loss can be considered as constant in the full range of the MEAF score. The chance to develop transplant

loss with any given score can therefore be calculated using the following function: $p = \exp(\alpha + \beta * \text{MEAF}) / (1 + \exp(\alpha + \beta * \text{MEAF}))$, where p is the probability of transplant loss, α ($= -4.8184$) the intercept, and β ($= 0.4374$) is the log of the regression coefficient from the univariable logistic regression model for transplant loss. Applying this function to our dataset, we plotted the probability of transplant loss within 3 months posttransplant for each possible MEAF value (Fig. 4). As an example, the probability of transplant loss within 3 months after transplantation with an MEAF score of 5 is 7% (95% confidence interval 5%-9%), this increases to 22% (15%-30%) with an MEAF of 8 and 37% (24%-53%) for the maximal MEAF of 10.

Purely based on the data, there is no natural candidate cut-off for the MEAF score at which point the risk for transplant loss increases exponentially because of the linear association between MEAF and transplant outcome. This is also visually appreciated in the Kaplan-Meier curves of Fig. 1 where no natural categorisation of MEAF in function of transplant survival can be detected.

Discussion

The MEAF score has been proposed as a measure of severity of graft dysfunction early after liver transplantation.⁷ The objective variables on which the score is based are available within 3 days posttransplant. The score implies that the individual scoring of a recipient's initial posttransplant status can provide more precise information on the expected outcome than current categorical classifications like EAD.⁷

We could indeed confirm that MEAF is an independent predictor of not only 3 but also 6 and 12-month transplant survival. Furthermore, MEAF predicts transplant loss more

accurately than EAD, either when considered as a stand-alone parameter or when added to other independent predictors. The effect of MEAF seems most pronounced early after transplantation. As it encompasses biochemical results obtained in the first 3 days posttransplant, this is not an entirely unsurprising finding as many more factors that will influence transplant survival come into play in the later postoperative setting. These factors (eg, length of ICU stay, the occurrence of ischemic type biliary strictures) were not considered in these analyses as we wanted to determine the potential of MEAF on top of other confounders available at the same time in the posttransplant setting.

As always, scoring systems are imperfect (the *c*-indices of MEAF and EAD are well below 0.8, the regarded minimum for an efficient and reliable predictor) and should therefore be used to complement and aid clinical judgment and decision-making, not drive it. One could argue that MEAF, with *c*-indices at univariable level between 0.609 and 0.644, does not seem to perform much better than EAD, which has *c*-indices between 0.640 and 0.727 (Table 2). Nevertheless, the NRI and IDI scores indicate that MEAF is indeed a better predictor than EAD.

Although it is unlikely for any score based on just a few variables to ever be a perfect predictor of (long term) outcome, a major benefit of the MEAF score is its continuity. This continuity and the absence of nonlinearity allow the clinician to estimate the probability of transplant loss based on MEAF alone and for any given score. This additional information provides an objective -albeit imperfect- measure of the risk of transplant loss shortly after transplantation. The MEAF score provides additional information on prognosis that is of use when informing patients and caregivers about expected posttransplant outcomes. When risk-

benefits are being weighed eg, when considering the need for a retransplant, this information could be used to aid clinical decision-making, as suggested by Pareja et al.⁷ In addition, the MEAF score might prove to be a predictor of other posttransplant complications such as acute kidney injury. Indeed, ischemia-reperfusion injury and liver dysfunction early after transplantation have recently been shown to be associated with the development of acute kidney injury after liver transplantation.^{16,17} When specific interventions to alleviate parenchymal injury, suboptimal synthetic function, or acute kidney injury have been identified, the MEAF score – objectifying the grade of liver dysfunction – could be seen as a trigger to initiate these therapies.

Pareja et al, who designed the MEAF score, also envisaged the MEAF score as a potential warning tool where high MEAF scores could signify a risk for the development of primary non-function. Although we do indeed see poor outcomes with MEAF scores above 8 (Fig. 1), with a probability for transplant loss within 3 months for this score of 22%, transplant survival was still well above 50% until 5 years posttransplant. Of the 5 cases that did develop primary non-function, MEAF was indeed above 8 (median 9), however, another 48 recipients with an MEAF above 8 did not develop primary non-function. These findings stress the importance of using the MEAF in addition to other clinical assessment tools and to not let it drive decision-making.

As the MEAF score associates independently with transplant loss within the first year after transplantation, it has the potential to serve as a proxy endpoint in clinical trials. As 1-year transplant survival in liver transplantation exceeds 90% in many centers, studies that have transplant loss as primary endpoint need large sample sizes to achieve adequate study

power. A surrogate endpoint that occurs more frequently but is at the same time a good predictor of transplant outcome, could be valuable. The Food and Drug Administration increasingly uses validated surrogate endpoints to approve new therapeutic interventions¹⁸ and has recently held a public workshop to identify potential surrogate endpoints in kidney transplantation.¹⁹ Furthermore, several clinical trials now use liver graft dysfunction as primary outcome. Of 137 currently open randomised controlled trials in liver transplantation registered at clinicaltrials.gov, 8% have graft dysfunction (5/12) or evolution of transaminases (7/12) as primary endpoint, 15% look primarily at patient or graft survival (SDC, <http://links.lww.com/TP/B456>). As MEAF is a better predictor of transplant survival and provides a continuous scoring, it might be a more valuable surrogate endpoint than EAD. However, it remains to be shown whether the reduction of graft dysfunction posttransplant effectively influences transplant survival.

There are limitations to this study. It concerns a relatively small single-center dataset. It might be that in larger datasets there would be evidence of a non-linear relationship between MEAF and transplant survival, allowing the search for a cut-off MEAF score that identifies an exponentially higher risk for transplant loss. There were also few cases of primary non-function and as such the usefulness of the MEAF score to predict primary non-function could not be investigated in detail.

In conclusion, MEAF outperforms EAD as an independent predictor of transplant survival after liver transplantation. It grades severity of posttransplant graft dysfunction and provides the clinician with additional information that might help to determine further

treatment options. Nevertheless, its predictive accuracy is still moderate and therefore other clinically relevant variables remain important to provide context to the score.

ACCEPTED

Acknowledgments

We thank Tine Wylin for management of the research database.

ACCEPTED

References

1. Deschênes M, Belle SH, Krom RA, Zetterman RK, Lake JR. Early allograft dysfunction after liver transplantation: a definition and predictors of outcome. National Institute of Diabetes and Digestive and Kidney Diseases Liver Transplantation Database. *Transplantation*. 1998; 66(3):302-310.
2. Olthoff KM, Kulik L, Samstein B, et al. Validation of a current definition of early allograft dysfunction in liver transplant recipients and analysis of risk factors. *Liver Transpl*. 2010; 16(8):943-949.
3. Hudcova J, Scopa C, Rashid J, Waqas A, Ruthazer R, Schumann R. Effect of early allograft dysfunction on outcomes following liver transplantation. *Clin Transplant*. 2017; 31(2).
4. Rosen HR, Martin P, Goss J, et al. Significance of early aminotransferase elevation after liver transplantation. *Transplantation*. 1998; 65(1):68-72.
5. Tilg H, Vogel W, Aulitzky WE, et al. Evaluation of cytokines and cytokine-induced secondary messages in sera of patients after liver transplantation. *Transplantation*. 1990; 49(6):1074-1080.
6. Busuttil RW, Tanaka K. The utility of marginal donors in liver transplantation. *Liver Transpl*. 2003; 9(7):651-663.
7. Pareja E, Cortes M, Hervas D, et al. A score model for the continuous grading of early allograft dysfunction severity. *Liver Transpl*. 2015; 21(1):38-46.
8. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials*. 1996; 17(4):343-346.
9. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *J R Stat Soc Series C*. 1999; 48(3):313-329.

10. Harrell FE. *Regression Modeling Strategies*. New York, NY: Springer; 2001.
11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44(3):837-845.
12. 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(4):685-703.
13. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011; 30(1):11-21.
14. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008; 27(2):157-172 discussion 207-112.
15. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med*. 2013; 32(14):2430-2442.
16. Jochmans I, Meurisse N, Neirynck A, Verhaegen M, Monbaliu D, Pirenne J. Hepatic ischemia-reperfusion injury associates with acute kidney injury in liver transplantation: prospective cohort study. *Liver Transpl*. 2017; 23(5):634-644.
17. Wadei HM, Lee DD, Croome KP, et al. Early allograft dysfunction after liver transplantation is associated with short- and long-term kidney function impairment. *Am J Transplant*. 2016; 16(3):850-859.

18. U.S. Food & Drug Administration. FDA facts: biomarkers and surrogate endpoints. <http://www.fda.gov/aboutfda/innovation/ucm512503.htm>. Updated July 22 2016. Accessed 19 November 2016.
19. U.S. Food & Drug Administration. Surrogate endpoints for clinical trials in kidney transplantation; public workshop. <http://www.fda.gov/Drugs/NewsEvents/ucm449248.htm>. Updated April 1 2016. Accessed 19 November 2016.

Figure Legends

Figure 1: Kaplan-Meier curves for transplant survival during the first posttransplant year split per MEAF (Panel A) and EAD (Panel B).

Figure 2: Forest plots representing adjusted odds ratios for the Model of Early Allograft Function (MEAF) score (Panel A) and early allograft dysfunction (EAD) (Panel B) in 2 separate multivariable logistic regression models for transplant survival within 3, 6, and 12 months after liver transplantation with either MEAF or EAD as additional variable. Other variables in the models were: donor and recipient gender, donor and recipient BMI, donor type (brain death, circulatory death), trauma donor, preservation solution, recipient age, model of end stage liver disease, indication, cold ischemia time, surgery time, cytomegalovirus mismatch.

CI, confidence interval.

Figure 3: Net reclassification improvement (NRI) and integrated discrimination improvement (IDI) for MEAF versus EAD to predict transplant survival at 3, 6, and 12 months.

CI, confidence interval.

Figure 4: Graphic representation of the probability of transplant loss within 3 months. The dashed lines represent the 95% confidence interval for the predicted value. The predication was obtained from a logistic regression model assuming linearity on the logit scale.

MEAF, model for early allograft function.

Figure 1

Figure 2

Figure 3

Figure 4

Table 1: Donor, transplant, and recipient characteristics

Donor		
Age (y)		52 (40.5-62.0)
Gender	Male	378 (57%)
	Female	282 (43%)
Body mass index (kg/m ²)		24 (22-26)
Cause of death ^a	Trauma	277 (42%)
	CVA	313 (47%)
	Anoxia	46 (7%)
	Other	19 (3%)
Donor type	DBD	601 (91%)
	DCD	59 (9%)
Donor warm ischemia time (min) ^b	In DCD transplants	21 (16-26)
Recipient		
Age (y)		58 (50-64)
Gender	Male	421 (64%)
	Female	239 (36%)
Body mass index (kg/m ²)		26 (23-29)
Lab MELD		15 (11-21)
Indication for transplant ^c	Cholestatic disease	98
	HBV cirrhosis	98
	HCV cirrhosis	102
	Cholestatic disease	98
	Hepatocellular carcinoma	205

	NASH	42
	Postalcoholic cirrhosis	225
	Polycystic disease	30
	Metabolic diseases	38
	Retransplant	59
Transplant		
Preservation fluid ^a	HTK	154 (23%)
	UW	459 (70%)
	Other	41 (6%)
Cold ischemia time (h)		8.0 (6.4-9.7)
Surgery time (h)		5.5 (4.7-6.7)
Outcome		
Primary nonfunction		5 (0.8%)
Patient death	within 3 months	34 (5.2%)
	within 6 months	49 (7.4%)
	within 12 months	89 (9.2%)
Transplant loss	within 3 months	46 (7.0%)
	within 6 months	63 (9.5%)
	within 12 months	79 (12.0%)

^a missing values; ^b donor warm ischemia is defined as the time between withdrawal of therapy and start of cold flush of the organs; ^c not mutually exclusive

DBD, donation after brain death; DCD, donation after circulatory death; CVA, cerebrovascular accident; HBV, hepatitis B virus; HCV, hepatitis C virus; HTK, histidine-tryptophan-ketoglutarate; NASH, non-alcoholic steatohepatitis; MELD, model for end stage liver disease; UW, University of Wisconsin solution

Table 2: Comparison of the discriminative ability (Harrell's *c*-indices) of EAD and MEAF as a predictor of transplant survival after liver transplantation

	c-index EAD model (95% CI)	c-index MEAF model (95% CI)	p value of the difference
Univariable models			
within 3 months	0.644 (0.569;0.719)	0.727 (0.643;0.812)	0.036
within 6 months	0.628 (0.564;0.693)	0.693 (0.618;0.767)	0.060
within 12 months	0.609 (0.551;0.667)	0.640 (0.567;0.713)	0.35
Multivariable models^a			
within 3 months	0.740 (0.668;0.812)	0.782 (0.709;0.856)	0.13
within 6 months	0.732 (0.669;0.794)	0.752 (0.687;0.816)	0.41
within 12 months	0.747 (0.693;0.802)	0.747 (0.690;0.804)	0.99

CI, confidence interval; EAD, early allograft dysfunction; MEAF, model for early allograft function

^a Other variables in the models were: donor and recipient gender, donor and recipient; BMI, donor type (brain death, circulatory death), trauma donor, preservation solution, recipient age, model of end stage liver disease, indication, cold ischemia time, surgery time, cytomegalovirus mismatch.

Table 3: NRI and IDI of the MEAF and EAD multivariable models for transplant survival quantifying the improvement of the discriminatory capacity of MEAF over EAD.

	NRI (95% CI) ^a	p value NRI	IDI (95% CI)	p value IDI
within 3 months	0.232 (0.084-0.379)	0.002	0.052 (0.034-0.069)	0.003
within 6 months	0.170 (0.042-0.299)	0.01	0.042 (0.026-0.057)	0.002
within 12 months	0.025 (-0.093-0.144)	0.67	0.015 (0.002-0.028)	0.12

CI, confidence interval; IDI, integrated discrimination improvement; NRI, net reclassification improvement

^a the continuous NRI is reported as the average of the ‘event NRI’ and the ‘non-event NRI’ and quantifies the average improvement in discrimination.¹³

Covariables in the multivariable models were obtained after backward selection (final models are presented in Table S2-S4). MEAF and EAD were then entered the models separately. Table S5-S6 show additional information on NRI and IDI development.

**Supplementary Digital Content to “Model for Early Allograft Function” outperforms
“Early Allograft Dysfunction” as a predictor of transplant survival**

Supplementary Methods: Systematic Search of ClinicalTrials.gov

To identify primary endpoints of current randomised controlled clinical trials in liver transplantation, the ClinicalTrials.gov database was queried on 19th November 2016.

Using the advanced search option on www.clinicaltrials.gov the following search criteria were entered:

Search Terms: “liver transplantation”

Recruitment: Open Studies

Study Type: Interventional Studies

Study Results: All Studies

All other search fields were not altered.

Of 247 identified studies, 137 were randomised controlled trials. Primary outcome in these 137 studies was classified as:

‘survival’ –patient or graft survival or survival in general (disease free survival was counted as other)

‘graft dysfunction’ – any description of graft dysfunction

‘transaminases’ –the use of asparate aminotransferase or alanine aminotransferase

‘other’ – all others

Table S1: Results from univariable logistic regression models for transplant survival after liver transplantation in a dataset of 660 adult only liver transplants.

		OR (95%CI) 3 months	p value	OR (95%CI) 6 months	p value	OR (95%CI) 12 months	p value
Donor age (y)		1.007 (0.990;1.025)	0.43	1.004 (0.989;1.019)	0.62	1.011 (0.997;1.025)	0.13
Donor gender	Female	1.654 (0.906;3.019)	0.10	1.902 (1.125;3.216)	0.016	1.522 (0.950;2.438)	0.081
	Male	#	.	#	.	#	.
Donor BMI (kg/m ²)		0.956 (0.871;1.049)	0.34	0.960 (0.885;1.040)	0.32	0.961 (0.893;1.033)	0.28
Donor type			.		.		.
	DCD	0.444 (0.105;1.881)	0.27	0.483 (0.147;1.591)	0.23	0.819 (0.340;1.972)	0.66
	DBD	#	.	#	.	#	.
Donor WIT if DCD		1.061 (0.912;1.233)	0.44	1.065 (0.939;1.207)	0.33	1.043 (0.947;1.149)	0.39
Donor Cause of Death ^a			0.89		0.79		0.69
	Anoxia	1.370 (0.442;4.248)	0.58	1.004 (0.332;3.040)	0.99	0.784 (0.263;2.340)	0.66
	CVA	1.088 (0.571;2.073)	0.80	1.159 (0.662;2.027)	0.61	1.241 (0.752;2.049)	0.40
	Other	1.693 (0.363;7.905)	0.50	1.977 (0.537;7.269)	0.31	1.544 (0.425;5.608)	0.51
	Trauma	#	.	#	.	#	.

		OR (95%CI) 3 months	p value	OR (95%CI) 6 months	p value	OR (95%CI) 12 months	p value
Preservation solution ^b			0.04		0.15		0.028
	HTK	1.677 (0.469;5.994)	0.43	2.000 (0.566;7.067)	0.28	1.600 (0.576;4.442)	0.37
	UW	0.730 (0.211;2.528)	0.62	1.143 (0.337;3.878)	0.83	0.783 (0.292;2.095)	0.63
	Other	#	.	#	.	#	.
Recipient age (y)		1.030 (0.998;1.062)	0.06	1.030 (1.003;1.058)	0.028	1.027 (1.003;1.052)	0.025
Recipient gender	Female	1.260 (0.685;2.319)	0.46	1.177 (0.692;2.004)	0.55	1.090 (0.671;1.770)	0.73
	Male	#	.	#	.	#	.
Recipient BMI (kg/m ²)		1.033 (0.974;1.095)	0.28	1.031 (0.980;1.085)	0.24	1.033 (0.986;1.082)	0.17
Recipient MELD		1.035 (1.007;1.063)	0.01	1.033 (1.008;1.059)	0.011	1.042 (1.018;1.067)	0.0006
Transplant indications ^c							
	Retransplant	1.940 (0.826;4.553)	0.13	1.823 (0.850;3.910)	0.12	2.035 (1.028;4.029)	0.041
	Metabolic disease	1.154 (0.341;3.906)	0.82	1.473 (0.554;3.920)	0.44	1.725 (0.733;4.061)	0.21
	HCC	1.199 (0.638;2.253)	0.57	1.122 (0.646;1.949)	0.68	0.965 (0.579;1.607)	0.90
	Polycystic disease	0.448 (0.060;3.367)	0.44	0.316 (0.042;2.359)	0.26	0.244 (0.033;1.817)	0.17
	Alcoholic cirrhosis	0.664 (0.337;1.310)	0.24	0.632 (0.350;1.142)	0.13	0.622 (0.364;1.062)	0.082

	OR (95%CI) 3 months	p value	OR (95%CI) 6 months	p value	OR (95%CI) 12 months	p value
HCV cirrhosis	0.650 (0.251;1.686)	0.38	0.903 (0.431;1.892)	0.79	0.675 (0.326;1.398)	0.29
Cholestatic disease	1.660 (0.795;3.466)	0.18	1.566 (0.816;3.006)	0.18	1.274 (0.684;2.374)	0.44
NASH cirrhosis	1.444 (0.492;4.238)	0.50	1.640 (0.663;4.060)	0.28	1.517 (0.650;3.541)	0.34
HBV cirrhosis	0.357 (0.048;2.663)	0.32	0.251 (0.034;1.865)	0.18	0.194 (0.026;1.436)	0.11
Cold ischemia time (h)	1.137 (0.995;1.301)	0.06	1.128 (1.004;1.267)	0.043	1.115 (1.003;1.239)	0.044
Time of surgery (h) ^d	1.210 (1.019;1.436)	0.03	1.152 (0.986;1.344)	0.074	1.161 (1.009;1.337)	0.037
CMV mismatch	0.762 (0.417;1.390)	0.38	0.949 (0.565;1.596)	0.84	0.839 (0.524;1.342)	0.46
Length ICU (d)	1.023 (0.991;1.057)	0.16	1.011 (0.980;1.043)	0.50	1.011 (0.986;1.038)	0.39
Length of stay (d)	1.005 (0.995;1.015)	0.30	1.004 (0.996;1.013)	0.33	1.004 (0.997;1.011)	0.24
EAD (y/n)	3.466 (1.887;6.365)	<.0001	3.076 (1.815;5.213)	<.0001	2.649 (1.639;4.282)	<.0001
MEAF score	1.549 (1.339;1.791)	<.0001	1.440 (1.270;1.632)	<.0001	1.321 (1.181;1.478)	<.0001

^a 5 missing values, ^b 6 missing values, ^c indications are not mutually exclusive ^d, 7 missing values.

reference category.

Significant OR are denoted in bold.

BMI, body mass index; CMV, cytomegalo virus; CVA, cerebro vascular accident; DBD, donation after brain death; DCD, donation after circulatory determination of death; EAD, early allograft dysfunction; HBV, hepatitis B; HCC, hepatocellular carcinoma; HCV, hepatitis C; HTK, histidine-tryptophan-ketoglutarate; ICU, intensive care units; MEAF, model of early allograft function; MELD, model of end stage liver disease; NASH, nonalcoholic steatohepatitis; OR, odds ratio; UW, University of Wisconsin solution.

Table S2: Results from a multivariable logistics regression model for transplant loss within 3 months.

	OR (95%CI)	p value
Donor Gender (Female vs Male)	1.719 (0.929;3.180)	0.0843
Preservation solution		0.0543
HTK	1.305 (0.348;4.897)	0.6928
UW	0.590 (0.161;2.154)	0.4242
Recipient age (y)	1.036 (1.003;1.070)	0.0342
Recipient MELD	1.034 (1.005;1.063)	0.0211
Cold ischemia time (h)	1.146 (0.991;1.326)	0.0654

The model is based on 649 observations using a backward regression procedure with $p=0.157$ as critical p value to stay in the model. This value corresponds to using the Aikake Information Criterion for model selection that requires that the increase in model Chi^2 has to be larger than 2 times the degrees of freedom.¹ The p values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally too liberal. Warm ischemia time, the length of ICU and hospital stay are not included as potential predictors.

AUC, area under the curve; HTK, histidine-tryptophan-ketoglutarate; MELD, model of end stage liver disease; University of Wisconsin solution

AUC original=0.692 (0.614;0.771)

Corrections for over-optimism:

- Optimism-corrected AUC based on simple bootstrap=0.650007
- Optimism in original AUC evaluated by advanced bootstrap=0.092431999999999
- Optimism-corrected AUC based on advanced bootstrap=0.600068000000001

Table S3: Results from a multivariable logistics regression model for transplant loss within 6 months.

	Odds ratio (95%CI)	p value
Donor gender (Female vs Male)	1.827 (1.067;3.130)	0.0282
Recipient age (y)	1.037 (1.009;1.067)	0.0091
Recipient MELD	1.037 (1.011;1.064)	0.0056
Alcoholic cirrhosis	0.620 (0.338;1.136)	0.1215
Cold ischemia time (h)	1.115 (0.985;1.262)	0.0841

The model is based on 649 observations using a backward regression procedure with $p=0.157$ as critical p value to stay in the model. This value corresponds to using the Aikake Information Criterion for model selection that requires that the increase in model Chi^2 must be larger than 2 times the degrees of freedom.¹ The p values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally too liberal. Warm ischemia time, the length of ICU and hospital stay are not included as potential predictors.

AUC, area under the curve; MELD, model for end stage liver disease

AUC original=0.691 (0.626;0.756)

Corrections for over-optimism:

- Optimism-corrected AUC based on simple bootstrap=0.644096
- Optimism in original AUC evaluated by advanced bootstrap=0.076168
- Optimism-corrected AUC based on advanced bootstrap=0.614732

Table S4: Results from a multivariable logistics regression model for transplant loss within 12 months.

	Odds ratio (95%CI)	p value
Donor gender (Female vs Male)	1.523 (0.929;2.498)	0.0953
Preservation solution		0.1255
HTK	1.273 (0.431;3.760)	0.6629
UW	0.731 (0.256;2.088)	0.5578
Recipient age (y)	1.037 (1.011;1.063)	0.0044
Recipient MELD	1.041 (1.016;1.066)	0.0013
Alcoholic cirrhosis	0.489 (0.277;0.864)	0.0138
HCV cirrhosis	0.443 (0.199;0.989)	0.0470
HBV cirrhosis	0.175 (0.023;1.327)	0.0918

	Odds ratio (95%CI)	p value
Cold ischemia time (h)	1.109 (0.987;1.246)	0.0805

The model is based on 649 observations using a backward regression procedure with $p=0.157$ as critical p value to stay in the model. This value corresponds to using the Aikake Information Criterion for model selection that requires that the increase in model Chi^2 must be larger than 2 times the degrees of freedom.¹ The p values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally too liberal.

AUC, area under the curve; HBV, hepatitis B virus; HCV, hepatitis C virus; HTK, histidine-tryptophan-ketoglutarate; MELD, model of end stage liver disease; University of Wisconsin solution

AUC original=0.724 (0.665;0.782)

Corrections for overoptimism:

- Optimism-corrected AUC based on simple bootstrap=0.66302
- Optimism in original AUC evaluated by advanced bootstrap=0.068049999999999
- Optimism-corrected AUC based on advanced bootstrap=0.655850000000001

Table S5: Net Reclassification Improvement

	Events			Nonevents			Total NRI (95%CI)
	with prediction			with prediction			
	up	down	NRI events	up	down	NRI nonevents	
within 3 months	0.587	0.413	0.174	0.355	0.645	0.289	0.232 (0.084;0.379), p=0.0020
within 6 months	0.556	0.444	0.111	0.385	0.615	0.229	0.170 (0.042;0.299), p=0.0095
within 12 months	0.487	0.513	-0.026	0.462	0.538	0.076	0.025 (-0.093;0.144), p=0.6738

Evaluation based on the continuous net reclassification improvement (NRI).²

NRI: net reclassification improvement:

- Events: improvement in discrimination within group of subjects with an event ie, the probability of having a higher risk score with the new prediction model minus the probability of having a lower risk score (the higher, the better).
- Nonevents: improvement in discrimination within group of subjects without an event ie, the probability of having a lower risk score with the new prediction model minus the probability of having a higher risk score (the higher, the better).
- Total NRI: average of the events NRI and the nonevents NRI.

Table S6: Integrated Discrimination Improvement

	Discrimination (Yates) slope		IDI
	Model EAD	Model MEAF	
within 3 months	0.056 (0.038;0.074)	0.108 (0.084;0.131)	0.052 (0.034;0.069), p=0.0026
within 6 months	0.066 (0.046;0.085)	0.107 (0.084;0.130)	0.042 (0.026;0.057), p=0.0017
within 12 months	0.085 (0.062;0.107)	0.100 (0.077;0.123)	0.015 (0.002;0.028), p=0.1186

Evaluation of the added value of the diagnostic information based on the integrated discrimination improvement (IDI).²

The discrimination slope (Yates slope) is the difference in predicted probability between events and nonevents. The IDI is the difference in discrimination slope between 2 models.

Figure S1: Kaplan-Meier curves for transplant survival split per MEAF (Panel A) and EAD (Panel B).

References

1. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *J R Stat Soc Series C*. 1999; 48(3):313-329.
2. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008; 27(2):157-172 discussion 207-112.