Received: 14 November 2016 | Accepted: 23 January 2017

DOI: 10.1111/liv.13378

LIVER TRANSPLANTATION

WILEY Liver INTERNATIONAL

External validation of the Donor Risk Index and the Eurotransplant Donor Risk Index on the French liver transplantation registry

Audrey Winter[1,2] | Cyrille Féray[3] | Etienne Audureau[4] | René Écochard[5] | Christian Jacquelinet[6] | Françoise Roudot-Thoraval[3] | Christophe Duvoux[3] | Jean-Pierre Daurès[1,2] | Paul Landais[1,7]

[1]Department of Biostatistics, UPRES EA2415, Clinical Research University Institute, University of Montpellier, Montpellier, France

[2]Beau Soleil Clinic, Languedoc Mutualité, Montpellier, France

[3]Department of Hepatology, Henri Mondor University Hospital, Créteil, France

[4]Department of Biostatistics and Public Health, Henri Mondor University Hospital, Créteil, France

[5]Laboratory Biostatistics-Health, CNRS 5558 - LBBE, Lyon, France

[6]Agence de la Biomédecine, Saint Denis, Villejuif, France

[7]Department of Biostatistics & Public Health, Nîmes University Hospital, Montpellier, France

Correspondence
Audrey Winter, Clinical Research University Institute, Montpellier CEDEX 5, France.
Email: audrey.winter89@gmail.com

Funding information
The present study belongs to the program entitled « Optimatch ». It was funded by the French Ministry of Health within the framework of the national Clinical Research Hospital Program.

Handling Editor: Alessio Aghemo

Abstract

Background & Aims: A major limitation to liver transplantation is organ shortage leading to the use of non-optimal liver grafts. The Donor Risk Index has been validated and recommended to select donors/organs. The Eurotransplant Donor Risk Index was derived from the Donor Risk Index. The objective of our study was to perform an external validation of both Donor Risk Index and Eurotransplant-Donor Risk Index against the French liver transplantation Cristal registry according to recommendations of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Methods: Liver transplantations performed in France between 2009 and 2013 were used to perform the validation study for the Donor Risk Index and the Eurotransplant-Donor Risk Index respectively. We applied on the French data the models used to construct the Donor Risk Index and the Eurotransplant-Donor Risk Index respectively.

Results: Neither the Donor Risk Index nor the Eurotransplant-Donor Risk Index were validated against this dataset. Discrimination and calibration of these scores were not preserved according to our data. Important donor and candidates differences between our dataset and the Organ Procurement and Transplantation Network or the Eurotransplant datasets may explain why the Donor Risk Index and the Eurotransplant-Donor Risk Index appeared unadapted to the French transplant registry.

Conclusion: Neither of these risk indexes were suitable to optimize the French liver allocation system. Thus, our next step will be to propose a general adaptive model for a Donor Risk Index.

KEYWORDS

donor risk, external validation, liver transplantation, prognostic score

1 | INTRODUCTION

Liver transplantation (LT) is a life-saving procedure in cases of liver failure or liver malignancy. In France, the "Agence de la Biomédecine" (ABM) is responsible for managing the waiting list and distributing the grafts. Similarly to other countries, the number of newly registered patients continues to increase in France: +12.3% in 2012, +5.7% in 2013. As a consequence, the past decade has seen a growing number

Abbreviations: ABM, Agence de la Biomédecine; BMI, body mass index; COD, cause of death; DCD, donation after cardiac death; DRI, Donor Risk Index; ECD, expanded criteria donor; ET-DRI, Eurotransplant-Donor Risk Index; GammaGT, gamma glutamyl transpeptidase; HCC, hepatocellular carcinoma; HR, hazard ratio; ICU, intensive care unit; LT, liver transplantation; MELD, model for end stage liver disease; OPTN, Organ Procurement and Transplantation Network; PI, prognostic index; SE, standard error; TRIPOD, transparent reporting of a multivariable prediction model for individual prognosis Or diagnosis.

of LT candidates placed on a waiting list. Despite a progressive increase in the number of liver transplants conducted and substantial efforts to promote organ donation, the number of candidates by far exceeds the number of organs available. Patients have to wait longer and mortality whilst on the waiting list persists.

To decrease organ shortage, surgical techniques have been improved and donor selection criteria were expanded (expanded criteria donor, ECD). However, an unequivocal definition of these ECD livers[1] was not available. Several risk factors must be considered such as donor age >65 or high level of alanine amino-transferase.[1,2] If any one of these variables is present then the graft is classified as ECD. Considering the stringent parameters set, it is likely that a significant number of grafts are classified as ECD. To address this issue, the Donor Risk Index (DRI) was developed in 2006 from the American Organ Procurement and Transplantation Network (OPTN) database.[3] The donor covariates were: age, cause of death (COD), race, donation after cardiac death (DCD), partial/split liver, height, location (local/regional/national) and cold ischaemia time.

Recently, Blok et al.[4] validated the DRI on the Eurotransplant dataset and created the Eurotransplant-Donor Risk Index (ET-DRI)[1] derived from the DRI. The ET-DRI includes all the covariates of the DRI except race and height. However, two covariates were added (latest serum gamma glutamyl transpeptidase [GammaGT] and rescue offer) and one covariate, location, was modified. Originally, the DRI recorded location as: local/regional/national; for the ET-DRI it was modified to local/regional/extra-regional. The mean value of the DRI was 1.45 in the OPTN database vs 1.70 for the Eurotransplant database.

The DRI or ET-DRI have been defined for the OPTN or Eurotransplant databases respectively. However, an external validation of these scores is necessary to explore whether they are applicable to other populations, the French population in the present case.

In France, liver allocation is conducted at the national level by the "score foie", which takes into account the LT indication, the model for end stage liver disease (MELD) score and other variables such as the waiting-time, the distance between organ procurement and recipient's centre, and the alpha-protein score.[5] However, the quality of the graft is not taken into account despite the impact this has on post-transplant survival.[1,3]

The DRI can be applied to assess the quality of a graft, and whilst the DRI alone is not sufficient for matching the donor and recipient, it might provide a complementary support for a better French allocation system. The goal is to use it as a donor indicator for optimal matchings between donors and recipients according to new modelling approaches,[6] as currently the only existing matching is ABO group. The DRI has never been validated on the French database, thus the aim of our study was to perform an external validation of the DRI on the French database. An external validation of the ET-DRI was also performed.

Key points

- Validating DRI is a key step prior to apply it to match a donor to a recipient within a liver transplantation framework.
- We performed an original statistical procedure to achieve the validation of the DRI and the ET-DRI, to assess the model performance in several steps, including discrimination and calibration.
- Neither of them were validated against the French dataset.
- Given these results we propose to improve the DRI approach creating a general adaptive model of the DRI to improve liver transplantation matching.

2 | MATERIAL

2.1 | Data

Information relating to LT performed in France between January 4, 2009 and December 31, 2013 was obtained from the ABM. The patients' informed consent was required. The study was conducted according to the approval given by the Comité de Protection des Personnes and by the Comité Consultatif pour le Traitement de l'Information en matière de Recherche dans le domaine de la Santé. Authorization was also obtained from the Commission Nationale de l'Informatique et des Libertés (agreement No. 915206). The data provided were de-identified beforehand.

In accordance with previous works,[1,3] recipients under 18 years or multiple organ transplants were not included. Follow-up began at LT and ended with one of the following events: lost to follow-up, death, graft loss (re-transplantation) or end of the study, as of December 31, 2014. The outcome was death or graft loss. The final analysis was performed in 3677 LT for the external validation of the DRI and 4558 for the ET-DRI. The DRI and the ET-DRI were calculated according to refs (3) and (1) respectively. According to French law, and similarly to the Eurotransplant database, race was not registered in our database, therefore all patients were considered as Caucasians. The definition of "national" sharing within OPTN differed to that used for Eurotransplant. For Eurotransplant, all countries except Germany were regarded as a single region.[1] Consequently, "national" was replaced by "extra-regional". In France, we considered a transplant as a "local" share if the estimated time between the procurement and transplantation centres was ≤1 hour and "regional" otherwise. For the "rescue allocation" variable of Eurotranplant, we used a similar covariate in France named an "Hors Tour" donor, whereby, after at least five consecutive refusals, the graft is supplied to a transplant team that chooses the candidate according to the graft proposed which is not necessarily poor.

2.2 | Definitions of the variables

In France, all donors are hospitalized in intensive care. The ICU length of stay is based on the number of nights spent, rather than the number of days. Therefore, a stay of 0 days is still observable. The estimated distance between donor and recipient was calculated on the basis of a geographic model taking into account road distances in minutes.

2.3 | The adjustment covariates

According to the DRI,[3] the Cox model was adjusted on all covariates, whether available and similar (Table 1) in the French database. A similar procedure was performed for the ET-DRI[1] (Table 2).

3 | METHODS

3.1 | Comparison of the datasets

A comparison of the OPTN and Eurotransplant datasets with the French dataset was performed using Chi^2 tests or ANOVA, when appropriate. The data were extracted from the tables of the two original articles.[1,3]

3.2 | External validation

We performed an external validation of the DRI, and the ET-DRI with our database, according to Royston and Altman.[7] Model performance was evaluated considering two fundamental aspects: discrimination and calibration. Discrimination, known as "separation", allows differentiation of patients' prognoses through risk estimates from the model. Calibration reflects the prediction accuracy; if well-calibrated, a score assigns the appropriate probability at each level of predicted risk.[7] We used the same model, namely a Cox model, in exactly the same way as in refs (3) and (1) respectively. For the sake of simplification, we shall present the whole procedure for the DRI. The same procedure was performed for validating the ET-DRI.

3.2.1 | Regression on the prognostic index in the validation data

In order to obtain a prognostic index (PI), we used ln(DRI) and ln(ET-DRI). Then, for each DRI, we built a Cox model with the PI and all the variables of adjustment. We used the following model:

$$\beta_{DRI} \ln(DRI) + \beta_{adjustment} X, \quad (1)$$

where X represents the adjustment covariates (see the "Adjustment covariates" section).

An accepted way of performing an external validation is to estimate the coefficient of the PI on the validation data.[8,9] The discrimination is considered good when this coefficient is equal to 1 and poor if the slope is lower than 1; a coefficient>1 is considered very good.

In order to test $\hat{\beta}_{DRI} = 1$, we used a likelihood ratio test between the model (Equation 1) and the same model with the coefficient of ln(DRI) set at the slope obtained minus 1. Then we were able to test the following equations: $\hat{\beta}_{DRI} - 1 = 0$ or $\hat{\beta}_{DRI} = 1$.

3.2.2 | Check model misspecification/fit

A possible reason for a PI coefficient <1 is poor adjustment of one or more covariates. To test whether one or more of the DRI covariates needed an adjustment we used model 2:

$$\beta_{DRI} \ln(DRI) + \beta_{adjustment} X + \beta^* Z, \quad (2)$$

where the adjustment covariates X and Z were those used for the construction of the DRI.

In these models, the β_{DRI} was set at 1, and the $\beta_{adjustment}$ was estimated using model 1, and fixed at the estimated value in model 2. Next, we performed a likelihood ratio test between these two models to test the following equation: $\beta^*=0$. The proportional hazards risk assumption was checked using the Schoenfeld residuals.

3.2.3 | Measures of discrimination

We retained three discrimination indexes[7]: Harrell C-index,[10] Gönen and Heller K statistic,[11] and Royston and Sauerbrei R_D^2[12] to independently evaluate the model discrimination.

Harrell C-index is defined as the proportion of all usable patient pairs in which the predictions and outcomes are concordant.[10] Gönen and Heller's K statistic is used to evaluate the discriminatory power and the predictive accuracy of nonlinear statistical models. It is a function of the regression parameters and the covariate distribution only and is therefore asymptotically unbiased.[11] Royston and Sauerbrei's R_D^2 is a measure of the proportion of explained variation, based on D, a measure of the ability of a model to discriminate between good and poor patient outcomes.[12]

When these measures were available in the construction model, we compared them with those obtained from the validation model.

3.2.4 | Kaplan–Meier curves for groups at risk

We plotted the survival curves according to the groups at risk using the Kaplan–Meier estimates. When possible, we did a visual comparison of these curves with those of the construction dataset which was sufficient to evaluate the calibration. Indeed, if the survival curves of the construction and validation data for each group at risk were superimposable, then the visual calibration was considered as preserved. Discrimination can also be evaluated from the curves, with more separated survival curves indicating better discrimination.

3.2.5 | Hazard ratios between groups at risk

According to Royston and Altman,[7] in contrast to P-values for comparing groups at risk, evaluating hazard ratio (HR) is a sensible check of discrimination. The more the Kaplan-Meier curves are separated, the larger the HR. HR were then estimated for each group at risk using Cox model. All analyses were performed using R software, version 3.3.0 (R Development Core Team, 2016).

4 | RESULTS

4.1 | Comparing the databases

To perform the external validation of the DRI and ET-DRI models, we applied the models following the exact procedures proposed

TABLE 1 Differences in donor, transplantation and recipient characteristics between the OPTN dataset[3] and the French dataset (*P*-values, Chi[2] tests and ANOVA when appropriate)

	OPTN: 1998-2002	French: 2009-2013	*P*-value
Donor characteristics			
Age (%)			
0-17	12.0	1.9	<.001
18-39	39.2	19.8	
40-49	18.7	16.4	
50-59	16.3	18.8	
60-69	9.5	19.8	
70	4.3	23.2	
Race (%)			
Afro-American	11.7	0.0	-
Other	2.6	0.0	
White	85.6	100.0	
COD (%)			
CVA	43.6	59.5	<.001
Trauma	44.6	25.3	
Anoxia	8.6	12.2	
Other	3.0	3.0	
Donation after cardiac death (DCD) (%)	1.1	0.0	-
Partial/split liver (%)	2.0	5.0	<.001
Gender (%)			
Male	59.5	55.3	<.001
Female	40.5	44.7	
Age mean (SD)	39	54.2 (18.4)	-
Length mean (SD)	171.3 (12.4)	169.5 (10.0)	<.001
HBcAb status (%)	4.9	4.5	.23
Sodium>170 mEq/L (%)	2.6	0.7	<.001
Transplantation characteristics			
ABO compatibility (%)			
Identical	90.6	97.36	<.001
Compatible	8.4	2.61	
Incompatible	1.1	0.03	
Sharing (%)			
Local	73.3	42.6	<.001
Regional	21.2	57.4	
National	5.5	0.0	
Cold ischaemia time, mean (SD)	8.2 (3.8)	8.4 (4.4)	.08
DRI, mean (SD)	1.34 (-)	1.65 (0.40)	-
Recipient characteristics			
Age (%)			
18-24	2.4	2.1	<.001
25-34	4.4	3.8	
35-44	16.8	10.9	
45-54	41.7	32.5	
55-64	26.8	41.9	
65+	7.9	8.8	
Female (%)	35.5	25.6	<.001
Race (%)			
Afro-American	7.7	0.0	-
Other	6.0	0.0	
White	86.3	100.0	
Hepatitis C status (%)	39.2	24.5	<.001
Status at time of LT (%)			
ICU	21.8	13.4	<.001
Hospitalized not in ICU	13.8	14.1	
Not hospitalized	64.4	72.5	
Angina (%)	2.3	-	-
Cerebrovascular disease (%)	0.5	-	-
Previous transfusion (%)	34.9	-	-
Status 1 (%)	8.3	3.7	<.001
On life support (%)	9.4	-	-
Previous liver transplant (%)	8.8	7.4	.005
CMV positive (%)	58.3	58.7	.68
Tumour found at transplant (%)	3.8	-	-
Previous abdominal surgery (%)	36.8	29.8	<.001
Inotropes for BP support (%)	5.4	-	-
On dialysis (%)	4.6	2.9	<.001
Diabetes (%)	-	22.8	-
Portal vein thrombosis (%)	2.6	-	-
Grade III ou IV encephalopathy (%)	28.3	5.4	<.001
Hepatitis B positive (%)	6.9	20.0	<.001
Total bilirubin (mg/dL), mean (SD)	7.09 (9.83)	7.00 (9.96)	.61
Serum creatinine (mg/dL), mean (SD)	1.31 (1.08)	0.99 (0.60)	<.001
Serum albumin (g/dL), mean (SD)	2.84 (0.77)	3.26 (2.07)	<.001
BMI, mean (SD)	28.11 (11.51)	25.85 (4.82)	<.001
SGOT/AST, mean (SD)	269.87 (1000.90)	-	-
Follow-up, median (years)	3	2.5	

in the original articles.[1,3] We rendered our dataset compatible with the DRI and ET-DRI datasets after preliminary assessment of the available variables. We had to take into account: dates, criteria of inclusion or non-inclusion, coding of the variables, graft procurement and transplantation locations, and length of follow-up. The comparisons of the datasets appear on Tables 1 and 2 for the OPTN and Eurotransplant vs the French dataset, respectively, and several differences are apparent. In France for instance, organ removal necessitated an ICU stay, whilst in the OPTN and the Eurotransplant datasets it was not the case. Age distribution differed between the French dataset and both the OPTN and Eurotransplant sets. In the OPTN database, 51.2% of donors were aged <39 (Table 1), vs 28.9% in the Eurotransplant set (Table 2), while in France this population only accounted for around 20%. Mean donor age was 39 years in the OPTN database, and 54 in the French one. The French dataset had a much higher burden of older patients in comparison with the previous datasets; whilst 9% of donors were aged over 70 years in the Eurotransplant dataset and 4.3% in the OPTN dataset, whereas 23% of the French dataset were more than 70 years old.

The distribution of the DRI from the OPTN dataset and the ET-DRI from the Eurotransplant dataset, vs the French dataset are presented on Figures 1 and 2 respectively. For the OPTN dataset, the percentage of patients peaked at a DRI of 1.0-1.2 and then steadily dropped with increasing DRI. In contrast, in the French dataset there was an 9% jump in the percentage of patients with a DRI of <1.0 to a DRI 1.0-1.2, with a more gradual increase followed by a plateau at a DRI of 1.4-1.6. However, 21% of donors had a DRI >2, against 6% in the OPTN dataset (Figure 1). In the Eurotransplant dataset, a similar trend was seen for both datasets, with a low percentage patients with an ET-DRI of <1.0, gradually increasing as the ET-DRI increased to 1.6-1.8, followed by a slight drop at ET-DRI 1.8-2.0. However, in the ET-DRI >2 category, the percentage of patients in the Euro-transplant dataset jumped to nearly 30%, whilst the French dataset decreased to just over 15%. Of the two validation datasets, the ET-DRI showed a more similar distribution of the Eurotransplant dataset to the French dataset, except at the two extremes.

4.2 | External validation of the Donor Risk Index and Eurotransplant-Donor Risk Index

4.2.1 | Regression on prognostic index in the French validation dataset

The slopes on the PI were 0.57 (SE 0.15) and 0.64 (SE 0.16) for the DRI and ET-DRI respectively. The slope was significantly different from 1 (*P*-value<.001), thus, the discrimination should be considered poor.

4.2.2 | Checking model misspecification/fit

For DRI, the β^* was 0, therefore the *P*-value was not significant (*P*=.28). No adjustment on the DRI covariates was needed. One variable did not meet the proportional hazards assumption: hepatitis C status.

For ET-DRI, as the *P*-value was significant (*P*<.001), an adjustment of the β^* coefficients on the ET-DRI covariates was needed. Three variables did not meet the proportional hazards assumption: the recipient gender, diagnosis and highest level of GammaGT.

4.2.3 | Measures of discrimination

For the DRI, the Harrell C-index was 0.601 (SD: 0.010), the Gönen and Heller *K* statistic was 0.608 (SD: 0.006) and the Royston and Sauerbrei R_D^2 was 0.442 (SD: 0.037). For the ET-DRI, the Harrell C-index was 0.622 (SD: 0.009), the Gönen and Heller *K* statistic was 0.623 (SD: 0.006) and the Royston and Sauerbrei R_D^2 was 0.498 (SD: 0.030).

4.2.4 | Kaplan–Meier curves for groups at risk

Failure free survival per DRI and ET-DRI category is given on Table 3 at month 3, year 1 and year 3. There was no survival decrease according to the groups at risk.

4.2.5 | Hazard ratios between groups at risk

Hazard ratio between groups at risk are presented on Table 4 for the DRI and ET-DRI respectively. Each HR is expressed according to the lowest group at risk (Group 0). For both of them, survival was not different between the different groups at risk. Moreover, the risk did not increase as expected when comparing group 0 with the other groups.

5 | DISCUSSION

Looking for the best match between a graft and its recipient is a challenge. Within the scope of a national project called OPTIMATCH, our aim was to optimize the current liver allocation system in France. A donor risk index, which scores the characteristics of the graft, is one of the key factors for a better matching. Rather than developing a new donor risk score, we wondered whether the existing scores, the DRI and ET-DRI, were valid for the French dataset. In this study, we performed an external validation of these two scores according to the current methodology proposed by the TRIPOD statement.[14]

External validation of a model allows the assessment of its performance on an independent database. Through the external validation steps,[7] the two DRIs did not meet the requirements to be validated on the French database. This result should not be interpreted as a failure of the model, although it does underline that to be used in clinical practice, a prognostic score has to be validated and demonstrate its effectiveness.[7]

Considering the differences between the three populations studied, it might seem inevitable that we would be unable to validate either of the previous DRIs against our dataset. However, in our opinion, our demonstration was necessary prior to developing a general adaptive model for a donor risk index.

For the DRIs, according to the slopes on the PI, the discrimination was not sufficient. For the DRI, since the estimation of the β^*

TABLE 2 Differences in donor, transplantation and recipient characteristics between the Eurotransplant dataset[1] and the French dataset (*P*-values Chi[2] tests and ANOVA when appropriate)

	EuroTransplant: 2003-2007	French: 2009-2013	*P*-value
Donor characteristics			
Age (%)			
<40	28.9	21.9	<.001
40-49	23.1	16.7	
50-59	22.9	18.9	
60-69	16.5	19.5	
70	8.7	23.0	
Gender (%)			
Male	53.8	55.2	.17
Female	46.2	44.8	
COD (%)			
CVA	63.0	60.4	<.001
Trauma	26.7	24.8	
Anoxia	6.9	11.8	
Other	3.4	3.1	
On dialysis (%)	5.0	7.5	<.001
Hypertension (%)	26.7	35.4	<.001
Malignancy (%)	0.4	1.8	<.001
Alcohol (%)	9.0	15.2	<.001
Smoking (%)	30.9	38.1	<.001
Drugs (%)	2.0	4.1	<.001
Hepatitis C virus antibody (%)	0.9	0.4	.006
Hepatitis B core antibody (%)	5.8	4.5	.005
Resuscitation (%)	10.1	100.0	-
Hypotension (%)	16.9	-	-
Inotropes (%)	81.0	38.8	<.001
DCD (%)	2.1	0	-
Partial/split liver (%)	4.4	4.9	.16
Age, mean (SD)	47.6 (16.5)	54.0 (18.3)	<.001
Length, mean (SD)	173.5 (9.4)	169.5 (9.8)	<.001
Weight, mean (SD)	75.9 (13.7)	72.7 (15.0)	<.001
BMI, mean (SD)	25.1 (3.7)	25.2 (4.6)	.11
Sodium: latest (mmol/L), median (range)	147 (78-196)	146 (115-180)	.03
Sodium: highest (mmol/L), median (range)	149(121-199)	149 (120-180)	.24
Creatinine: latest (μmol/L), median (range)	92.2 (4.4-849)	76 (20-1000)	<.001
Creatinine: highest (μmol/L), median (range)	103 (2.5-1186)	89 (20-1000)	<.001
ASAT/SGOT: latest (U/L), median (range)	67.5 (1-2684)	39 (0-2000)	<.001
ASAT/SGOT: highest (U/L), median (range)	96 (1-7366)	51 (9-2000)	<.001
ALAT/SGPT: latest (U/L), median (range)	54.5 (1-5300)	27 (0-2000)	<.001
ALAT/SGPT: highest (U/L), median (range)	75.1 (1-13 572)	33 (1-2000)	<.001
Total bilirubin: lastest (μmol/L), median (range)	13 (1.5-102)	10 (0-150)	<.001
Total bilirubin: highest (μmol/L), median (range)	14.3 (1.5-102)	12 (1-150)	<.001

(Continues)

TABLE 2 (Continued)

	EuroTransplant: 2003-2007	French: 2009-2013	*P*-value
Alkaline phosphatase: latest (U/L), median (range)	86.0 (3-6617)	63 (0-400)	<.001
Alkaline phosphatase: highest (U/L), median (range)	92.3 (3-6617)	72 (0.8-400)	<.001
GammaGT: latest (U/L), median (range)	68.4 (1-1970)	30 (0-1477)	<.001
GammaGT: highest (U/L), median (range)	76.8 (1-1970)	36 (1-1835)	<.001
ICU (days), median (range)	4.7 (0.5-72)	2 (0-85)	<.001
Transplantation characteristics			
ABO compatibility (%)	99.97	99.91	.46
Type of allocation (%)			
Normal	77.5	94.1	<.001
Rescue	22.5	5.9	
Allocation (%)			
Local	10.3	40.1	<.001
Regional	35.2	59.9	
Extra-regional	54.5	-	
Cold ischaemia, mean (SD)	9.7 (2.9)	8.4 (4.9)	<.001
DRI, mean (SD)	1.71 (0.42)	1.65 (0.40)	<.001
ET-DRI, mean (SD)	-	1.63 (0.37)	
Recipient characteristics			
Age (%)			
18-39	14.9	11.0	<.001
40-49	24.8	20.7	
50-59	35.8	39.4	
60-69	23.6	28.4	
70	0.9	0.5	
Gender (%)			
Male	64.4	73.5	<.001
Female	35.6	26.5	
Diagnosis (%)			
Primary biliary cirrhosis	4.3	1.6	<.001
Primary sclerosing cholangitis	6.6	2.7	
Biliary atresia	0.2	0.2	
Other cholestatic diagnosis	1.1	0.2	
Autoimmune cirrhosis	2.3	1.9	
Cryptogenic cirrhosis	6.1	1.4	
Post-alcoholic cirrhosis	18.8	29.9	
Hepatitis B cirrhosis	3.5	2.6	
Hepatitis C cirrhosis	9.0	10.6	
Post-hepatitis cirrhosis	8.0	0.1	
Other cirrhosis	5.9	2.3	
Metabolic liver disease	3.6	2.4	
Acute liver failure	8.5	4.9	
Hepatocellular carcinoma/malignant tumours	12.7	28.2	
Vascular liver disease	1.9	0.3	

(Continues)

TABLE 2 (Continued)

	EuroTransplant: 2003-2007	French: 2009-2013	*P*-value
Other/unknown	7.5	10.8	
Highly urgent (%)	15.8	8.6	<.001
MELD (%)			
6-14	14.6	36.9	-
15-24	13.0	26.6	
25	13.6	36.5	
Unknown	58.8	0	
Retransplantation (%)	14.4	10.6	<.001
Age, mean (SD)	51.0 (11.2)	53.1 (10.5)	<.001
MELD, mean (SD)	20.3 (10.0)	20.8 (11.1)	.03
Follow-up, median (years)	2.5	2.5	

coefficients were not different from 0, we hypothesized that this insufficient discrimination was probably because of variables not included in the DRI. For ET-DRI, since one or more of the β* coefficients were different from 0, it meant that there was a lack of fit for the corresponding included covariates.

For both DRIs, the assumption of proportional hazards was not verified for all covariates. This means that the assumption required for the model was not met, thus impairing the interpretation of the model. On one hand, lacking the discrimination measures on the construction set for the DRI, we were unable to draw any conclusions about this result. However, these measures appeared acceptable for survival data. On the other hand, on the ET-DRI construction data set, the C-index was 0.624 (SD=0.008) and the cross-validated C-index was 0.613. Our C-index was very similar (0.622, SD=0.009). Lacking the *K* statistic and R^2_D we could not draw any conclusions for those two measures except that they appeared reasonable.

For the DRI and ET-DRI groups, Kaplan-Meier curves were not available from the construction data sets. However, survival at month 3, year 1 and year 3 were available, grouped according to DRI (or ET-DRI) categories. Again, there was a lack of discrimination. Indeed, there was no survival decrease according to the different DRI groupings or ET-DRI groupings (Table 3). Moreover, there was a lack of apparent calibration: survival, according to DRI (or ET-DRI) categories, was not consistent in both datasets (construction and validation datasets). Finally, HRs for the various DRI groups (or ET-DRI) confirmed the results obtained in the previous steps. The survival among the different DRI (or ET-DRI) groups was not different; moreover, the risk did not increase as expected when comparing group 0 with the other groups. There was no significant difference of mortality/graft loss between groups.

Since the distributions of the DRI and the ET-DRI both differed between the construction and validation datasets, we performed a direct standardization on the joint populations to remove the DRI (or the ET-DRI, respectively) effect. Interestingly, after standardization, survival analyses at month 3, year 1 and year 3 showed a better survival for the French dataset in patients grafted with donors with high DRI (or high ET-DRI, respectively) (See Tables S1 and S2).

The calibration and baseline hazard function method[7] was not applied since baseline survival was not available in the construction files.

The Eurotransplant study had validated the DRI on its own database,[4] however, the model used in the validation dataset was not based on the same covariates as in the construction dataset, contrary to what is recommended.[7] Indeed, the adjustment variables used for validation differed from those used in the construction set (extra- and missing variables). In addition, the ET-DRI score was built by changing

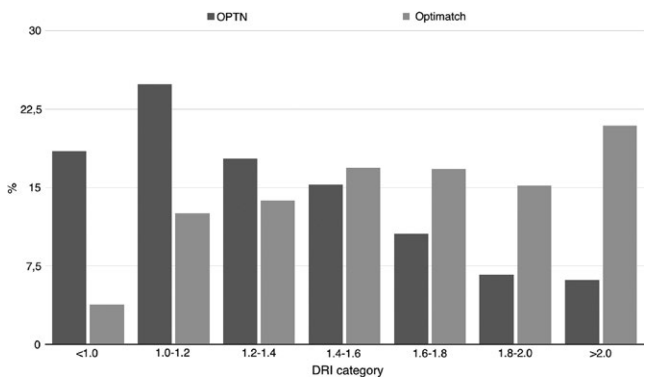

FIGURE 1 Distribution of the DRI for the OPTN and the French datasets

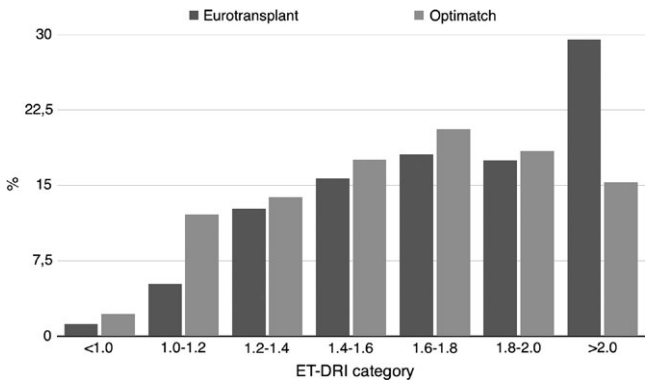

FIGURE 2 Distribution of the ET-DRI for the Eurotransplant and the French datasets

TABLE 3 Three-month, 1-year and 3-year failure free survival per DRI and ET-DRI category on the French dataset

			Graft survival (95% confidence interval)		
Group	**DRI**	**%**	**3 months**	**1 year**	**3 years**
0	0.0<DRI≤1.0	3.8	91.5 (87.0-96.2)	85.8 (80.2-91.8)	78.3 (71.3-85.9)
1	1.0<DRI≤1.1	6.3	90.4 (86.7-94.3)	85.6 (81.1-90.2)	80.1 (75.0-85.6)
2	1.1<DRI≤1.2	6.3	88.3 (84.3-92.6)	84.0 (79.4-88.8)	76.8 (71.3-82.7)
3	1.2<DRI≤1.3	6.6	91.7 (88.3-95.3)	87.6 (83.5-91.9)	77.0 (71.4-83.0)
4	1.3<DRI≤1.4	7.2	91.3 (88.0-94.8)	82.6 (78.2-87.3)	75.0 (69.6-80.8)
5	1.4<DRI≤1.5	8.1	90.2 (86.9-93.7)	84.5 (80.4-88.7)	77.1 (72.2-82.4)
6	1.5<DRI≤1.6	8.8	92.6 (89.8-95.5)	88.3 (84.9-91.9)	80.4 (76.0-85.2)
7	1.6<DRI≤1.8	16.8	90.9 (88.7-93.2)	85.8 (83.0-88.6)	78.4 (75.1-81.9)
8	1.8<DRI≤2.0	15.2	86.4 (83.6-89.3)	80.6 (77.4-84.0)	70.9 (66.9-75.1)
9	2.0<DRI	20.9	88.8 (86.6-91.1)	81.2 (78.4-84.0)	74.3 (71.1-77.6)
			Graft survival (95% confidence interval)		
Group	**ET-DRI**	**%**	**3 months**	**1 year**	**3 years**
0	0.0<ET-DRI≤1.0	2.4	89.1 (83.5-95.1)	82.7 (76.0-90.1)	78.0 (70.3-86.5)
1	1.0<ET-DRI≤1.2	12	89.9 (87.4-92.5)	84.6 (81.6-87.6)	77.1 (73.5-80.9)
2	1.2<ET-DRI≤1.4	14	90.7 (88.5-93.0)	84.3 (81.5-87.2)	75.8 (72.3-79.4)
3	1.4<ET-DRI≤1.6	17.5	88.8 (86.7-91.1)	83.8 (81.3-86.4)	76.2 (73.1-79.3)
4	1.6<ET-DRI≤1.8	20.8	87.4 (85.3-89.5)	81.2 (78.8-83.8)	74.7 (71.8-77.6)
5	1.8<ET-DRI≤2.0	18.2	87.8 (85.6-90.1)	80.5 (77.8-83.2)	72.3 (69.2-75.6)
6	2.0 <ET-DRI	15.1	87.4 (84.9-89.9)	79.9 (77.0-83.0)	71.3 (67.8-75.0)

the coefficients of the DRI and adding variables, which is not recommended.[7] When adjustment is needed, if the original data is available, it is preferable to refit the original model, but that is not the aim of an external validation procedure per se.

Reichert et al.[15] did not validate the ET-DRI on their own database. However, the number of patients was limited (n=291 LTs) for a validation study, limiting its power as a validation tool.

There are several possibilities for why the external validation of the scores did not work on our dataset. First, several differences were identified between the validation database and construction datasets, especially in terms of donor age which is one of the strongest and most variable graft failure risk factors. Second, for the DRI, some variables were not registered in our dataset, such as donor race, and for recipient: race, angina, cerebrovascular disease, transfusion at time of listing, serum glutamic oxaloacetic transaminase, tumour identified during LT, on life support, inotropic support, and portal vein thrombosis. Third, we had to create new variables such as local/regional/national sharing. Thus, it appears that the model had not been applied exactly in the same way as in refs (1) and (3) as recommended. Four, the number of LTs used for OPTN was 20 023 against 3681 LT in our database, given greater statistical power to the OPTN dataset. Finally, our dataset shows a large difference in distribution of the DRI (Figure 1); most donors have low scores in the OPTN data set, whereas we observed higher scores in ours.

For the ET-DRI, two variables were not registered in our database: the organ perfusion solution and the occurrence of a hypotensive period. We also had to create new variables such as local/regional/extra-regional sharing and a "rescue allocation" variable that was not totally identical to the one used by Eurotransplant. Conclusions identical to the DRI were then drawn. Finally, we noticed a large proportion of recipients with missing MELD in the Eurotransplant database.

The periods of the studies were not identical: for the OPTN database the inclusion period was between 1998 and 2002, for the Eurotransplant database it was between 2003 and 2007, and for the French database between 2009 and 2013. This covariate impacts on the DRIs as the number of recipients has increased in recent years without a corresponding increase in the number of donors, resulting in an increase in the use of poorer grafts with higher DRI (Figures 1 and 2). It should, however, be noted, that there is no consensus in the literature about selecting the number of groups at risk or in positioning the cut-off points to delineate these groups.[16] Too many groups could be unstable and consequently discrimination becomes insufficient. The recommendation is to create 3-5 groups, not necessarily of the same size, in order to highlight extreme groups. That was not the case for either the DRI or for the ET-DRI, which instead formed 10 and 7 groups respectively. Moreover, we did not observe the expected matching,[6] for instance between a high DRI with either a low MELD or a hepatocellular carcinoma (HCC). For example, for an ET-DRI >2, the percentage of HCC was 38%, vs 32% for an ET-DRI lower than 2 (See Tables S3 and S4).

It could be considered that data derived from the French data would not be applicable to transplant recipients outside of France.

TABLE 4 Hazard ratios between the DRI groups and between the ET-DRI groups on the French dataset

	DRI HR	Confidence interval
HR: group 0 vs 1	0.87	0.54-1.41
HR: group 0 vs 2	0.98	0.61-1.57
HR: group 0 vs 3	0.93	0.58-1.49
HR: group 0 vs 4	1.05	0.66-1.67
HR: group 0 vs 5	1.18	0.74-1.88
HR: group 0 vs 6	1.23	0.77-1.96
	ET-DRI HR	**Confidence interval**
HR: group 0 vs 1	0.89	0.57-1.39
HR: group 0 vs 2	1.02	0.66-1.57
HR: group 0 vs 3	1.02	0.67-1.58
HR: group 0 vs 4	1.18	0.78-1.78
HR: group 0 vs 5	1.02	0.67-1.54
HR: group 0 vs 6	0.91	0.60-1.38
HR: group 0 vs 7	0.99	0.68-1.44
HR: group 0 vs 8	1.34	0.92-1.95
HR: group 0 vs 9	1.29	0.90-1.87

However, validation against such a dataset is obligatory, as validation of a prognostic score requires both an internal and external validation.[14] Afterwards, an impact analysis is recommended if the first two steps are validated. This three–step procedure is necessary to reach a further step in order to study the appropriate matching between a graft and its recipient. A DRI, if validated, is a support for a better matching. In this perspective, our approach is not a model restricted to French patients but serves a broader purpose to define a general framework for ultimately improving the matching of the graft to its host.

Re-calibrating a model must be considered.[8,17] However, the discrimination cannot be altered.[7] Then the current scores cannot be used on our dataset. As mentioned previously, the period of study has an impact on the DRIs, so it is essential that a DRI evolves over time through the revaluation of coefficients, including addition or removal of some covariates, in order to accurately define the quality of a graft. Thus, our next step will be to propose a general adaptive model for a DRI.

ACKNOWLEDGEMENTS

We warmly thank Mrs Sarah Kabani for her constructive and sensitive editorial assistance. We thank all the professionals who contributed to gather the information collected in the LT Cristal database.

CONFLICT OF INTEREST

The authors do not have any disclosures to report.

REFERENCES

1. Braat A, Blok J, Putter H, et al. The Eurotransplant donor risk index in liver transplantation: ET-DRI. *Am J Transplant*. 2012;12:2789–2796.
2. Tector AJ, Chestovich P, Vianna R, et al. Use of extended criteria livers decreases wait time for liver transplantation without adversely impacting posttransplant survival. *Ann Surg*. 2006;244:439–450.
3. Feng S, Goodrich N, Bragg-Gresham J, et al. Characteristics associated with liver graft failure: the concept of a donor risk index. *Am J Transplant*. 2006;6:783–790.
4. Blok JJ, Braat AE, Adam R, et al. Validation of the donor risk index in orthotopic liver transplantation within the Eurotransplant region. *Liver Transpl*. 2012;18:112–119.
5. Duvoux C, Roudot-Thoraval F, Decaens T, et al. Liver transplantation for hepatocellular carcinoma: a model including α-fetoprotein improves the performance of Milan criteria. *Gastroenterology*. 2012;143:986–994.
6. Schaubel D, Sima C, Goodrich N, Feng S, Merion R. The survival benefit of deceased donor liver transplantation as a function of candidate disease severity and donor quality. *Am J Transplant*. 2008;8:419–425.
7. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;13:33.
8. Van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19:3401–3415.
9. Miller ME, Hui SL, Tierney WM. Validation techniques for logistic regression models. *Stat Med*. 1991;10:1213–1226.
10. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543–2546.
11. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005;92:965–970.
12. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med*. 2004;23:723–748.
13. R Development Core Team *A Language and Environment for Statistical Computing*. Vienna: Development Core Team; 2016.
14. 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–W73.
15. Reichert B, Kaltenborn A, Goldis A, Schrem H. Prognostic limitations of the Eurotransplant-donor risk index in liver transplantation. *J Negat Results Biomed*. 2013;12:1.
16. Altman D. Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest*. 2009;27:235.
17. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Winter A, Féray C, Audureau E, et al. External validation of the Donor Risk Index and the eurotransplant Donor Risk Index on the french liver transplantation registry. *Liver Int*. 2017;37:1229-1238.