

# The Model for End-stage Liver Disease accurately predicts 90-day liver transplant wait-list mortality in Atlantic Canada

Paul Douglas Renfrew MD MSc FRCSC<sup>1,2</sup>, Hude Quan MD MPH PhD<sup>2</sup>, Christopher James Doig MD MSc FRCPC<sup>2</sup>,  
Elijah Dixon MD MSc FRCSC<sup>2</sup>, Michele Molinari MD<sup>1</sup>

PD Renfrew, H Quan, CJ Doig, E Dixon, M Molinari. The Model for End-stage Liver Disease accurately predicts 90-day liver transplant wait-list mortality in Atlantic Canada. *Can J Gastroenterol* 2011;25(7):359-364.

**OBJECTIVE:** To determine the generalizability of the predictions for 90-day mortality generated by Model for End-stage Liver Disease (MELD) and the serum sodium augmented MELD (MELDNa) to Atlantic Canadian adults with end-stage liver disease awaiting liver transplantation (LT).

**METHODS:** The predictive accuracy of the MELD and the MELDNa was evaluated by measurement of the discrimination and calibration of the respective models' estimates for the occurrence of 90-day mortality in a consecutive cohort of LT candidates accrued over a five-year period. Accuracy of discrimination was measured by the area under the ROC curves. Calibration accuracy was evaluated by comparing the observed and model-estimated incidences of 90-day wait-list failure for the total cohort and within quantiles of risk.

**RESULTS:** The area under the ROC curve for the MELD was 0.887 (95% CI 0.705 to 0.978) – consistent with very good accuracy of discrimination. The area under the ROC curve for the MELDNa was 0.848 (95% CI 0.681 to 0.965). The observed incidence of 90-day wait-list mortality in the validation cohort was 7.9%, which was not significantly different from the MELD estimate of 6.6% (95% CI 4.9% to 8.4%;  $P=0.177$ ) or the MELDNa estimate of 5.8% (95% CI 3.5% to 8.0%;  $P=0.065$ ). Global goodness-of-fit testing found no evidence of significant lack of fit for either model (Hosmer-Lemeshow  $\chi^2$  [df=3] for MELD 2.941,  $P=0.401$ ; for MELDNa 2.895,  $P=0.414$ ).

**CONCLUSION:** Both the MELD and the MELDNa accurately predicted the occurrence of 90-day wait-list mortality in the study cohort and, therefore, are generalizable to Atlantic Canadians with end-stage liver disease awaiting LT.

**Key Words:** End-stage liver disease; Liver transplantation; Mortality; Statistical models; Validation study; Wait list

The Model for End-stage Liver Disease (MELD) is a prognostic model for end-stage liver disease (ESLD) that generates a continuous disease severity score reflective of an individual's risk of ESLD-related mortality based on the status of three objective laboratory variables: the international normalized ratio for prothrombin time (INR), serum creatinine concentration and serum total bilirubin concentration (1). The MELD score has been shown to accurately predict the risk of 90-day mortality in patients awaiting liver transplantation (LT) in the United States, where it has subsequently been used to prioritize deceased donor liver allograft allocation since 2002 (2). At present in Canada, there is no centrally administered national organ allocation system, and there is ongoing discussion as to whether a MELD-based

**Le modèle de maladie hépatique en phase terminale prédit avec précision la mortalité au bout de 90 jours des personnes sur la liste d'attente d'une transplantation hépatique dans la région atlantique du Canada**

**OBJECTIF :** Déterminer la généralisabilité des prédictions de mortalité dans les 90 jours au moyen du modèle de maladie hépatique en phase terminale (MMHT) et du MMHT augmenté par le sérum sodique (MMHTNa) chez les Canadiens de l'Atlantique atteints d'une maladie hépatique en phase terminale en attente d'une transplantation hépatique (TH).

**MÉTHODOLOGIE :** Les auteurs ont évalué l'exactitude prédictive du MMHT et du MMHTNa en mesurant la discrimination et la calibration des évaluations de chaque modèle à l'égard de l'occurrence de mortalité au bout de 90 jours dans une cohorte consécutive de candidats à la TH compilée sur une période de cinq ans. Ils ont mesuré l'exactitude de la discrimination par la zone sous les courbes ROC et l'exactitude de la calibration par la comparaison des incidences observées et évaluées par modèles des échecs de la liste d'attente au bout de 90 jours dans l'ensemble de la cohorte et au sein de quantiles de risque.

**RÉSULTATS :** La zone sous la courbe ROC du MMHT correspondait à 0,887 (95 % IC 0,705 à 0,978), ce qui est indicateur d'une très bonne précision de la discrimination. La zone sous la courbe ROC du MMHTNa correspondait à 0,848 (95 % IC 0,681 à 0,965). L'incidence observée de la mortalité des personnes sur la liste d'attente au bout de 90 jours s'élevait à 7,9 %, ce qui n'est pas significativement différent de l'évaluation du MMHT de 6,6 % (95 % IC 4,9 % à 8,4 %;  $P=0,177$ ) ou de l'évaluation du MMHTNa de 5,8 % (95 % IC 3,5 % à 8,0 %;  $P=0,065$ ). Le test de validité de l'ajustement global n'a révélé aucune donnée probante significative de non-ajustement dans l'un ou l'autre modèle ( $\chi^2$  [de Hosmer-Lemeshow trois degrés de liberté] de 2,941,  $P=0,401$  pour le MMHT, et 2,895,  $P=0,414$  pour le MMHTNa).

**CONCLUSION :** Tant le MMHT que le MMHTNa prédisent avec précision l'occurrence de mortalité au bout de 90 jours des personnes sur la liste d'attente de la présente cohorte. Par conséquent, ces modèles peuvent être généralisés aux Canadiens de l'Atlantique atteints d'une maladie hépatique en phase terminale qui sont en attente d'une TH.

allocation system for LT should be adopted. For this deliberation to be evidence based, it is important that the predictive performance of the MELD be characterized for the Canadian LT wait-list population. In the absence of a national wait-list database to facilitate a national validation study, external regional validation studies have an important role to play in the evaluation of the MELD. The sole study to date, which was published in abstract form only (3,4), evaluated the discriminative performance of the MELD in a cohort of western Canadians. The primary objective of the present study was to expand the evidence base by comprehensively assessing the generalizability of the MELD prognostic model to the population of adults with ESLD from Atlantic Canada awaiting LT. Additionally, the current study evaluated the predictive

<sup>1</sup>Atlantic Multi-Organ Transplant Program, Department of Surgery, Dalhousie University, Halifax, Nova Scotia; <sup>2</sup>Department of Community Health Sciences, University of Calgary, Calgary, Alberta

Correspondence: Dr Paul Douglas Renfrew, Room 6-203, Victoria Building, Queen Elizabeth II Health Sciences Centre, 1276 South Park Street, Halifax, Nova Scotia B3H 2Y9. Telephone 902-473-5296, fax 902-473-5297, e-mail p.d.renfrew@dal.ca

Received for publication November 25, 2010. Accepted December 20, 2010

accuracy of the serum sodium augmented form of the MELD – the MELDNa – and compared its performance to that of the MELD (5).

## METHODS

### Ethical considerations

Before initiation of the present study, ethics approval of the protocol was received from the Research Ethics Boards of the Capital District Health Authority, Halifax, Nova Scotia (CDHA-RS/2009-075) and the University of Calgary, Calgary, Alberta (E-23065).

### Study setting and sample

The study sample was identified from a prospective institutional database that contains demographic and clinical information on all individuals referred to the Atlantic Multi-Organ Transplant Program for consideration of LT. This program provides LT services for adults who reside in the four provinces of Atlantic Canada: Nova Scotia, New Brunswick, Newfoundland and Labrador, and Prince Edward Island. This catchment region has a population of approximately 2.3 million (6). The validation cohort used for the present analysis consisted of all individuals 18 years of age or older whose primary indication for LT was decompensated ESLD, and who were actively assigned to the wait list for first LT between December 1, 2004, and December 1, 2009. Individuals whose primary indication for LT was acute liver failure, malignancy or other non-ESLD indications were excluded from the validation cohort, as were individuals who had undergone at least one previous LT.

### Independent and calculated variables

The laboratory investigation results necessary for the calculation of the MELD and MELDNa risk scores, and their date of assay were extracted by retrospective review of the candidate's institutional medical record. The set of investigations that were obtained temporally nearest to the date of active wait-list assignment was the one selected to calculate the risk scores. Serum creatinine and total bilirubin concentrations (in  $\mu\text{mol/L}$ ) were divided by 88.4 and 17.1, respectively, to convert the molar concentrations to the mg/dL units required for the calculation of the MELD risk score. The MELD score was calculated using the formula defined by Kamath et al (7) and the constraints applied by United Network for Organ Sharing (8,9). The formula is as follows:

$$(11.20 \times \ln[\text{INR}]) + (9.57 \times \ln[\text{creatinine}]) + (3.78 \times \ln[\text{total bilirubin}]) + 6.43,$$

with the lower limit for the three independent variables set at 1.0 and maximum serum creatinine concentration capped at 4.0 mg/dL for observations with values exceeding that amount or those receiving hemodialysis. The resultant MELD risk score was rounded to the nearest integer and capped at 40. The MELDNa risk score was calculated using the following formula:

$$\text{MELD score} + (140 - \text{serum Na}) - (0.025 \times \text{MELD score} \times (140 - \text{serum Na})),$$

with the serum sodium concentration bounded between 125 mmol/L and 140 mmol/L (5). The probability estimates for 90-day mortality corresponding to each observation's MELD and MELDNa risk scores were calculated using survival functions kindly provided by investigators at the Mayo Clinic College of Medicine in Rochester, Minnesota (USA) (Joanne Benson, Division of Biomedical Statistics and Informatics, personal communication, April 28, 2009).

### Follow-up and outcomes

Wait-list observation began on the date of active assignment to the LT wait list. Wait-list outcome was followed-up until March 1, 2010, to permit a minimum 90-day period of wait-list observation for all patients in the cohort. Wait-list mortality was the failure outcome used in the present study, and was defined as death of the candidate while on the wait list or withdrawal of the candidate from the wait list because they were deemed to have become too ill to survive LT (ie, terminally ill) due to progression of, or complications related to, their

ESLD. These failure outcomes were recorded to have occurred, respectively, on the date of death or withdrawal of the candidate from the wait list. Individuals that were removed from the wait list due to performance of LT; improvement in condition to the point that LT was no longer deemed to be necessary; revocation of candidacy due to detection of non-ESLD-related medical comorbidities or psychosocial conditions that rendered them unsuitable for LT; or transfer to the wait list of another transplant program, were defined to have nonfailure outcomes and had their observation time censored on the date of wait-list withdrawal. Candidates who were still awaiting LT on March 1, 2010, had their wait-list follow-up censored at that time.

Observations with noncensored 90-day wait-list outcomes were used to evaluate the predictive performance in the MELD and MELDNa models. Individuals transplanted within 90 days of wait-list registration have the natural history of their ESLD interrupted by LT and become no longer at risk for ESLD-related mortality; therefore, these individuals were excluded from the validation analysis. Additionally, candidates withdrawn from the wait list within 90 days for other reasons (ie, improvement and revocation) were also excluded from the validation analysis because this group of patients was not systematically followed after withdrawal and, therefore, their outcomes were not consistently known.

### Statistical methods

All statistical analyses were performed using Stata/IC 11 (StataCorp LP, USA). To supplement ROC curve analytic capabilities, package st0154 was installed from [http://www.stata-journal.com/software/sj9-1\(10\)](http://www.stata-journal.com/software/sj9-1(10)).

For continuous variables, unless otherwise specified, the median was used as the measure of central location, and the range and/or interquartile range (IQR) as the corresponding measure of variation. Categorical variables were expressed as proportions. The Kaplan-Meier product limit estimator of the survival function was used to describe censored time-to-event data (11).

For comparisons between continuous variables, the nonparametric Wilcoxon rank-sum test was used. An exception to this was in the comparison between the mean model estimated probability of 90-day wait-list mortality and the observed incidence of 90-day wait-list mortality, for which a one-sample *t* test was used. For comparisons between proportions, the Pearson's  $\chi^2$  test was used unless a cell contained less than five observations, in which case the Fisher's exact test was used. All tests of hypothesis were two sided, and a less than 5% probability of type I error was the threshold for significance.

The validation analysis assessed both the discrimination and calibration accuracy of the MELD models. Because the primary function of the MELD, as applied in a liver allograft allocation system, is to rank candidates according to their medical need based on the severity of their ESLD and risk of ESLD-related mortality, the principle performance measure of relevance in the present analysis was discrimination accuracy. This was measured by the area under the nonparametric ROC curve. This assumption-free method for the estimation of the area under the ROC curve is acknowledged to be robust and generalizable when the classification variable has five or more potential values (12-14). In light of the small sample size and low event occurrence, bootstrap CIs (bias-corrected and accelerated) for the area under the ROC curve were calculated using 1000 replications. Compared with nonparametric CIs, the bootstrap methodology has been found by simulation study to produce more appropriate CI estimates when outcome groups are smaller than 30 (15). The area under the ROC curve can range from 0.0 to 1.0. An area of 1.0 indicates perfect discrimination (model separates observations into appropriate outcome group 100% of the time), while an area of 0.5 indicates the instrument possesses no intrinsic discriminative ability (ie, the model separates observations into appropriate outcome group 50% of time – equivalent to random chance). Areas less than 0.5 indicate sorting into outcome groups opposite of what is expected, implying that outcomes have been incorrectly coded. Discrimination can be empirically qualified as

**TABLE 1**  
**Characteristics of adult candidates whose primary indication for first liver transplant was decompensated end-stage liver disease (ESLD) (n=121)**

Variable	
Sex, n (%)	
Male	64 (52.9)
Female	57 (47.1)
Age at wait-list assignment, years, median (range)	55.0 (21.6–68.9)
Primary liver disease, n (%)	
HCV-related ESLD	26 (21.5)
Alcohol-related ESLD	24 (19.8)
Primary biliary cirrhosis	21 (17.4)
Primary sclerosing cholangitis	17 (14.1)
Cryptogenic cirrhosis	11 (9.1)
Autoimmune hepatitis	9 (7.4)
Nonalcoholic steatohepatitis	7 (5.8)
Others (alpha-1 antitrypsin deficiency, congenital hepatic fibrosis, cystic fibrosis, HBV-related ESLD, sarcoidosis, Wilson's disease)	1 each
Wait-list laboratory data	
International normalized ratio for prothrombin time, median (range)	1.3 (1.0–4.2)
Serum creatinine concentration, mg/dL, median (range)	1.0 (0.2–2.9)
Serum total bilirubin concentration, mg/dL, median (range)	2.3 (0.4–40.9)
Serum sodium concentration, mmol/L, median (range)	136 (116–148)
Serum sodium concentration <135 mmol/L, n (%)	43 (35.5)
Wait-list registration risk scores, median (IQR; range)	
MELD	13 (10–18; 6–33)
MELDNa	17 (12–21; 8–34)

HBV Hepatitis B virus; HCV Hepatitis C virus; IQR Interquartile range; MELD Model for ESLD; MELDNa Serum sodium augmented form of the MELD

the following: 'unacceptable' for areas under the ROC curve from 0.50 to 0.69; 'acceptable' for areas of 0.70 to 0.79; and 'good to excellent' for models generating predictions associated with areas under the ROC curve of 0.80 or greater (16). Comparison of the discriminatory performance of the MELD and MELDNa was conducted by evaluating the difference between the areas under nonparametric ROC curves for the respective models using the nonparametric method, which has been shown to be acceptable even when the sample size is small (13,15).

The calibration accuracy of the MELD and MELDNa predictions for 90-day mortality were evaluated using three methods. The observed to model-estimated incidence of wait-list failure for the entire cohort was compared via a one-sample Student's *t* test, with the null hypothesis being that the mean estimate equalled the observed incidence of 90-day mortality in the validation cohort. The Hosmer-Lemeshow goodness-of-fit test was used to assess global calibration of the model. Typically, this methodology entails subdivision of the study sample into deciles; however, in light of the size of the study cohort in the present analysis, the sample was divided into three risk score-based quantiles. The Hosmer-Lemeshow test statistic is distributed as  $\chi^2$ , and when applied for the external validation of a (fixed) model, the degrees of freedom are equivalent to the number of subgroups (16). Quantile-of-risk calibration curves were also constructed to allow subjective evaluation of calibration performance across the range of observed risk scores. Methods for the comparison of calibration accuracy between two models are, in general, sparse and, in the current analysis, were limited to comparison of the  $\chi^2$  statistics and related *P* values from the Hosmer-Lemeshow goodness-of-fit tests and subjective comparison of calibration curves.

**TABLE 2**  
**Overall wait-list outcomes and duration of wait time for adult candidates whose primary indication for first liver transplant was decompensated end-stage liver disease (n=121)**

Outcome	Frequency, n (%)	Time to event, days		
		Median	IQR	Range
Transplanted	76 (62.8)	133	57.5–292	1–1012
Withdrawn – unnecessary	13 (10.7)	554	255–698	122–1510
Withdrawn – unsuitable	10 (8.3)	301.5	86–515	29–1112
Transferred to another program	2 (1.7)		144 and 235	
Still awaiting transplantation	8 (6.6)	777	532–1200	382–1839
Died	5 (4.1)	34	27–111	24–128
Withdrawn – terminal	7 (5.8)	63	45–236	13–577
Total	121 (100.0)	157	63–398	1–1839

IQR Interquartile range

To evaluate the effect that might be generated by the inclusion of the influence of low serum sodium in the prediction of 90-day wait-list mortality, a risk score reclassification table was constructed comparing MELD- and MELDNa-derived risk scores. This method of analysis is contingent on the existence of clinically meaningful risk thresholds. In the case of the MELD, under the Share MELD 15 policy, a risk score of 15 or greater has the significance that it avails the candidate to eligibility for allografts recovered from a broader geographical area. Therefore, reclassification was analyzed using this cut-off point, with the net effect summarized by calculation of the net reclassification improvement (NRI) and its related test of hypothesis as described by Pencina et al (17). Also described by Pencina et al as a measure of relative change in discrimination performance is the integrated discrimination improvement (IDI), which does not require the existence of risk thresholds. Consequently, the IDI was also calculated to further compare the performance of the MELDNa relative to the MELD.

## RESULTS

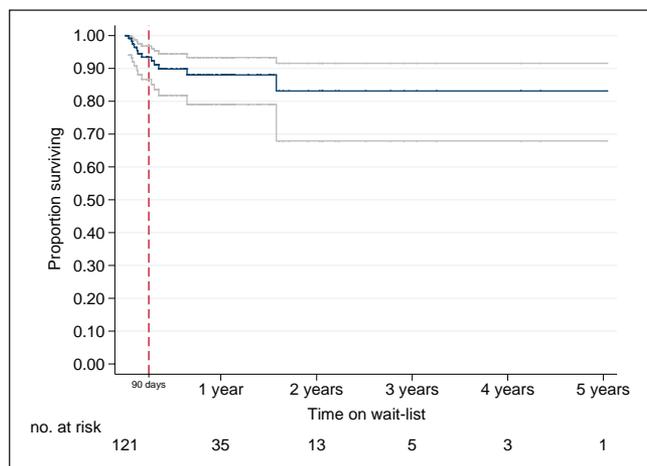
Between December 1, 2004, and December 1, 2009, 159 individuals 18 years of age or older were assigned to the wait list for primary LT. The principle indications for LT were as follows: decompensated ESLD for 121 candidates (76.1%); malignancy in 29 (18.2%); acute liver failure in five; and polycystic liver disease in four. The characteristics of the 121 candidates who met the study cohort inclusion criteria are summarized in Table 1.

The laboratory investigational data necessary for calculation of the MELD and MELDNa risk scores were available for all 121 candidates. The median time elapsed between the date of these investigations and the date of wait-list registration was two days before (range 57 days before to 35 days after). Thirty-eight per cent of candidates underwent laboratory investigations within seven days of wait-list registration, and 91.7% were obtained within 30 days. The median MELD score of the 121 candidates was 13 (IQR 10 to 18, range 6 to 33), and the median MELDNa score was 17 (IQR 12 to 21, range 8 to 34).

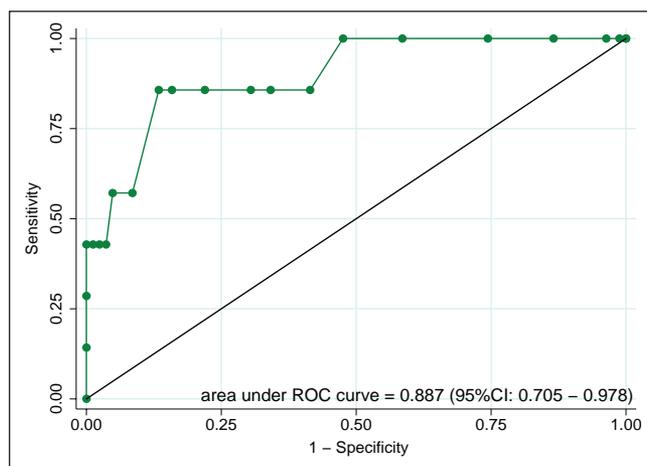
### Wait-list outcomes

The overall wait-list outcomes and duration of wait time for the cohort of 121 candidates are summarized in Table 2. The incidence rate for wait-list failure in the cohort was 12.3/100 person-years at risk (95% CI 7.0 to 21.6). The Kaplan-Meier estimates for 90-day and one-year mortality on the wait list were 6.6% (95% CI 3.2% to 13.3%) and 12.0% (95% CI 6.7% to 21.0%), respectively (Figure 1).

From the cohort of 121, a total of 32 candidates had their wait-list follow-up censored before 90 days of observation. Twenty-nine of these candidates ceased to be at risk for ESLD-related mortality when



**Figure 1)** Kaplan-Meier estimate of survival on the liver transplant wait list for the cohort of 121 individuals whose primary indication was decompensated end-stage liver disease (with 95% CI). no Number



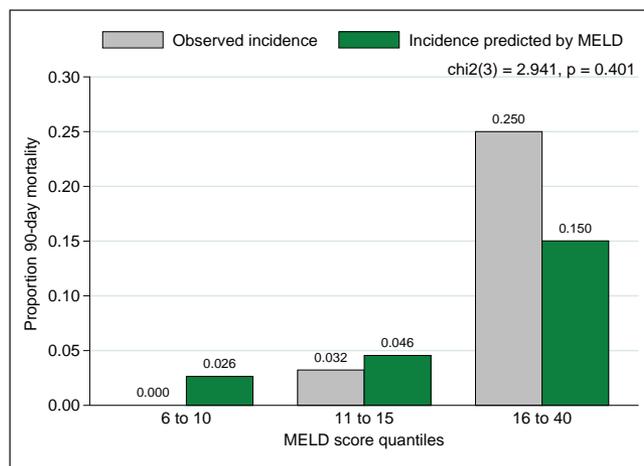
**Figure 2)** Nonparametric ROC curve for the Model for End-stage Liver Disease prediction of the occurrence of 90-day wait-list mortality for the 89 candidates with uncensored 90-day wait-list outcomes

they underwent LT, and three had their candidacy revoked due to detection of medical comorbidities in two and psychosocial issues in one, which rendered them unsuitable for LT. At the end of 90 days observation, 82 candidates were still awaiting LT, and a total of seven had either died ( $n=3$ ) or were delisted ( $n=4$ ) because they became too ill to undergo LT. The data from this group of 89 individuals with uncensored 90-day wait-list follow-up were used to evaluate the predictive accuracy of the MELD models.

The MELD and MELDNa scores of the seven individuals who experienced 90-day wait-list failure (median 20, range 12 to 33; and median 22, range 15 to 34, respectively) were significantly greater than the MELD and MELDNa scores of the 82 survivors (median 11, range 6 to 25; and median 14, range 8 to 28;  $P<0.001$  and  $P=0.001$ , respectively). The serum sodium concentrations for the seven wait-list mortalities (median 136 mmol/L, range 133 mmol/L to 148 mmol/L) were not found to be different from those of the 90-day wait-list survivors (median 137 mmol/L, range 123 mmol/L to 148 mmol/L) ( $P=0.860$ ), nor was the prevalence of hyponatremia (serum sodium concentration lower than 135 mmol/L) at the time of wait-list assignment (28.1% versus 28.6%;  $P=1.000$ ).

#### Predictive accuracy of the MELD

Figure 2 displays the nonparametric ROC curve for the MELD prediction of the occurrence of 90-day wait-list mortality. The estimate of the area under the ROC curve and its bootstrap CI was 0.887 (95% CI



**Figure 3)** Observed versus Model for End-stage Liver Disease (MELD)-predicted incidence of 90-day mortality for the 89 candidates with uncensored 90-day outcomes according to three risk-score quantiles

0.705 to 0.978). This value for the area under the ROC curve reflected very good discrimination accuracy.

The MELD-generated estimate for 90-day wait-list mortality in the group of 89 candidates was 6.6% (95% CI 4.9% to 8.4%), which was not found to be significantly different from the observed incidence of 7.9% in this group ( $P=0.177$ ). The test statistic from the Hosmer-Lemeshow goodness-of-fit test was 2.941, which on a  $\chi^2$  distribution with three degrees of freedom, corresponded to a P value of 0.401. While this result implies no evidence of significant lack of fit, the validity of this result must be interpreted with caution because it is generally accepted that cells must have a minimum of five observations for the distributional assumptions upon which the Hosmer-Lemeshow statistic is based to be met (16). Figure 3 graphically displays the results of the calibration analysis using the three score-based quantile grouping method.

#### Predictive accuracy of the MELDNa

The area under the ROC curve for the MELDNa's prediction for the occurrence of 90-day wait-list failure was 0.848 (95% CI 0.681 to 0.965) (Figure 4). This value for the area under the ROC curve implies good discrimination accuracy.

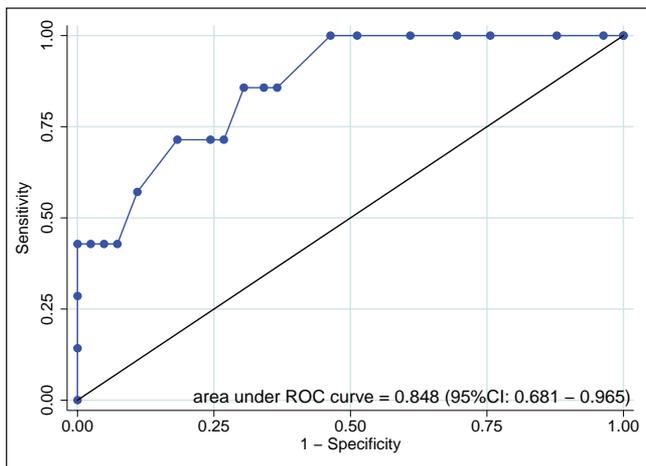
Based on the MELDNa-derived probability estimates, the predicted incidence of 90-day mortality in the cohort of 89 candidates was 5.8% (95% CI 3.5% to 8.0%). This estimate did not significantly differ from the observed incidence of 7.9% ( $P=0.065$ ). Figure 5 graphically displays the results of the calibration analysis using three risk-score quantiles. The Hosmer-Lemeshow global goodness-of-fit test statistic was 2.895, which on a  $\chi^2$  distribution with three degrees of freedom corresponded to a P value of 0.414. This result suggested no significant lack of fit, although, as stated in the preceding section, the validity of this finding can be questioned due to the low frequency of event occurrence.

#### Comparison between the predictive accuracy of the MELD and MELDNa

Nonparametric comparison between the areas under the ROC curve for the MELD and MELDNa predictions for the occurrence of 90-day mortality produced no evidence to indicate that the absolute difference of 0.038 was statistically significant ( $P=0.294$ ).

The  $\chi^2$  statistics and related P values from the Hosmer-Lemeshow tests of global goodness of fit were very similar, suggesting that the fit of the two models did not differ.

The analysis of reclassification using a risk score of 15 as a clinically relevant cut-off point is presented in Table 3. This analysis demonstrated that, if the MELDNa was used to derive their risk

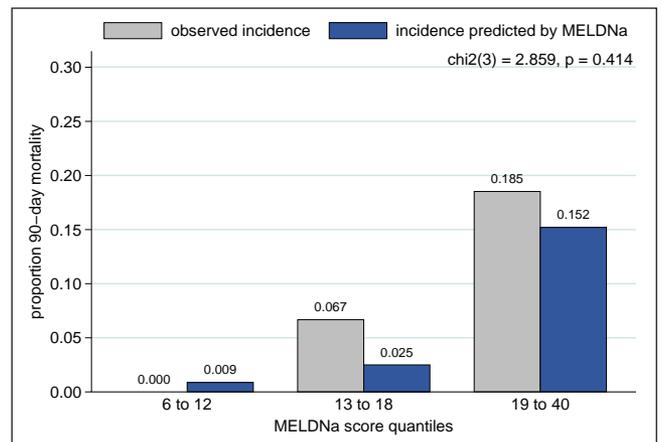


**Figure 4)** Nonparametric ROC curve for the sodium augmented form of the Model for End-stage Liver Disease prediction of the occurrence of 90-day wait-list mortality for the 89 candidates with uncensored 90-day wait-list outcomes

score, 14 candidates with MELD scores of 14 or less would have been reassigned a risk score of 15 or greater, thus affording them greater access to an allograft under the Share MELD 15 policy. Of these 14 reclassified, one was from the group of seven 90-day wait-list failures. This means that the MELDNa would have favourably reclassified 14.3% of the individuals who failed. At the same time, 13 of the 89 candidates who did not experience 90-day failure (14.6%) were unnecessarily, or unfavourably, reassigned to the higher risk-score group. The NRI was therefore  $-0.003$ , which implies a net 0.3% harmful reclassification when the MELDNa was used to assign the risk score, as opposed to the MELD – a difference that was not statistically significant ( $P=0.491$ ). A reclassification analysis using a risk score of 19 (median MELD score for the 29 candidates transplanted within 90 days) was also conducted and produced corroborative results (data not shown). The IDI, which does not rely on clinically relevant risk categories but is a less tangible measure than the NRI, was calculated to be 0.030. The interpretation of this result is that the predictions generated by the MELDNa resulted in a 3% improvement in average true-positive incidence (sensitivity) minus average false-positive incidence ( $1 - \text{specificity}$ ); this difference was not statistically significant ( $P=0.137$ ).

**DISCUSSION**

Until now, the MELD and MELDNa prognostic models have not been comprehensively validated in Canadian LT candidates. Although the present analysis was a modest regional validation study, it is hoped that it will be the impetus for further evaluation of the MELD in Canada. The current results regarding the MELD indicate that it has a very good accuracy of discrimination when applied to predict the occurrence of 90-day mortality in Atlantic Canadian LT candidates. The estimate for the area under the ROC curve of 0.887 was similar to the c-statistics reported in previous large validation studies performed in the United States, reported by Wiesner et al (2) (c-statistic 0.83 [95% CI 0.81 to 0.84]) and Kim et al (5) (c-statistic 0.868 [95% CI not reported]). This result was encouraging given that discrimination is the sole performance measure of relevance if the risk score is used exclusively for priority ranking of candidates (ie, specific score thresholds play no role in management decisions). Analysis of the calibration of the MELD produced no evidence to indicate that the model-derived probability estimates for 90-day mortality differed from the observed incidence, although it must be acknowledged that global goodness-of-fit testing was hampered by low event occurrence in the cohort. The only other study that evaluated the calibration of the MELD was reported by Kim et al (5). In that very large validation cohort, significant lack of fit was found for both the MELD and



**Figure 5)** Observed versus sodium augmented form of the Model for End-stage Liver Disease (MELDNa)-predicted incidence of 90-day mortality for the 89 candidates with uncensored 90-day outcomes according to three risk-score quantiles

**TABLE 3**  
**Reclassification analysis based on assignment to risk score of 14 or lower or, 15 and higher (clinically relevant threshold for the Share MELD 15 policy)**

MELD score	MELDNa score		Total
	6 to 14	15 to 40	
<b>6 to 14</b>			
n total	44	14	58
n events	0	1	1
n nonevents	44	13	57
<b>15 to 40</b>			
n total	0	31	31
n events	0	6	6
n nonevents	0	25	25
<b>Total</b>			
n total	44	45	89
n events	0	7	7
n nonevents	44	38	82

Green indicates favourable reclassification; Red indicates unfavourable reclassification. MELDNa Serum sodium augmented form of the Model for End-stage Liver Disease

MELDNa, although the authors indicated that due to the large number of individuals in the cohort, the analysis was powered to detect even small departures in fit (personal communication, Terry Therneau, Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, August 12, 2010).

The present study was the first to externally validate the MELDNa in LT candidates outside of the United States. Compared with the MELDNa derivation study reported by Kim et al (5), the performance characterizations of the MELDNa from the current study were less optimistic and less conclusive. This is undoubtedly and, at least partially, related to the fact that the current study's validation cohort was considerably smaller and, therefore, comparisons were relatively underpowered. Kim et al (5) found the c-statistic for the MELDNa prediction of 90-day wait-list mortality to be 0.883 (95% CI not reported), which is consistent with the point estimate of 0.848 generated in the present study. Unlike the MELDNa derivation study, the present analysis did not find the discrimination accuracy of the MELDNa to be superior to that of the MELD and, in fact, although not statistically significant, the absolute value of the area under the ROC curve for the MELDNa prediction was less than that of the MELD. Evaluation of the calibration of the MELDNa did not produce

any evidence to indicate that the estimates for 90-day wait-list failure differed from the observed, although the validity of the global goodness-of-fit analysis can be challenged based on the low event occurrence in the cohort. There was also no evidence that the MELDNa produced better (or worse) calibrated estimates of 90-day wait-list mortality than the MELD. The reclassification analysis failed to reveal any significant relative improvement in prioritization precipitated by the application of the MELDNa. Therefore, it appeared that in this cohort, the addition of the influence of low serum sodium concentration as an independent variable did not augment the predictive performance of the model.

The reason for this apparent lack of influence of hyponatremia in the present cohort of LT candidates is uncertain. At 35.5%, the prevalence of hyponatremia at the time of wait-list registration was similar to the prevalence of 30.9% observed in the MELDNa derivation study reported by Kim et al (5). Analysis of the current cohort found that serum sodium concentrations and prevalence of hyponatremia did not differ between 90-day wait-list failures and survivors. It seems improbable that the clinicians caring for LT candidates in Atlantic Canada have any greater success at correcting the hyponatremia and, thus, ameliorating this mortality risk in patients on the wait list than clinicians elsewhere. Therefore, in the absence of any plausible physiological or clinical explanation, it is speculated that the low number of 90-day mortality events may be responsible for the association between hyponatremia and wait-list failure being unapparent in the current study.

As was alluded to, a limitation of the present study was the low failure event occurrence and relatively small sample size. The study time frame and subsequent constraints on sample size corresponded to the first five-year experience of the Atlantic Multi-Organ Transplant Program's LT service after a hiatus of 3.5 years, which was imposed by

a lack of surgical manpower. Between May 2001 and November 30, 2004, individuals from Atlantic Canada were required to travel to more distant centres for assessment and provision of surgical LT service. It was believed by the investigators that the wait-list experience of candidates from that period was potentially confounded by accessibility. The period selected for the current study was believed to be reflective of more optimal patient care and current clinical practice in Atlantic Canada. To address this constraint, whenever possible, statistical methods that were deemed to be more appropriate for low event scenarios were used.

Overall, the present study found that the MELD produced accurate predictions for 90-day wait-list mortality in this five-year wait-list cohort and, therefore, appears to be generalizable to adult LT candidates with ESLD residing in Atlantic Canada. The MELDNa predictive model was also found to perform adequately; however, it did not appear to offer any prognostic improvement over the MELD. It is hoped that the current study will provide the impetus for further evaluation of the MELD in Canada and consideration of its adoption as the measure for prioritization of adult LT candidates with ESLD in a national organ allocation system. We propose to conduct a broader MELD validation study and encourage other Canadian LT programs to collaborate with us in this endeavour.

---

**ACKNOWLEDGMENTS:** The authors express their gratitude to Ray Kim, Joanne Benson and Terry Therneau of the Mayo Clinic College of Medicine for taking the time to respond to our inquiries pertaining to the MELD and MELDNa. The authors also thank Michael Pencina for his assistance with questions pertaining to the NRI and IDI.

---

## REFERENCES

- Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 2000;31:864-71.
- Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology* 2003;124:91-6.
- Burak KW. Does the MELD score predict mortality before and after liver transplantation in Alberta? Calgary: University of Calgary, 2005.
- Burak KW, Meeburg G, Myers RP, et al. Should Canada adopt the MELD score for allocation of liver transplants? *Hepatology* 2005;42(Suppl 1):322A. (Abst)
- Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. *N Engl J Med* 2008;359:1018-26.
- Bollman RD, Clemenson HA. Structure and change in Canada's rural demography: An update to 2006 with provincial detail. In: Agricultural Division, ed. Ottawa: Statistics Canada, 2008:6.
- Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001;33:464-70.
- Freeman RB Jr, Wiesner RH, Harper A, et al. The new liver allocation system: Moving toward evidence-based transplantation policy. *Liver Transpl* 2002;8:851-8.
- United Network for Organ Sharing (UNOS). UNOS Policy 3.6 Organ Distribution: Allocation of Livers.   
<[http://www.unos.org/PoliciesandBylaws2/policies/pdfs/policy\\_8.pdf](http://www.unos.org/PoliciesandBylaws2/policies/pdfs/policy_8.pdf)>. (Accessed on November 17, 2009).
- Pepe MS, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves (st0154). *Stata J* 2009;9:1-16.
- Hosmer DW, Lemeshow S, May S. Descriptive methods for survival data. In: Hosmer DW, Lemeshow S, May S, eds. *Applied Survival Analysis: Regression Modelling of Time-to-Event Data*. 2nd edn, Hoboken: John Wiley & Sons, Inc 2008:16-66.
- Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform* 2005;38:404-15.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988;44:837-45.
- Faraggi D, Reiser B. Estimation of the area under the ROC curve. *Stat Med* 2002;21:3093-106.
- Obuchowski NA, Lieber ML. Confidence intervals for the receiver operating characteristic area in studies with small samples. *Acad Radiol* 1998;5:561-71.
- Hosmer DW, Lemeshow S. Assessing the fit of the model. *Applied Logistic Regression*, 2nd edn. Hoboken: John Wiley & Sons, Inc. 2000:143-67,186-8.
- 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;27:157-72; discussion 207-12.



**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

