UNIVERSITA' DEGLI STUDI DI NAPOLI "FEDERICO II"

University of Naples "Federico II"

DOTTORATO DI RICERCA IN
TERAPIE AVANZATE BIOMEDICHE E CHIRURGICHE
XXXIII CICLO

PhD Program in Advanced Surgical and Biomedical Therapies - 33° cycle

TESI DI DOTTORATO

PhD Thesis

Development of a model to predict the risk of early graft failure after adult-to-adult living donor liver transplantation

CANDIDATO
PhD Candidate

Dr. Mariano Cesare Giglio

TUTOR
Prof. Giovanni Domenico De Palma

SUPERVISOR
Prof. Roberto Troisi

ANNO ACCADEMICO 2020/2021

SUMMARY

ABSTRACT

BACKGROUND

Graft survival is a critical endpoint in adult-to-adult living donor liver transplantation (LDLT), where graft procurement endangers the lives of healthy individuals. Therefore, LDLT must be responsibly performed in the perspective of a positive harm-to-benefit ratio. To define the likelihood of failure of an LDLT in the short-term, this study aimed to develop a risk prediction model of early graft failure after LDLT.

METHODS

Using 5201 LDLTs data available from the European Liver Transplant Registry, we studied donor and recipient factors associated with a 3-months graft failure. A risk prediction model of this event was developed using a dual approach, including the Least Absolute Shrinkage and Selecting Operator (LASSO) logistic regression and an artificial neural network (NNET) classification algorithm. Models were built on a training set of 2060 LDLTs and were validated on an independent random-split test sample (n=515). Model performance was assessed using discrimination measures, as the Area Under the Curve of the receiver operating characteristic (ROC) and discrimination slope, calibration plots, and decision curve analysis (DCA), which estimated the net benefit at different threshold probabilities. Prediction models were compared through reclassification indices and DCA.

RESULTS

A 3 months graft failure occurred in 913 of the 5201 LDLTs (17.5%). Multiple donor and recipient characteristics factors were associated with early LDLT failure, with the most important predictors selected (LASSO) being the type of graft, the graft weight, the UNOS status, and the severity of recipient liver disease. The LASSO and NNET model showed similar AUC values, of 0.65 (95% CI, 57 to 70) and 0.67 (95% CI, 60 to 73), respectively. However, the NNET model presented a higher discrimination slope(0.104 vs. 0.017), and compared to the LASSO model, yield a significant improvement in risk reclassification. Also, the NNET model was associated with a higher clinical benefit, resulting in a net reduction in early graft losses varying from 5 to 15 for 100 LDLTs, according to the threshold probability. The NNET risk prediction model is available as an online web application (http://ldlt.shinyapps.io/eltr_app).

CONCLUSIONS

Multiple donor and recipient characteristics are associated with early graft failure. Using a panel of easily available donor and recipient characteristics, an NNET was able to predict this risk with such a performance as to be associated with a significant net clinical benefit.

INTRODUCTION

Living donor liver transplantation (LDLT) is an effective therapeutic opportunity for patients with end-stage liver diseases. Due to the scarce availability of grafts from deceased donors, LDLT represents the first option for liver replacement in the eastern countries and remains the alternative to deceased-donor liver transplantation (DDLT) in the western countries, where it contributes to expanding the organ pool. Different from DDLT, LDLT has no problems with graft allocation priority and pauperization of the pool of organs. Nonetheless, graft survival remains a crucial ethical endpoint in LDLT, since grafts are procured risking the lives of healthy individuals. Indeed, LDLT remains associated with major donor morbidity[1] and has also a significant psychological[2,3] and financial impact[4] on the donors' lives. Therefore, LDLT must be carried out responsibly within the perspective of a favourable harm-to-benefit ratio.

In the years, several factors related to both the donor and the recipient have been associated with poor graft survival following LDLT. Adequate graft size has been first described as a determinant of LDLT outcomes[5]. Later improvements in surgical techniques[6], knowledge of physiopathology[7,8], as well as accurate recipient selection, have contributed to reducing the impact of graft size on LDLT outcomes[9]. Additional factors, including donor age[10], graft steatosis[11,12], the severity of recipient liver disease[13], and other characteristics of the donor-to-recipient matching[14] have been associated with graft survival over the years. However, due to the large number and complex interaction of risk factors, it remains difficult for the clinicians to compute the overall risk of failure for each potential LDLT.

A tool for an accurate assessment of this risk would be of clinical relevance considering its possible implications. Several risk prediction models have been developed in the field of DDLT, mainly to guide organ allocation and prioritize patients on the waiting list[15–17]. However, these models focus on specific characteristics of DDLT and ignore others relevant to LDLT[18]. In 2014 Goldberg et al. published the "Living donor risk index", a score to predict long-term graft survival in LDLT[19]. However, this prediction model was developed in a population of recipients with mildly severe liver disease and could not consider this important predictor. Also, the score stratification ability was limited, since the worst scores were associated with a graft survival of 82% and 60% at 1 and 5 years, respectively. Although suboptimal, such predicted outcomes are unlikely to alter the decisional process.

On this background, this study aimed to develop a risk prediction model of early graft failure, based on a panel of relevant donor and recipient characteristics, to aid clinicians to better estimate the likelihood of failure in the short term of an adult-to-adult LDLT.

MATERIAL AND METHODS

STUDY POPULATION

Data for the present study were obtained from the European Liver Transplant Registry (ELTR) with the consent of the board of the European Liver Intestine and Transplant Association (ELITA). All data available regarding adult-to-adult LDLT were extracted from the ELTR registry. Data regarding country and center of origin were anonymized. LDLT consisting of re-transplants after the failure of a previous DDLT or LDLT were excluded. All other available adult-to-adult LDLTs were included in the study, without any exclusion criteria based on donor and recipient characteristics.

DATA ANALYSIS AND ENDPOINTS

Donor characteristics analyzed were age, sex, height, weight, body mass index (BMI), and ABO group, while graft characteristics included the type of graft (right lobe, left lobe or left lateral lobe, or others), weight, and rate of micro- or macro-steatosis, characterized as none, mild (<30%), moderate (30-60%), and severe (>60%). Recipient factors analyzed were age, sex, height, weight, BMI, ABO group, date of transplant, the United Network for Organ Sharing (UNOS) status, transplant performed as urgent, multiorgan transplantation, need for dialysis (at list twice /week), presence of ascites or encephalopathy, the disease leading to LDLT, laboratory values regarding pre-transplant serum albumin, creatinine, bilirubin, international normalized ratio (INR), serum sodium, and pre-transplant Model of end-stage liver disease (MELD) score, MELD-Na score, Child Score and Child Class. Finally, follow-up data regarding graft and patient survival were also extracted, as well as the occurrence and time of re-transplantation.

According to the ELTR instructions, the MELD score available in the database was used for the analysis. In the case of missing values, MELD was calculated based on the available laboratory parameters. MELD-Na, as well as the Child-Pugh score, were also calculated. The graft to recipient body weight ratio (GWBWR) was calculated and the status of ABO compatibility and gender mismatch was established for each LDLT.

Centers were divided into quartiles according to the volume of LDLTs recorded into the ELTR. In addition, the early LDLTs for each centre, i.e. the first 15 cases recorded in the ELTR by each centre were identified for specific analysis, to control for a learning curve effect[20].

Graft loss was defined as recipient death or graft failure necessitating liver re-transplantation. The occurrence of graft loss at 3 months was the primary endpoint.

STATISTICAL ANALYSES

Categorical data are presented as frequency counts and percentages, and continuous data as medians and interquartile ranges (IQRs) in case of skewed distributions or as mean and standard deviation (SD) in presence of symmetrical distribution. Categorical variables were compared using χ^2 tests, while continuous variables were compared using parametric o non-parametric tests, as appropriate. The shape of the association between predictors and outcome was modelled using local weighted regression (LOESS)[21] between the variable and the probability of 3-months graft loss estimated using logistic regression.

For inference purposes, to study the independent association between variables and the outcome, a multivariable logistic regression model was fit with variables presenting a p-value <

0.10 at univariable analysis. In case of evidence of multicolliearity between two or more predictors (e.g. MELD and MELD-Na), to reduce statistical noise and avoid type II errors, only one was entered into the multivariable model. The strength of the association is reported as Odds Ratio along with a 95% confidence interval.

Model development was based on the following variables: donor and recipient age, sex, height, weight, BMI, ABO group, the type of graft, graft weight, the GWBWR, the liver disease, recipient UNOS status, urgent and multiorgan transplantation, the recipient Child Class, CHILD, MELD and MELD-Na scores and each item needed for their calculation.

Missing data were considered as missing at random, as also confirmed by the evidence of an almost identical incidence of graft failure between complete and non-complete cases. Therefore, models were built on cases with complete data for the abovementioned variables. The dataset was randomly split into a training (80%) and test (20%) set.

For prediction purposes, a dual approach was used for model development. As the first approach, we used the least absolute shrinkage and selection operator (LASSO) logistic regression, an extension of linear regression that automates the selection of a subset of variables and optimizes the predictive accuracy by adding a regularization penalty to the loss function during training[22]. The tuning parameter λ was selected using 10-fold cross-validation in the training set.

As the second approach, we trained a feedforward artificial neural network classification algorithm with stochastic gradient descent using back-propagation. Model tuning was performed using 10-fold cross-validation on the training set. Feature importance was calculated according to the method described by Gedeon[23].

Model performance was assessed both in the training set, using cross-validation, and in the test set. Model discrimination was measured as the area under the receiver-operating characteristic curve (AUC), which measures the concordance of predictions with actual outcomes, and by the discrimination slope, which measures how well subjects with and without the outcome are separated[24]. Discriminative capacity was also represented using box plots, showing the distribution of the predicted risks according to the actual occurrence of the event. Sensitivity and specificity, positive and negative predictive values were calculated at different risk cut-offs. Calibration was evaluated by comparing the observed with the predicted rate of events and graphically represented by calibration bar plots.

Models were compared using net reclassification indices such as the integrated discrimination improvement (IDI) and the net reclassification improvement (NRI)[25,26]. The IDI measures the improvement produced by a second model as the percentage increase in the difference of the predicted risks between cases with and without the outcome[25]. The continuous NRI refers to the proportion of cases with and without the outcome correctly assigned by the second model to a higher or lower risk, respectively[25].

Also, we used decision curve analysis to assess and compare the usefulness of the developed models by quantifying the resulting net benefit at different threshold probabilities[27,28]. The net benefit is calculated as the number of total true positive classifications minus the total false-positive classifications weighted by the odds of the thresholds probability. For better comprehension, the net benefit was also expressed in terms of net reductions, that is the net number of graft failures avoided without missing any successful transplant.

Although graft weight and GWBWR were used for model development, these data are accurately known only after graft procurement, and before surgery, only an estimate of the graft weight is available, subject to a variable error[29]. Therefore, we tested model performance using

these estimates. Since they were not available in the ELTR, starting from the actual graft weight, we backward simulated the estimate by applying the estimation errors observed in the daily practice[29]. Also, we developed and tested a simplified prediction model excluding graft weight and GWBWR from predictors. Subgroups analyses were also planned to test model performance within specific categories of LDLTs, like those performed in high volume centers (3rd and 4th quartile), those representing "late cases", and those performed on recipients having a MELD score > 24, and a GWBWR <1 and < 0.8.

The artificial neural network was developed within the *h2o* framework. The LASSO regression was performed using the *glmnet* package. All statistical analyses were performed using R version 3.6.1 (2019, The R Foundation for Statistical Computing).

RESULTS

STUDY POPULATION AND PREDICTORS OF EARLY GRAFT FAILURE

Between Oct 20, 1996, and Dec 30, 2019, 88 transplant centers from 18 European countries recorded data of 6403 LDLTs into the ELTR. Thirty-six of these were excluded as they were liver-retransplantation after the failure of a primary LDLT or DDLT. The 3-months graft survival status was known for 5201 LDLTs which were therefore included in the study. Table 1 summarizes donors' and recipients' characteristics. A 3-months graft failure occurred after 913 transplants (17.5%) and was associated with donor age, donor and recipient BMI, and left-sided grafts. Also, early graft failure was more frequent in recipients with more severe liver disease, with a higher UNOS status, higher MELD, MELD-NA, and CHILD scores, and in those necessitating urgent liver transplantation. Concerning the donor-recipient matching, graft failure was associated with ABO incompatibility, with gender mismatch (female donor and male recipient), and with a lower graft-to-recipient body weight ratio. Figure 1 shows the shape of the relation between continuous predictors and the probability of 3-months graft failure based on logistic regression. The risk of graft failure was higher (inverse relation) for GWBWR < 1, remained stable for GWBWR values between 1 and 1.5, and rise again, with a milder slope, for higher GWBWR values. Similarly, the risk of a worse outcome increased in presence of low (BMI < 18) and high (BMI > 30) recipient BMI values. Also, the probability of graft loss increased in presence of donor BMI > 30. On the contrary, donor age, MELD, MELD-Na, and Child score were linearly associated with an increased probability of early graft loss over the entire range of values. Data on graft steatosis were missing for a high number of grafts (81.6%), and macro-steatosis was reported as absent or mild (<30%) in 88.3% and 11.5% of the grafts, respectively. Table 2 summarizes the occurrence of graft failure according to time and center related factors. No significant differences in the rate of occurrence were observed according to the year of transplant (p=0.88) and according to the center volume of LDLTs (p=0.103). On the contrary, a higher incidence of early failure was observed during the first 15 cases in each center (20% vs 17.1%, p=0.0409). Table 3 shows the results of the univariable analysis for each candidate predictor. At multivariable analysis, MELD score, UNOS status (ICU bound or hospitalized), and graft weight resulted independently associated with early graft failure (Table 3).

MODELS DEVELOPMENT AND VALIDATION

Models were built on LDLTs with complete data (n=2575). There was no difference in the incidence of graft failure between cases with and without missing values (p=0.99). The dataset was randomly split into a training set (n=2060) and a test set (n=515). As expected, no difference in baseline characteristics and event rate was observed between the training and test set (Table 4). The logistic LASSO regression model selected four predictors (Table 5), including Child score, right Graft, UNOS status (Hospital or ICU), and the graft weight. This model yielded a cross-validated AUC of 0.66, while, at validation on the test set the AUC was 0.65 (95%CI, 0.57 to 0.70) and the discrimination slope 0.017. Figure 2 shows the discriminative ability of the LASSO model. A calibration bar plot is reported in Figure 3.

The final NNET architecture consisted of *5* hidden layers of *n* neurons per layer with rectified linear unit activations and a logistic output. The relative importance of each feature is

shown in Figure 4. The NNET model showed an AUC of 0.68 at cross-validation and of 0.67 (95%CI, 0.60 to 0.73) at validation on the test set. The discrimination slope was 0.104 (Fig 2). Model calibration is presented in Figure 3.

MODELS COMPARISON

Compared to the LASSO regression, the NNET model significantly improved patient reclassification ($p<0.001$), with an NRI of 47.7% (95% CI, 26 to 69.5) and IDI of 8.75% (95%, 4.9 to 12.5, $p<0.001$). Also, the NNET model showed a higher discriminative ability, with an almost 10-folds higher discrimination slope (0.104 vs. 0.017) and with a better risk distribution between LDLTs with and without graft failure (Figure 5).

The decision curve analysis showed a higher, net clinical benefit for the NNET model compared to the LASSO regression model, over a wide range of threshold probabilities (Figure 6). Also, a net benefit was evident for the NNET model compared to the strategy "transplant all" (i.e. no risk assessment). The net benefit is also expressed in terms of net reduction in graft losses for 100 LDLTs (Figure 6).

INFLUENCE OF THE GRAFT WEIGHT ESTIMATION ERROR

When predictions were made using the estimated rather than actual graft weights and GWBWR, changes in the individual patient risk estimation were -0.2% (-4.6%,4.3%) for the NNET model and +0.27% (0.18%,0.36%) for the LASSO model. No significant differences were observed in patient reclassifications (NRI), both for the NNET ($p=0.46$) and the LASSO ($p=0.18$) model, when compared to the respective predictions based on the actual values.

DEVELOPMENT OF NNET PREDICTION MODEL WITHOUT GRAFT WEIGHT AND GWBWR

An NNET was also trained with all, except graft weight and GWBWR, donor and recipient characteristics (simplified NNET model). Variable importance within this model is reported in Figure 7. This model had a cross-validated AUC of 0.68 and an AUC of 0.66 (95% CI, 0.59 to 0.73) on the test set. The discrimination slope was 0.116. Compared to the full NNET model, no significant changes were observed in terms of patient reclassifications (NRI and IDI, $p=0.46$ for both). Also, the decision curve showed a nearly equal net benefit (Figure 8). This model is made available as an online application at the web address http://ldlt.shinyapps.io/eltr_app

MODEL PERFORMANCE IN SUBGROUPS OF LDLT

The simplified NNET model showed consistent performance when tested in specific subgroups of LDLTs within the test set. In particular, the AUC values remained unaltered (0.66, 95% CI, 0.60 to 0.73) when the simplified NNET model was applied only to "late cases" (0.66, 95% CI, 0.60 to 0.73). Moreover, the simplified NNET model showed an higher performance in presence of recipients with a MELD > 24, with an AUC of 0.80 (0.67 to 0.93). Also, within the test set, performance tended to be higher in case of recipients presenting a GWBWR < 1 (n=174), with AUC of 0.67 (95% CI, 0.57 to 0.78) and in 39 recipients (test set) with a GWBWR < 0.8, AUC 0.81 (95% CI, 0.65 to 0.98).

RISK CUT-OFFS

The simplified NNET model predicted the risk of graft failure < 5% in 21.5% of the LDLTs, and an early graft failure occurred in 3% of them. The majority of the LDLTs (50.7%) had a predicted risk of failure < 12%, with effective failure, observed in 6% of them. Five percent of the LDLTs had a predicted risk of failure ≥ 50%, and the event occurred in 65% of them.

The sensitivity and specificity of the simplified NNET model at different risk cut-offs are shown in Figure 9 and Table 6. A cut-off of 22.4% presented the highest Youden index, with values of specificity and sensitivity of 0.80 and 0.60, respectively. Risk cut-offs of 30% and 50% presented specificity values of 0.85 and 0.97, and positive predictive values of 0.41 and 0.65, respectively. Cut-offs of 5 % and 10% were associated with sensitivity values of 0.96 and 0.85, respectively. The Fagan nomogram shows changes in the individual probability of graft failure after risk assessment using the simplified NNET model(Figure 10).

DISCUSSION

This study aimed to develop a predictive model of the risk of early graft failure after LDLT. For this purpose, an artificial neural network classification algorithm was trained using a panel of donor and recipient characteristics available from more than two thousand LDLTs. The developed model presented fair discriminative ability and calibration, and, mostly, was associated with a net clinical benefit. The developed neural network is deployed as an online risk calculator.

Three-months graft failure occurred in 17.5% of LDLT recorded into the ELTR, an incidence not far from that reported in other series[30]. The prediction of early graft failure has, therefore, a clinical relevance, also in consideration of the possible implications. Indeed, in the presence of multiple potential donors, a tool to identify the one that best matches the recipient and minimizes the risk of failure would be valuable. Furthermore, patients waiting for a DDLT could evaluate the likelihood of success of an LDLT, switching to this option in the presence of favorable conditions, or vice-versa. Also, since graft procurement endangers the life of a healthy individual, living donation is allowed within the perspective of adequate benefits for the recipient. Undeniably, it is difficult determining how long a graft must survive to justify a living donation. However, it is unquestionable that a graft surviving only 3-months does not offer benefits worthy of the risks of surgery on a healthy subject such as a living donor. In accordance, most liver donors expect their donation to improve recipient survival by at least 6 months[31,32]. In this perspective, an accurate estimate of the risk of early graft failure provides both the medical team and donors with objective data for making their decisions.

Determinants of graft failure are multiple and pertain to the donor, recipient, and the result of their matching. The in-depth analysis performed on a large cohort of ALLDTs confirmed the importance of most of the predictors reported to the literature so far, describing the shape and strength of their association with the occurrence of graft failure. Of note, in the context of a controversial literature[33], the present analysis confirmed that the severity of recipient liver disease is a determinant of LDLTs results. A further substantial finding is that type of graft and graft weight independently predict graft failure, supporting the evidence that the superior outcomes achieved with right lobes do not depend only on the larger size of these grafts. Also, for both LASSO and NNET models, the rough graft weight was a better predictor of graft failure than the GWBWR (Figure 4). This suggests that the recipient body weight is a good, but probably not the best comparator for the graft weight. Consistently, the NNET found more informative the height rather than the weight of the recipient (Figure 4).

The evidence that the NNET outperformed the LASSO regression model supports the presence of an intricate relation between the predictors and the outcome. Probably, the NNET could understand these complex, nonlinear interactions existing at multiple levels better than LASSO regression could. A further advantage is also that NNET can use all the information provided by a large set of clinical meaningful parameters, without necessitating any a priori selection of predictors, which likely leads to a loss of information. The superiority of NNETs over other machine learning algorithms and linear regression in predicting transplant outcomes has been also confirmed elsewhere[34].

To ensure good applicability, the prediction model was built on parameters that are easily available before surgery. The graft steatosis rate was therefore excluded since the need for liver biopsy results would limit the model usability. Although donor's liver biopsy is mandatory in most

transplant centres and the information gathered serves also to improve donor safety[35–38], this procedure is usually performed at a late workup stage, and not during the initial screening phase where a prediction model would be more useful. In addition, this information was missing for 80% of the patients. However, a surrogate for graft steatosis such as donor BMI was included in the model[39]. Graft weight and GWBWR were at first included as predictors, although only an estimate, subject to error, is available before surgery[29]. Surprisingly, the model's predictions remained accurate also when based on these estimates, possibly because the estimation error has a negligible impact on the overall risk computation. Nonetheless, a predictive model independent from graft weight (and GWBWR) was also developed. Even more surprisingly, this simplified NNET model hold the same performance as the full NNET model. This leads to hypothesize that the NNET, using donor's and recipient's anthropometric data, can probably estimate the grade of the graft size to recipient size mismatch with such approximation to efficiently predict the risk even in absence of information regarding the graft. As confirmation of this hypothesis, after excluding the graft weight from predictors, the donor's height gained a significant importance among the predictors of graft failure (Figures 4 and 7). The equivalence performance between the full and simplified NNET model is relevant, meaning that a risk assessment can be obtained even before performing any liver volumetric assessment on the donor.

The NNET model showed a discriminative capacity comparable to those of other prediction models widely employed in liver transplantation[17,18]. In this setting, suboptimal AUC values are probably due to the occurrence of unpredictable events, such as intraoperative and postoperative complications that also affect graft survival and distort the relationship between predictors and outcomes. In absence of alternatives, however, these tools remain the only valuable support to the clinical decision-making process in liver transplantation. Model calibration was also fair, particularly for risks < 50%, which encompassed most of the model predictions. Since discrimination and calibration do not fully account for the clinical utility of a prediction tool, we assessed this aspect using the decision curve analysis[27]. This quantified the resulting net clinical benefit over a range of threshold probabilities, each of these representing a different harm-benefit ratio, that is, a different weighing of risks versus benefits[40]. The higher the threshold, the more serious a graft loss is considered with respect to a successful transplant. For example, in DDLT, the goal of graft survival of 70% at 5 years is represented by a threshold probability of 0.7, corresponding to an accepted 7:3 odds of graft survival versus non-survival at 5 years. Indeed, it is difficult to identify such a threshold for LDLT, which also varies according to the clinical situations and donor's expectations. Nonetheless, a threshold less than 0.5 remains inadmissible, since it corresponds to an odds of 1, i.e. accepting one early graft failure for each non-failure. Remarkably, the NNET models were associated with a net clinical benefit for all threshold probabilities > 0.5.

Since the model provides the risk of graft failure as a continuous value between 0 and 100%, adopting a cut-off for taking clinical decisions is needed. Graft failure increases significantly for risks > 30% and a cut-off value of 22.4% was found to optimizes both sensitivity (0.60) and specificity (0.80). However, cut-offs that maximize either the specificity or sensitivity might be more useful in specific settings. For example, a high-sensitivity cut-off (e.g. 5%) might help in safely choosing between multiple donors minimizing the risk of false-negative predictions. On the contrary, a high-specificity cut-off (e.g. 40%) would select LDLTs at a very high risk of failure. This should prompt the evaluation of alternative options (e.g. DDLT), a careful discussion with the donor about the likelihood of a futile donation, or the decision to refraining from an LDLT due to a very unfavourable harm-to-benefit ratio. In this view, risk predictions would be best used as a

continuous value, rather than dichotomized in a low/high risk, and should be interpreted in the light of the specific clinical setting.

This study has some limitations. Treatment effect on outcomes could not be assessed, as frequently occurs in risk prediction modelling. For example, the impact on outcomes of surgical[7,41,42] or pharmacological treatments[43] to moderate portal hypertension in presence of a low GWBWR could not be appraised, since such procedures are not recorded in the ELTR. Similarly, the effect of target therapies in presence of ABO incompatibility could not be assessed[44]. As a consequence, the best available treatment options, either surgical or medical, should be ensured to patients despite the predicted risk of failure. The model was built on a large, representative cohort of patients from different centers and countries, and showed a consistent performance at validation in a random internal independent set. Albeit this suggests optimal performance in new cohorts of patients, external validation remains indispensable. Finally, this model cannot compute that part of risk of failure depending on graft anatomy, e.g. the necessity of performing multiple biliary or arterial anastomoses, venous reconstructions, which needs to be evaluated case by case by the surgical team.

Despite these shortcomings, this study has several strengths points and adds some valuable information to the literature. An unprecedented sample of LDLTs allowed a powered statistical analysis which confirmed most of the information individually reported so far to the literature. Compared to the living donor risk score[19], this prediction model moves some steps forward, as it incorporates important predictors as the indices of liver disease severity (i.e MELD score, CHILD score). Furthermore, in contrast to the prediction of medium to long-term graft survival offered by the living donor risk score[19], the present model focuses on a robust short-term endpoint which indisputably equates to zero benefits from LDLT. As such, this model provides consistent information for clinical decision making. Notably, the NNET showed a higher performance in those LDLTs traditionally judged to be at greater risk of failure, like those performed on recipients with high MELD scores or in the presence of a low GWBWR. Despite being based on a complex algorithm, this prediction model is made easily usable by everyone through a web application.

In conclusion, several donor and recipient related factors determine the risk of graft failure following LDLT through a complex interaction. A comprehensive evaluation of all these parameters could improve donor and recipient selection and contribute to further improve recipient outcomes while avoiding at the same time futile surgery on healthy individuals. Given the amount of information to process, artificial intelligence can play a role by overcoming the limits of the human mind and conventional statistics.

ACKNOWLEDGMENTS

I warmly thank Dr. Pasquale Dolce, from the Department of Public Health, Federico II University of Naples, for his invaluable contribute to this work, and the board members of the European Liver and Intestine Transplant Association for endorsing the project.

My deep and sincere gratitude goes also to Professor Giovanni Domenico De Palma, who made me discover my interest in surgical research early in my surgical training and strongly supported my PhD.

Finally, this work would have not been possible without the support, encouragement, and scientific supervision of Professor Roberto Troisi. Every young surgeon who is passionate about hepatobiliary and pancreatic surgery deserves an academic mentor like him.

REFERENCES

1. Ghobrial RM, Freise CE, Trotter JF, Tong L, Ojo AO, Fair JH, et al. Donor Morbidity After Living Donation for Liver Transplantation. Gastroenterology. 2008;135(2):468–76.

2. Wang SH, Lin PY, Wang JY, Huang MF, Lin HC, Hsieh CE, et al. Mental health status after living donor hepatectomy. Med (United States). 2017;96(19).

3. Butt Z, Dew MA, Liu Q, Simpson MA, Smith AR, Zee J, et al. Psychological Outcomes of Living Liver Donors From a Multicenter Prospective Study: Results From the Adult-to-Adult Living Donor Liver Transplantation Cohort Study2 (A2ALL-2). Am J Transplant. 2017;17(5):1267–77.

4. DiMartini A, Dew MA, Liu Q, Simpson MA, Ladner DP, Smith AR, et al. Social and Financial Outcomes of Living Liver Donation: A Prospective Investigation Within the Adult-to-Adult Living Donor Liver Transplantation Cohort Study 2 (A2ALL-2). Am J Transplant. 2017;17(4):1081–96.

5. Kiuchi T, Kasahara M, Uryuhara K, Inomata Y, Uemoto S, Asonuma K, et al. Impact of graft size mismatching on graft prognosis in liver transplantation from living donors. Transplantation. 1999;67(2):321–7.

6. Troisi R, de Hemptinne B. Clinical relevance of adapting portal vein flow in living donor liver transplantation in adult patients. Liver Transplant. 2003 Sep;9(9):S36–41.

7. Ito T, Kiuchi T, Yamamoto H, Oike F, Ogura Y, Fujimoto Y, et al. Changes in portal venous pressure in the early phase after living donor liver transplantation: pathogenesis and clinical implications. Transplantation. 2003;75:1313–7.

8. Man K, Fan ST, Lo CM, Liu CL, Fung PCW, Liang TB, et al. Graft Injury in Relation to Graft Size in Right Lobe Live Donor Liver Transplantation: A Study of Hepatic Sinusoidal Injury in Correlation with Portal Hemodynamics and Intragraft Gene Expression. Ann Surg. 2003;237:256–64.

9. Wong TC-L, Fung JYY, Cui TYS, Sin SL, Ma KW, She BWH, et al. The Risk of Going Small. Ann Surg. 2020;Publish Ahead of Print.

10. Kubota T, Hata K, Sozu T, Ueda Y, Hirao H, Okamura Y, et al. Impact of Donor Age on Recipient Survival in Adult-to-Adult Living-donor Liver Transplantation. Ann Surg. 2018;267(6):1126–33.

11. Soejima Y, Shimada M, Suehiro T, Kishikawa K, Yoshizumi T, Hashimoto K, et al. Use of steatotic graft in living-donor liver transplantation. Transplantation. 2003;76(2):344–8.

12. Chu MJJ, Dare AJ, Phillips ARJ, Bartlett ASJR. Donor Hepatic Steatosis and Outcome After Liver Transplantation: a Systematic Review. J Gastrointest Surg. 2015;19(9):1713–24.

13. Marubashi S, Dono K, Asaoka T, Hama N, Gotoh K, Miyamoto A, et al. Risk Factors for Graft Dysfunction After Adult-to-Adult Living Donor Liver Transplantation. Transplant Proc. 2006;38(5):1407–10.

14. Giovanardi F, Melandro F, Laureiro ZL, Merli M, Lattanzi B, Hassan R, et al. Donor-to-recipient gender match in liver transplantation: A systematic review and meta-analysis.

World J Gastroenterol. 2018;24(20):2203–10.

15. Rana A, Hardy MA, Halazun KJ, Woodland DC, Ratner LE, Samstein B, et al. Survival Outcomes Following Liver Transplantation (SOFT) score: A novel method to predict patient survival following liver transplantation. Am J Transplant. 2008;8(12):2537–46.

16. Braat AE, Blok JJ, Putter H, Adam R, Burroughs AK, Rahmel AO, et al. The eurotransplant donor risk index in liver transplantation: ET-DRI. Am J Transplant. 2012;12(10):2789–96.

17. Dutkowski P, Oberkofler CE, Slankamenac K, Puhan MA, Schadde E, Müllhaupt B, et al. Are there better guidelines for allocation in liver transplantation?: A novel score targeting justice and utility in the model for end-stage liver disease era. Ann Surg. 2011;254(5):745–53.

18. De Boer JD, Putter H, Blok JJ, Alwayn IPJ, Van Hoek B, Braat AE. Predictive Capacity of Risk Models in Liver Transplantation. Transplant Direct. 2019;5(6).

19. Goldberg DS, French B, Abt PL, Olthoff K, Shaked A. Superior survival using living donors and donor-recipient matching using a novel living donor risk index. Hepatology. 2014;60(5):1717–26.

20. Olthoff KM, Abecassis MM, Emond JC, Kam I, Merion RM, Gillespie BW, et al. Outcomes of adult living donor liver transplantation: Comparison of the adult-to-adult living donor liver transplantation cohort study and the national experience. Liver Transplant. 2011;17(7):789–97.

21. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74(368):829–36.

22. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B. 1996;

23. Gedeon TD. Data mining of inputs: analysing magnitude and functional measures. Int J Neural Syst. 1997;

24. Yates JF. External correspondence: Decompositions of the mean probability score. Organ Behav Hum Perform. 1982;30(1):132–56.

25. Pencina MJ, D'Agostino RB, D'Agostino RB, 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):157–72.

26. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Vol. 21, Epidemiology. 2010. p. 128–38.

27. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. Vol. 313, JAMA - Journal of the American Medical Association. 2015. p. 409–10.

28. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: Guidance for correct interpretation and appropriate use. J Clin Oncol. 2016;

29. Radtke A, Sotiropoulos GC, Nadalin S, Molmenti EP, Schroeder T, Lang H, et al. Preoperative volume prediction in adult living donor liver transplantation: How much can we rely on it? Essen experience based on virtual three-dimensional computed tomography-

volume assessment. Am J Transplant. 2007;7(3):672–9.

30. Olthoff KM, Merion RM, Ghobrial RM, Abecassis MM, Fair JH, Fisher RA, et al. Outcomes of 385 adult-to-adult living donor liver transplant recipients: A report from the A2ALL consortium. In: Annals of Surgery. 2005. p. 314–25.

31. Molinari M, Matz J, Decoutere S, El-Tawil K, Abu-Wasel B, Keough V. Live liver donors' risk thresholds: Risking a life to save a life. Hpb. 2014;16(6):560–74.

32. Lieber SR, Schiano TD, Rhodes R. Should living donor liver transplantation be an option when deceased donation is not? Vol. 68, Journal of Hepatology. 2018. p. 1076–82.

33. Selzner M, Kashfi A, Cattral MS, Selzner N, McGilvray ID, Greig PD, et al. Live donor liver transplantation in high meld score recipients. Ann Surg. 2010;251(1):153–7.

34. Kantidakis G, Putter H, Lancia C, Boer J de, Braat AE, Fiocco M. Survival prediction models since liver transplantation - comparisons between Cox models and machine learning techniques. BMC Med Res Methodol. 2020;20(1).

35. Tan HP, Patel-Tom K, Marcos A. Adult living donor liver transplantation: Who is the ideal donor and recipient? Vol. 43, Journal of Hepatology. 2005. p. 13–7.

36. Akabayashi A, Slingsby BT, Fujita M. The first donor death after living-related liver transplantation in Japan [1]. Vol. 77, Transplantation. 2004. p. 634.

37. Nadalin S, Malagó M, Valentin-Gamazo C, Testa G, Baba HA, Liu C, et al. Preoperative donor liver biopsy for adult living donor liver transplantation: Risks and benefits. Liver Transplant. 2005;11(8):980–6.

38. Yamamoto K, Takada Y, Fujimoto Y, Haga H, Oike F, Kobayashi N, et al. Nonalcoholic steatohepatitis in donors for living donor liver transplantation. Transplantation. 2007;83(3):257–62.

39. Liu ZJ, Gong JP, Yan LN. Quantitative estimation of the degree of hepatic macrovesicular steatosis in a disease-free population: A single-center experience in mainland China. Liver Transplant. 2009;15(11):1605–12.

40. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352.

41. Troisi R, de Hemptinne B. Clinical relevance of adapting portal vein flow in living donor liver transplantation in adult patients. Liver Transpl. 2003;9:S36--41.

42. Boillot O, Delafosse B, Méchet I, Boucaud C, Pouyet M. Small-for-size partial liver graft in an adult recipient; a new transplant technique. Lancet. 2002;359:406–7.

43. Troisi RI, Vanlander A, Giglio MC, van Limmen J, Scudeller L, Heyse B, et al. Somatostatin as inflow modulator in liver-transplant recipients with severe portal hypertension: A randomized trial. Ann Surg. 2019;269(6):1025–33.

44. Egawa H, Teramukai S, Haga H, Tanabe M, Mori A, Ikegami T, et al. Impact of rituximab desensitization on blood-type-incompatible adult living donor liver transplantation: A japanese multicenter study. Am J Transplant. 2014;14(1):102–14.

TABLES

Table 1. Donor and recipient characteristics

		Total (N=5201)	**3-months Graft loss**		
	n[†]	Summary measure [a]	No (N=4288)	Yes (N=913)	*p* value
Donor age (years)	5073	33.0 (26.0, 42.0)	32.0 (26.0, 41.0)	34.0 (27.0, 44.0)	< 0.001
Donor gender (Female)	5059	2097 (41.5%)	1678 (80.0%)	419 (20.0%)	< 0.001
Donor BMI (kg/m^2)	4580	24.3 (22.2, 26.6)	24.2 (22.1, 26.5)	24.5 (22.3, 26.9)	0.014
Type of graft	4001				< 0.001
LLS		33 (0.8%)	21 (63.6%)	12 (36.4%)	
Left		156 (3.9%)	102 (65.4%)	54 (34.6%)	
Right		3807 (95.2%)	3231 (84.9%)	576 (15.1%)	
Other		5 (0.1%)	2 (40.0%)	3 (60.0%)	
Graft weight (g)	3375	795.0 (690.0, 904.0)	800.0 (700.0, 910.0)	760.0 (650.0, 880.2)	
GWBWR	3317	1.1 (0.9, 1.3)	1.1 (0.9, 1.3)	1.0 (0.9, 1.3)	0.001
≤ 0.8		265 (8.0%)	192 (72.5%)	73 (27.5%)	< 0.001
> 0.8		3052 (92.0%)	2585 (84.7%)	467 (15.3%)	
Recipient Age	5201	51.9 (42.0, 58.7)	51.9 (42.0, 58.7)	51.8 (42.0, 58.9)	0.778
Recipient gender (Female)	5200	1796 (100.0%)	1456 (81.1%)	340 (18.9%)	0.059
Recipient BMI	4500	25.3 (22.5, 28.4)	25.4 (22.6, 28.4)	24.9 (22.4, 28.0)	0.044
Recipient ABO Group	5180				0.019
A		2212 (42.7%)	1863 (84.2%)	349 (15.8%)	
AB		319 (6.2%)	259 (81.2%)	60 (18.8%)	
B		811 (15.7%)	647 (79.8%)	164 (20.2%)	
O		1838 (35.5%)	1501 (81.7%)	337 (18.3%)	
ABO incompatible (Yes)	5065	63 (1.2%)	44 (1.1%)	19 (2.1%)	0.009
Liver disease	5201				< 0.001
Hep B		1430 (27.7%)	1231 (86.1%)	199 (13.9%)	
Alcohol		1064 (20.6%)	853 (80.2%)	211 (19.8%)	
HCC		690 (13.4%)	587 (85.1%)	103 (14.9%)	
Hep C		658 (12.7%)	542 (82.4%)	116 (17.6%)	
Cholestasis		387 (7.5%)	316 (81.7%)	71 (18.3%)	
Metabolic		229 (4.4%)	189 (82.5%)	40 (17.5%)	
Autoimmune		169 (3.3%)	135 (79.9%)	34 (20.1%)	
Acute / Subacute liver failure		151 (2.9%)	94 (62.3%)	57 (37.7%)	
Budd-Chiari		96 (1.9%)	69 (71.9%)	27 (28.1%)	
Congenital biliary disease		33 (0.6%)	25 (75.8%)	8 (24.2%)	
Other		257 (5.0%)	215 (83.7%)	42 (16.3%)	
MELD score	3758	15.0 (11.0, 20.0)	15.0 (11.0, 20.0)	17.0 (12.0, 22.0)	< 0.001
<24		585 (15.6%)	2640 (83.2%)	533 (16.8%)	< 0.001
≥24		3173 (84.4%)	441 (75.4%)	144 (24.6%)	
MELD-Na score	3422	18.1 (13.7, 23.8)	17.7 (13.0, 23.0)	20.6 (15.6, 26.2)	< 0.001
CHILD score	3448	9.0 (7.0, 10.0)	9.0 (7.0, 10.0)	9.0 (8.0, 11.0)	< 0.001
CHILD class	3448				< 0.001
A		555 (16.1%)	500 (90.1%)	55 (9.9%)	
B		1552 (45.0%)	1326 (85.4%)	226 (14.6%)	
C		1341 (38.9%)	1072 (79.9%)	269 (20.1%)	
Urgent transplant (Yes)	4836	191 (3.9%)	122 (63.9%)	69 (36.1%)	< 0.001
Multiorgan transplant (Yes)	5201	25 (0.5%)	22 (88.0%)	3 (12.0%)	0.464
UNOS status	4603				< 0.001
ICU		181 (3.9%)	107 (59.1%)	74 (40.9%)	
Hospitalized – Non-ICU		1082 (23.5%)	836 (77.3%)	246 (22.7%)	
Continuous Medical care		2150 (46.7%)	1833 (85.3%)	317 (14.7%)	
Home		1190 (25.9%)	1032 (86.7%)	158 (13.3%)	
Gender mismatch (Yes)	5058	2408 (47.6%)	1948 (80.9%)	460 (19.1%)	0.008
Female-to-male		1376 (27.0%)	1103 (80.2%)	273 (19.8%)	0.011
Male-to-female		1032 (20.0%)	845 (81.9%)	187 (18.1%)	0.602

Number of observations available [a] Continuous data are reported as median and interquartile range. Percentage counts refers to rows for data regarding the occurrence of 3-months graft loss in case of multiple (>2) categories. BMI, body mass index; GWBWR , graft weight to recipient body weight ratio; MELD, model for end stage liver disease; UNOS, United Network for Organ Sharing, ICU, Intensive Care Unit

Table 2. Graft failure according to time and center specific characteristics

	Total (N=5201)	**3-months Graft loss**		
		No (N=4288)	Yes (N=913)	*p value*
Transplant year				0.884
≤2006	1519	1262 (83.1%)	257 (16.9%)	
2007-2010	813	667 (82.0%)	146 (18.0%)	
2011-2013	1643	1353 (82.3%)	290 (17.7%)	
≥ 2014	1226	1006 (82.1%)	220 (17.9%)	
Center Volume[a]				0.103
1st quartile (≤3 LDLTs)	33	30 (90.9%)	3 (9.1%)	
2nd quartile (4-11 LDLTs)	130	98 (75.4%)	32 (24.6%)	
3rd quartile (12-62 LDLTs)	715	589 (82.4%)	126 (17.6%)	
4th quartile (>63 LDLTs)	4323	3571 (82.6%)	752 (17.4%)	
Early center cases [b]				0.049
Yes	822	658 (80.0%)	164 (20.0%)	
No	4379	3630 (82.9%)	749 (17.1%)	

.

[a] According to the total number of cases reported to the European Liver Transplant Registry for each center
[b] First 15 cases of LDLT for each center reported to the European Liver Transplant Registry for each center
LDLT, adult-to-adult living donor liver transplantation

Table 3. Univariable and multivariable analysis.

		Univariable analysis				Multivariable analysis		
	missing	OR	95% CI	*p value*	AUC	OR	95% CI	*p value*
Donor age (years)	128	1.02	(1.01, 1.02)	< 0.001	0.546	1.00	(0.99-1.01)	0.326
Donor gender (M)	142	0.76	(0.66, 0.88)	< 0.001	0.534	0.77	(0.53-1.11)	0.167
Donor BMI (kg/m^2)	621	1.03	(1.01, 1.05)	0.006	0.527	1.00	(0.97-1.03)	0.930
Type of graft (Left)	1205	1.08	(0.48, 2.33)	0.848	0.533	0.89	(0.29, 2.88)	0.854
Type of graft (Right)		0.34	(0.24, 0.48)	< 0.001		0.29	(0.10, 0.88)	0.022
Recipient gender (M)	1	0.87	(0.75, 1.01)	0.059	0.516	0.92	(0.68,1.24)	0.587
Recipient BMI (kg/m^2)	701	0.99	(0.97, 1.00)	0.127	0.523			
Recipient blood group AB	21	1.24	(0.91, 1.66)	0.170	0.531			
Recipient blood group B		1.35	(1.10, 1.66)	0.004				
Recipient blood group O		1.20	(1.02, 1.41)	0.031				
Acute liver Failure		2.97	(2.11,4.14)	< 0.001	0.521	0.87	(0.46,1.59)	0.657
Metabolic		0.99	(0.69,1.39)	0.962	0.503			
Hep C		1.00	(0.81,1.24)	0.974	0.502			
Hep B		0.69	(0.58,0.82)	< 0.001	0.546	0.85	(0.64-1.12)	0.261
Budd Chiari		1.86	(1.17,2.88)	0.007	0.513	1.62	(0.84-2.96)	0.127
Autoimmune		1.19	(0.80,1.72)	0.379	0.503			
Cholestasis		1.06	(0.80,1.37)	0.682	0.502			
Alcohol		1.21	(1.02,1.43)	0.031	0.516	1.18	(0.88-1.59)	0.249
HCC		0.80	(0.64,1.00)	0.049	0.512	0.92	(0.57,1.43)	0.725
Congenital biliary disease		1.50	(0.63,3.20)	0.317	0.501			
Other		0.91	(0.64,1.27)	0.592	0.502			
MELD score †	1443	1.04	(1.03, 1.05)	< 0.001	0.577	1.03	(1.01,1.04)	< 0.001
MELD-Na score†	1779	1.05	(1.04, 1.06)	< 0.001	0.602			
CHILD score†	1753	1.17	(1.12, 1.22)	< 0.001	0.597			
Child-class (B) †	1753	1.55	(1.14, 2.13)	0.006	0.572			
Child-class (C) †		2.28	(1.69, 3.13)	< 0.001				
Urgent transplant (Yes)	365	2.79	(2.05, 3.78)	< 0.001	0.525	1.14	(0.66, 1.91)	0.634
UNOS status (Hosp)	598	1.92	(1.54, 2.40)	< 0.001	0.590	2.55	(2.02, 3.21)	< 0.001
UNOS status (ICU)		4.52	(3.21, 6.34)	< 0.001		4.09	(2.44, 6.84)	< 0.001
UNOS status (Med)		1.13	(0.92, 1.39)	0.245				
Graft weight*****	1826	0.99	(0,99-0,99)	< 0.001	0.571			
GWBWR*****	1884	0.61	(0.42, 0.86)	0.006	0.544			
GWBWR < 0.8*****	1884	2.10	(1.57, 2.79)	< 0.001	0.533	1.57	(1.08-2.27)	0.001
ABO incompatibility (yes)	136	2.04	(1.16, 3.46)	0.010	0.505	1.95	(0.63, 5.40)	0.213
Gender mismatch	143	1.22	(1.05, 1.41)	0.008	0.524			
Female-to-male	99	1.23	(1.05, 1.44)	0.011	0.521	1.11	(0.71, 1.73)	0.656
Early center case	-	1.21	(1.00, 1.46)	0.047	0.513	0.72	(0.42, 1.18)	0.220

† * Given the evidence of multi-collinearity and redundancy between these variables, solely the MELD and GWBWR </> 0.8 were entered into the multivariable analysis. BMI, body mass index, MELD, model for end-stage liver disease; UNOS, United Network for Organ Sharing, ICU, Intensive Care Unit, GWBWR, graft weight to recipient body weight ratio

Table 4. Donor and recipient characteristics in the training and test set

	Training set (N=2060)	Test set (N=515)	p value
Donor age (years)	32.0 (26.0, 39.0)	32.0 (25.0, 39.0)	0.796
Donor gender (Female)	779 (37.8%)	211 (41.0%)	0.188
Donor BMI (kg/m^2)	24.5 (22.3, 27.0)	24.8 (22.4, 27.4)	0.595
Type of graft			0.134
Left	66 (3.2%)	10 (1.9%)	
LLS	14 (0.7%)	1 (0.2%)	
Right	1980 (96.1%)	504 (97.9%)	
Recipient Age	51.6 (41.7, 58.6)	51.8 (42.1, 58.9)	0.961
Recipient gender (Female)	672 (32.6%)	153 (29.7%)	0.205
Recipient BMI	25.8 (23.1, 29.1)	25.6 (23.0, 29.1)	0.609
Recipient ABO Group			0.519
A	910 (44.2%)	246 (47.8%)	
AB	150 (7.3%)	37 (7.2%)	
B	339 (16.5%)	80 (15.5%)	
O	661 (32.1%)	152 (29.5%)	
MELD score	16.0 (12.0, 21.0)	15.0 (12.0, 20.0)	0.278
MELD-Na score	18.6 (14.0, 24.2)	17.8 (14.0, 23.4)	0.216
CHILD score	9.0 (7.0, 11.0)	9.0 (7.0, 10.0)	0.182
CHILD class			0.582
A	325 (15.8%)	90 (17.5%)	
B	934 (45.3%)	234 (45.4%)	
C	801 (38.9%)	191 (37.1%)	
Urgent transplant (Yes)	113 (5.5%)	26 (5.0%)	0.695
Associated transplant (Yes)	13 (0.6%)	5 (1.0%)	0.408
UNOS status			0.542
At home with normal function	486 (23.6%)	113 (21.9%)	
Continuous hospitalization	508 (24.7%)	117 (22.7%)	
Intensive care unit-bound	96 (4.7%)	25 (4.9%)	
Continuous medical care	970 (47.1%)	260 (50.5%)	
GWBWR	1.1 (0.9, 1.3)	1.1 (0.9, 1.3)	0.608
GWBWR <0.8	145 (7.0%)	38 (7.4%)	0.788
ABO incompatible (Yes)	11 (0.5%)	4 (0.8%)	0.517
Gender mismatch	943 (45.8%)	236 (45.8%)	0.984
Female-to-male	525 (25.5%)	147 (28.5%)	0.158
Male-to-female	418 (20.3%)	89 (17.3%)	0.124
MELD score >24	357 (17.3%)	72 (14.0%)	0.068
Graft loss 3 months	348 (16.9%)	96 (18.6%)	0.348
Early center cases	104 (5.0%)	19 (3.7%)	0.196
Center Volume			0.853
1st quartile	58 (2.8%)	16 (3.1%)	
2nd quartile	425 (20.6%)	99 (19.2%)	
3rd quartile	495 (24.0%)	121 (23.5%)	
4th quartile	1082 (52.5%)	279 (54.2%)	

Continuous data are presented as median and interquartile range.
BMI, body mass index, MELD, model for end stage liver disease; UNOS, United Network for Organ Sharing, ICU, Intensive Care Unit, GWBWR, graft weight to recipient body weight ratio

Table 5. Coefficients of predictors selected by the LASSO logistic regression

Predictor	Coefficient
Graft weight	-0.020
Type of graft = Right	-0.26
UNOS Status = Hospitalized	0.25
UNOS Status = ICU	0.63
Child score	0.53
Lamba	0.03

UNOS, United Network for Organ Sharing, ICU, Intensive Care Unit

Table 6. Sensitivity and specificity of the simplified NNET model at different risk cut-offs.

Cut-off	Sensitivity	Specificity	PPV	NPV	LR+	LR-
1%	1,00	0,01	0,17	1,00	1,01	0,00
2%	1,00	0,05	0,18	1,00	1,05	0,00
3%	0,99	0,12	0,19	0,99	1,13	0,05
5%	0,97	0,25	0,21	0,97	1,29	0,13
10%	0,86	0,50	0,26	0,94	1,70	0,29
20%	0,64	0,76	0,36	0,91	2,66	0,48
30%	0,50	0,85	0,41	0,89	3,40	0,59
40%	0,41	0,90	0,47	0,88	4,25	0,66
50%	0,21	0,98	0,66	0,86	9,24	0,81
60%	0,14	0,99	0,77	0,85	16,02	0,87
70%	0,09	1,00	0,82	0,84	21,97	0,91
80%	0,06	1,00	0,95	0,84	96,11	0,94

PPV, positive predictive value, NPV, negative predictive value, LR+, positive likelihood ratio, LR-, negative likelihood ratio

FIGURES

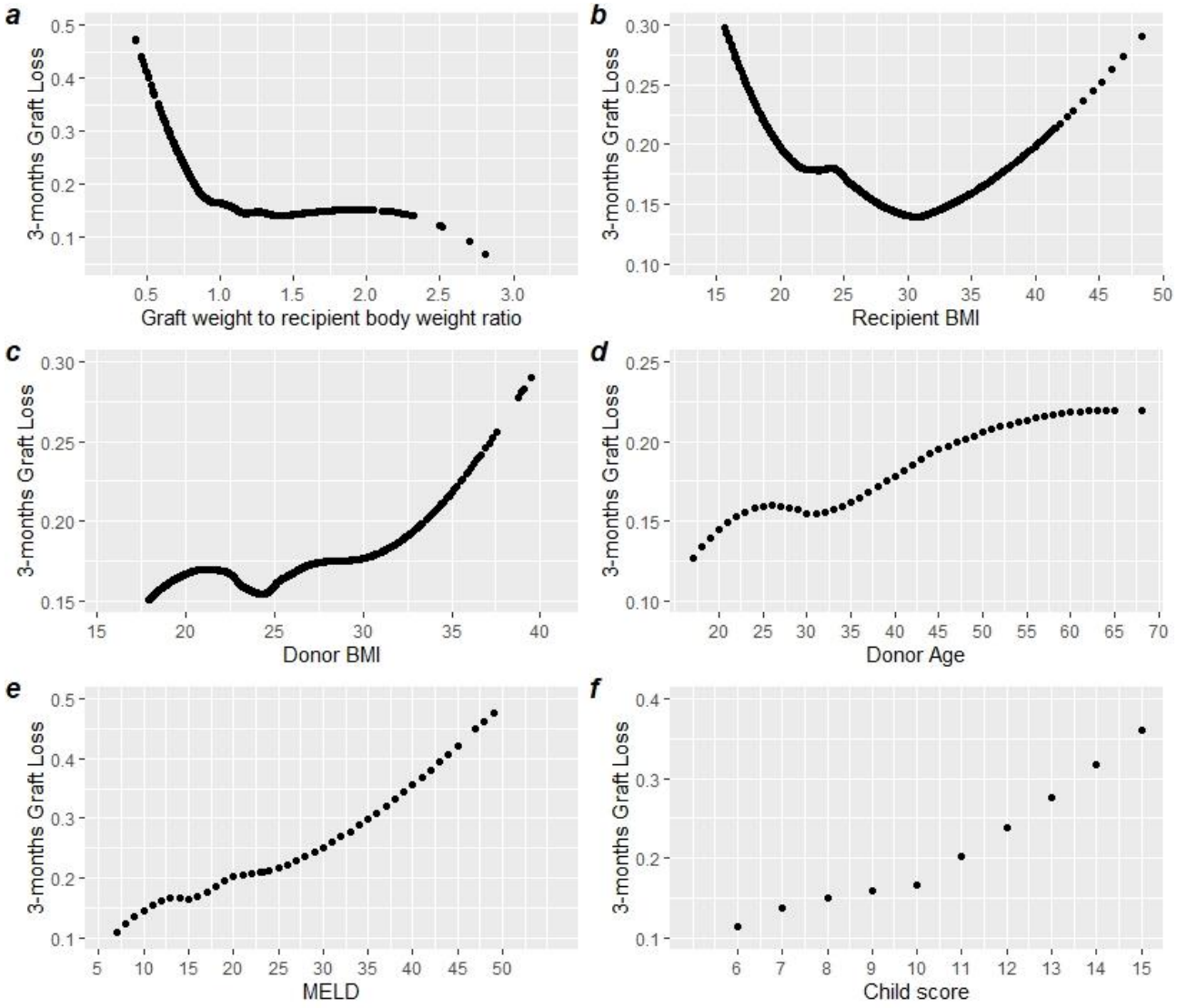


Figure 1. Plots showing the shape of the relation between continuous variables and the probability of early graft loss modelled using local weighted regression (LOESS).

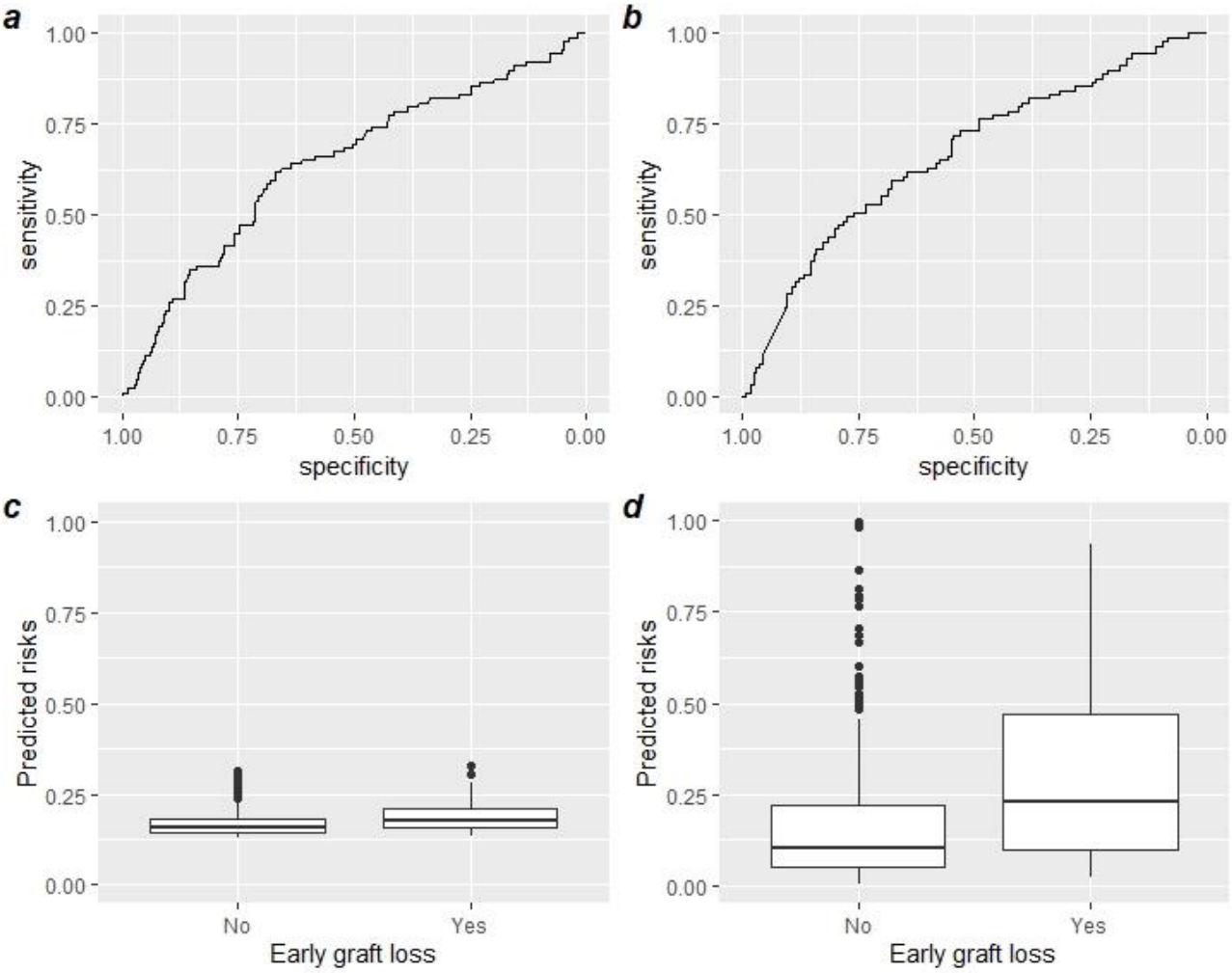
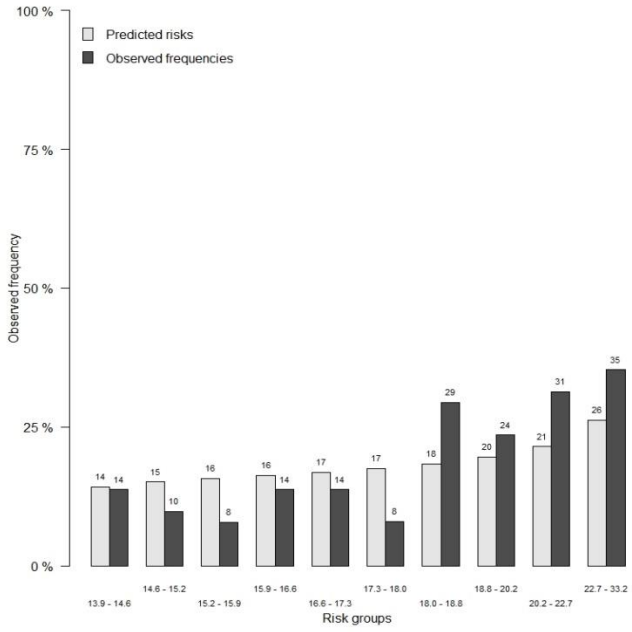
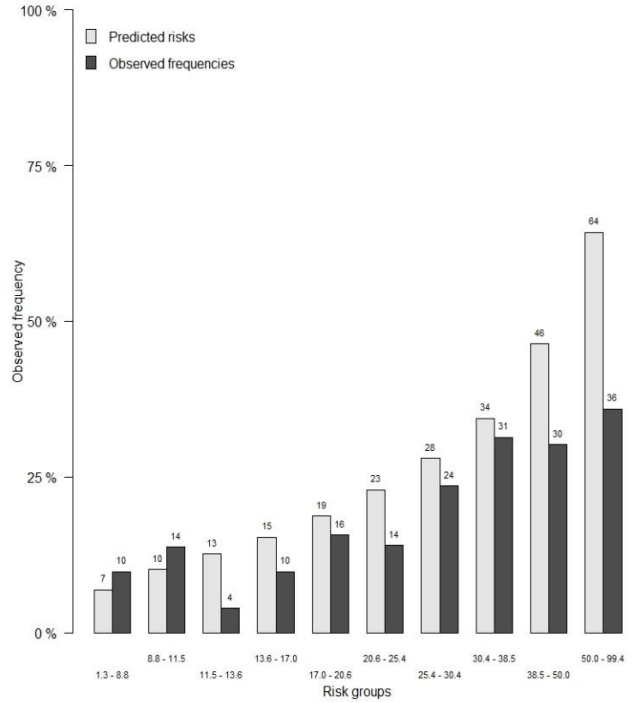


Figure 2. ROC curves showing the discrimantive accuracy of the LASSO regression model (a) and the NNET prediction model (b) on the test set. Box plots showing the distribution of predicted risk between LDLTs with and without occurrence of the early graft loss for the LASSO (c) and the NNET (d) prediction model.



a



b

Figure 3. Calibration plots comparing for the LASSO prediction model (a) and the NNET prediction model (b) the predicted vs observed frequencies of graft loss occurrence for different risk groups.

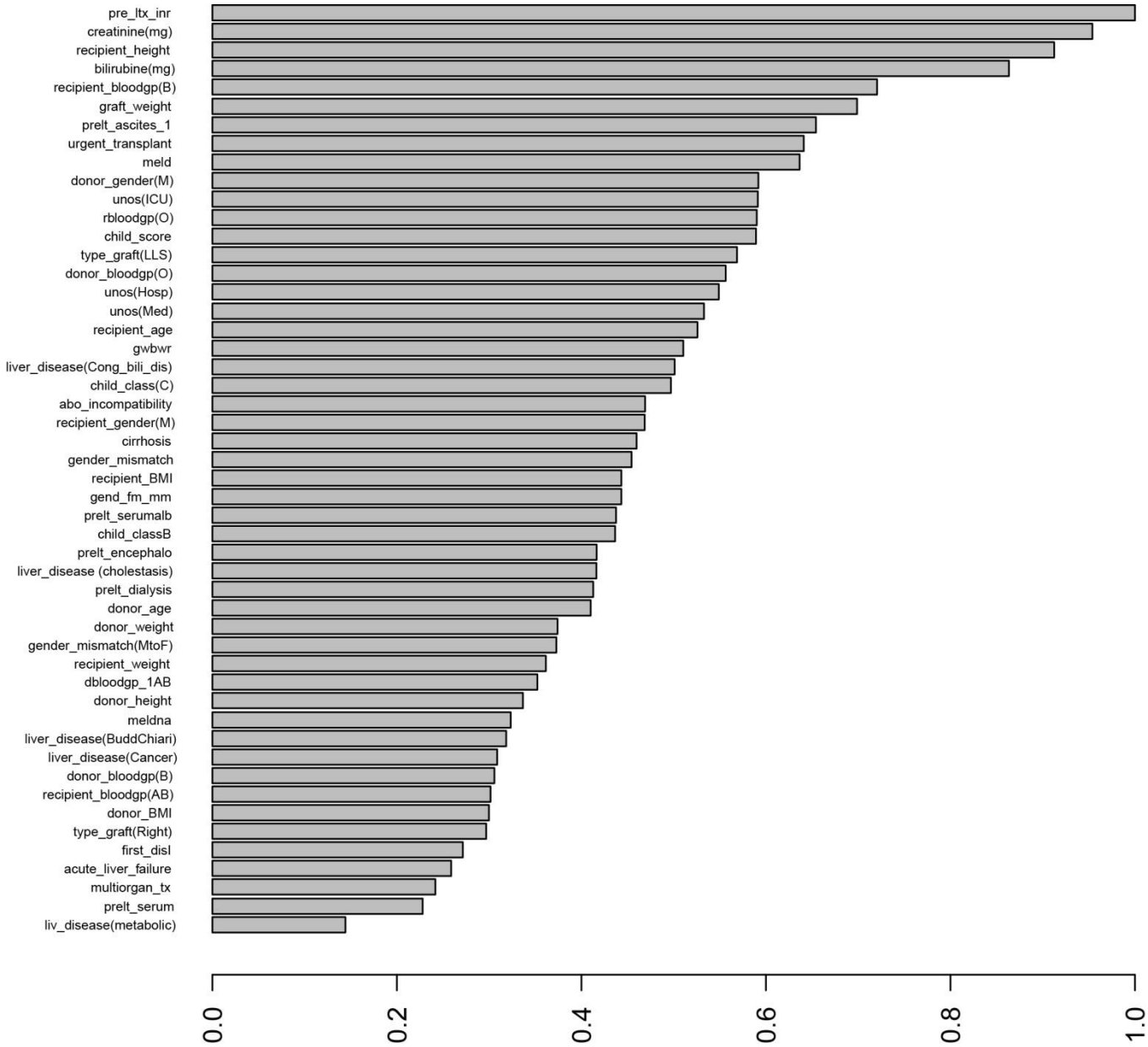


Figure 4. Features importance within the neural network model

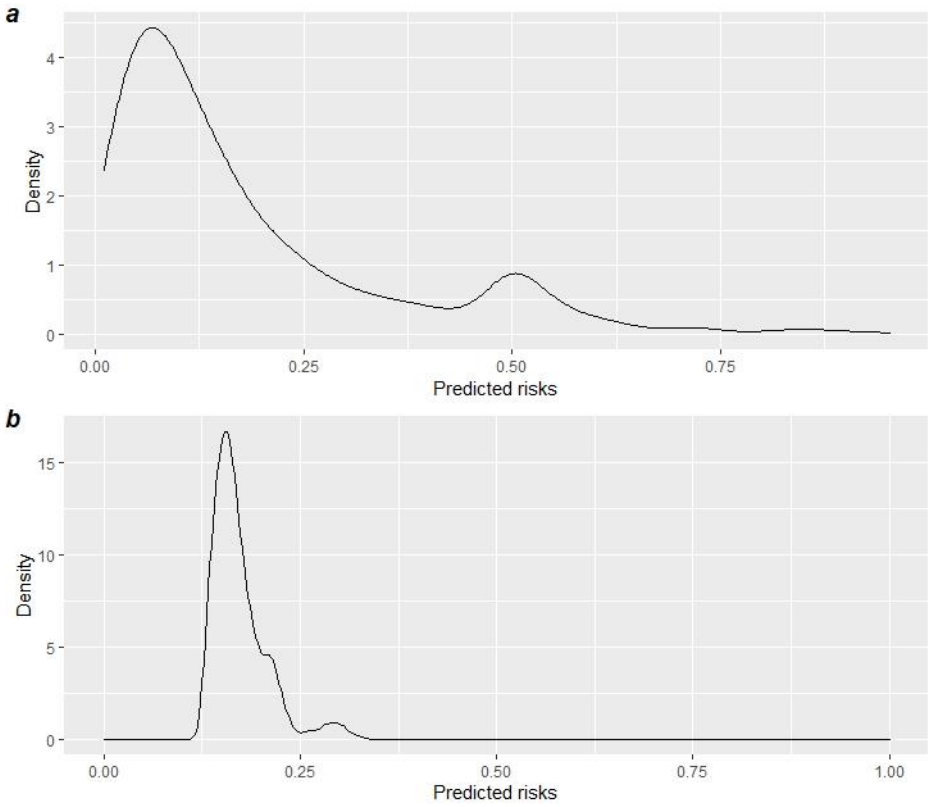


Figure 5. Density plots showing the distribution of the risks predicted by the NNET model (a) and the LASSO prediction model (b)

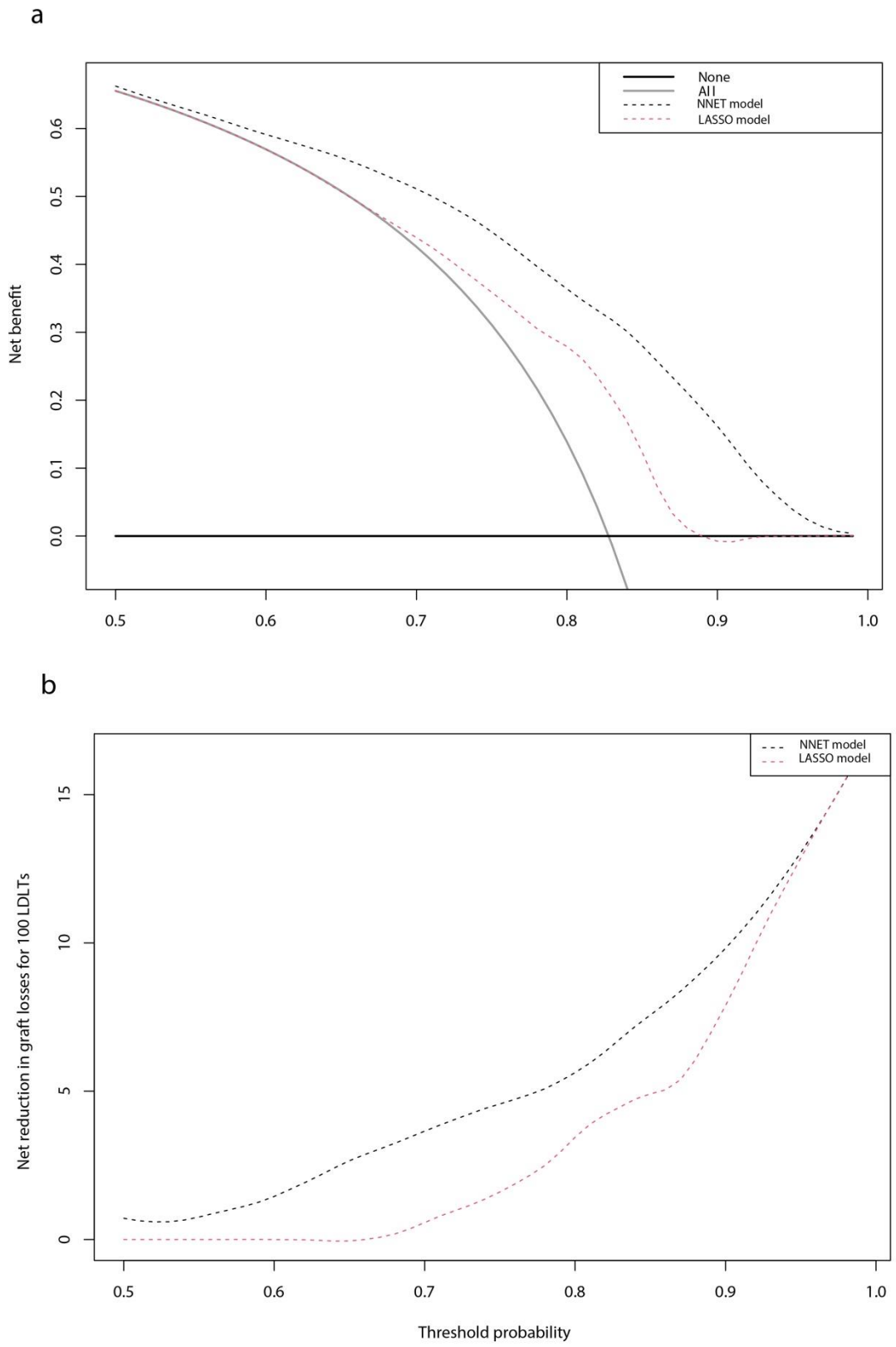


Figure 6. Decision curves showing the net benefit (a) and net reduction in graft losses for 100 LDTs resulting from the adoption of the NNET and LASSO prisk prediction models.

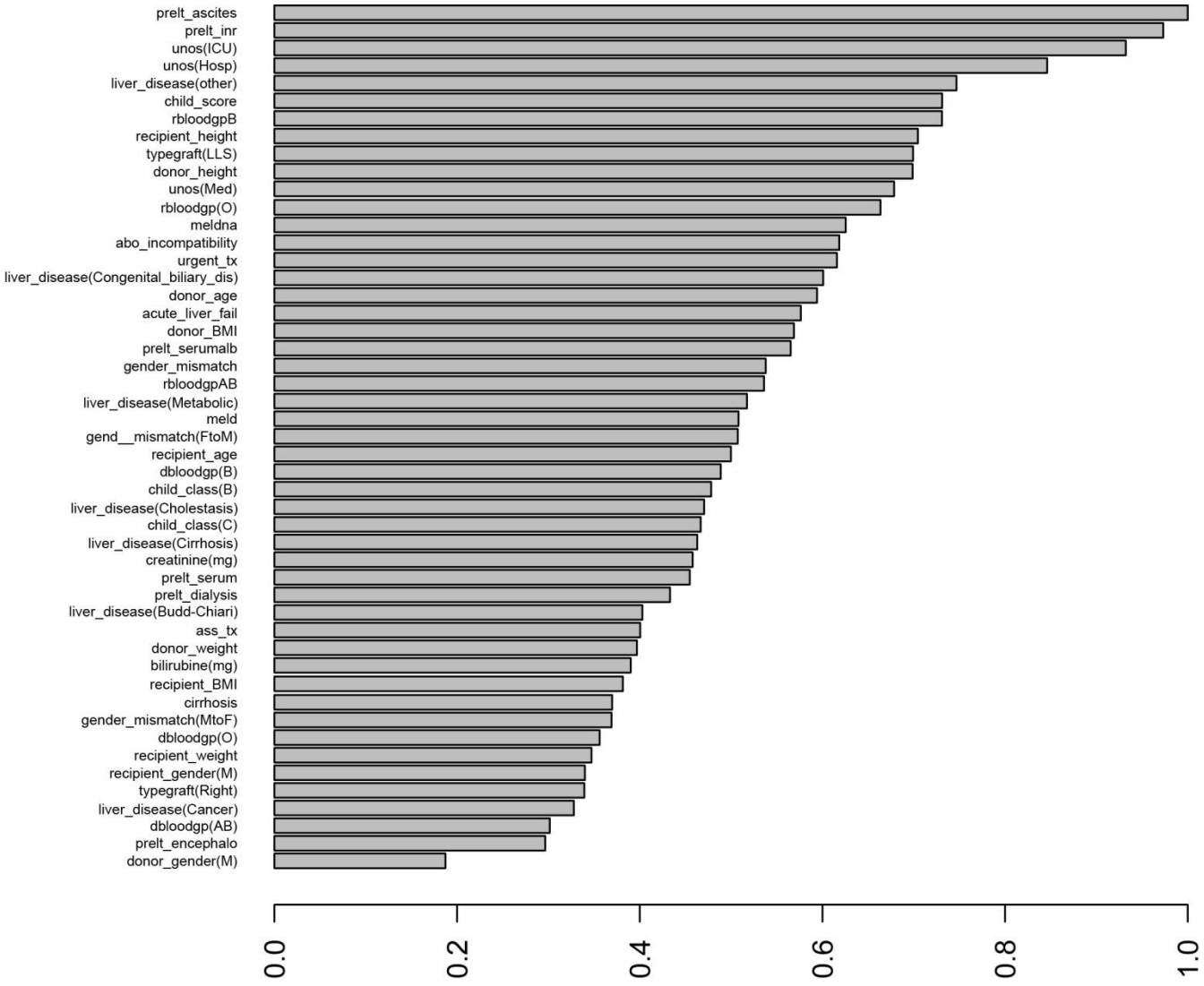


Figure 7. Variable importance within the simplified NNET prediction model

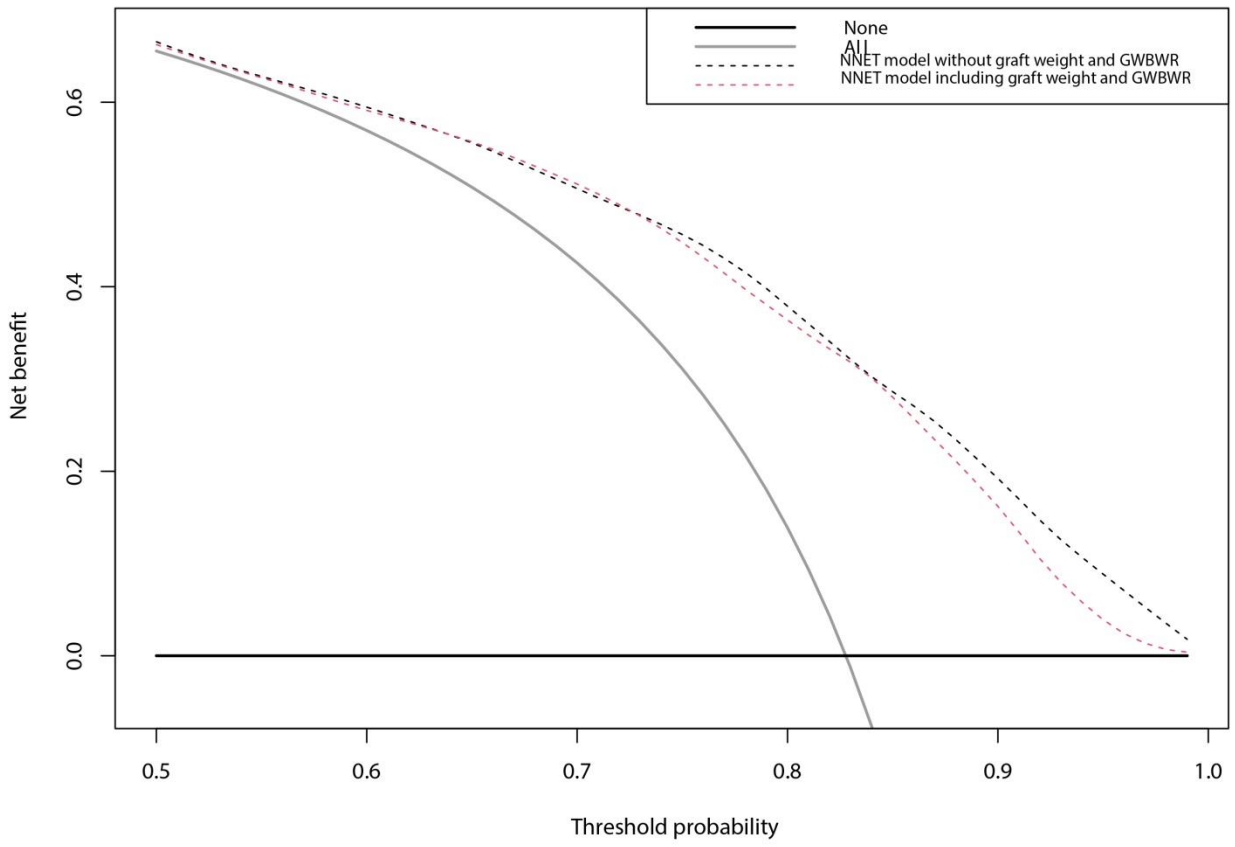


Figure 8. Decision curve comparing, at different threshold probabilities, the net benefit associated with the two developed NNET prediction models, including and not including the graft weight and the graft-to-recipient-body weight ratio as predictors.

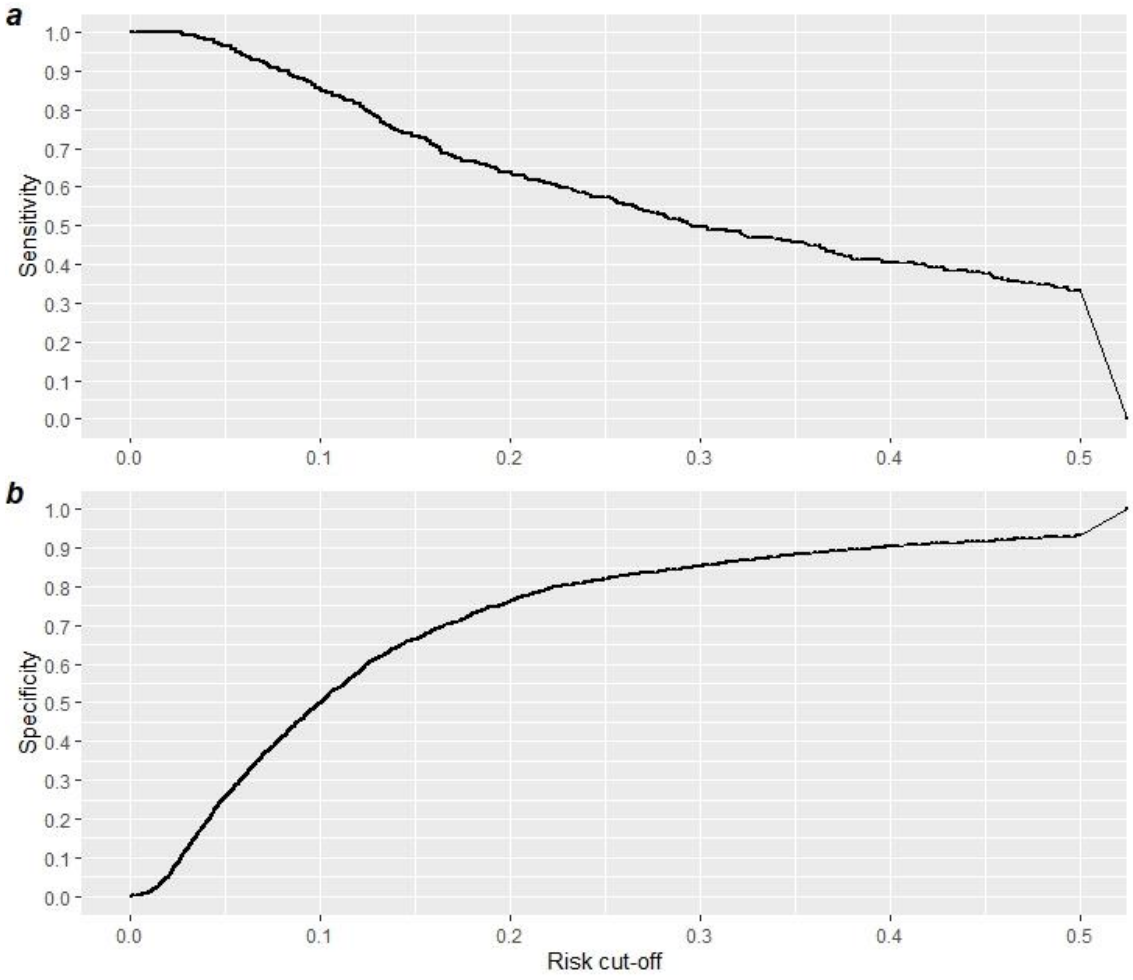


Figure 9. Sensitivity (a) and specificity (b) of the simplified neural ntetwork (not including graft weight and the graft-to-recipient-body weight ratio as predictors) risk prediction model at different risk-cut-offs.

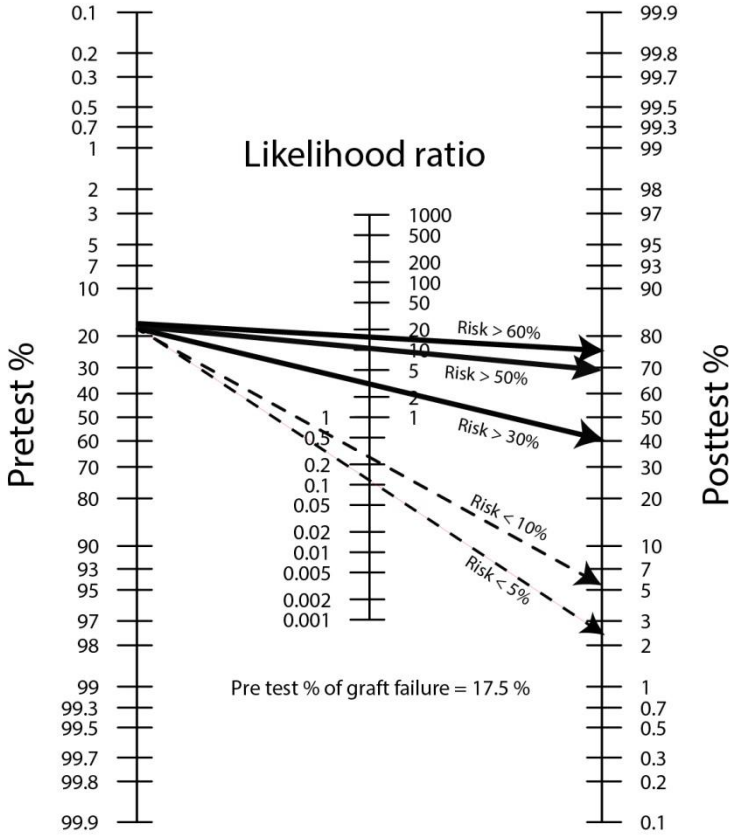


Figure 10. A fagan nomogram showing changes in the probability of early graft failure (3 months) after risk assessment with the simplified NNET model. Starting from a test probability of 17.5% (incidence of 3-months graft failure in the ELTR), predicted risks higher than 30%,50% and 60% are associated with individual post-test probabilities of graft failure at 3 months of 41.4%, 69% and 74.8%, respectively. On the contrary, predicted risks inferior to 10% and 5% are associated with individual post-test probabilities of graft failure at 3 months of 5.60% and 2.48%.