Development and validation of two-step algorithms to predict the outcome after kidney transplantation in the Eurotransplant region

Inaugural Dissertation

zur

Erlangung des Doktorgrades
philosophiae doctor (PhD) in Health Sciences
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von

Angela Johanna Ernst
aus Memmingen

Deutsche Zentralbibliothek für Medizin -
Informationszentrum Lebenswissenschaften
2025

Erstbetreuer:	Prof. Dr. rer. medic. Martin Hellmich
Zweitbetreuer:	PD Dr. med. Jan Ulrich Becker
Gutachterin/Gutachter:	Prof. Dr. med. Nicole Skoetz
	Prof. Dr. med. Christine Kurschat

Datum der Mündlichen Prüfung: 19.03.2025

Danksagung

Ich möchte mich bei allen Personen bedanken, die in irgendeiner Weise zu der in dieser Dissertation beschriebenen Arbeit beigetragen haben. Zwei Personen gilt zuallererst ein besonderer Dank: meinem Institutsleiter und akademischen Berater, Prof. Dr. rer. medic. Martin Hellmich, der mir nicht nur eine Anstellung gegeben, sondern mich auch zur Promotion motivert hat und dem Studienleiter und ebenfalls akademischen Berater, Jan Ulrich Becker, der das Projekt, in dem die Arbeit integriert wurde, überhaupt erst ins Leben gerufen hat. Während der gasamten Dauer hat Herr Hellmich zu einer lohnenden Erfahrung in der Graduiertenschule beigetragen, indem er mir intellektuelle Freiheit bei meiner Arbeit gewährte und mir die Teilnahme an zahlreichen Schulungen, Workshops und Konferenzen ermöglichte. Herr Becker hat nicht nur durch sein Networking die Kooperationspartner gefunden, mit deren Hilfe die benötigten Rohdaten zur Auswertung bereitgestellt werden konnten, sondern er hat selbst einen enormen Beitrag geleistet, indem er mit seiner nephropathologischen Expertiese Daten zur Auswertung generierte und Vorträge und Publikationen erarbeitete. Vielfach dankbar bin ich ihm zudem für den regelmäßigen Wissensaustausch, die Möglichkeit interdisziplinär mit Forschenden aus unterschiedlichen Instituten in Kontakt zu kommen und sein Vertrauen in meine Fähigkeiten. Zusätzlich möchte ich meinen Ausschussmitgliedern des IPHS, mit ihren damaligem Vorsitzenden Prof. Dr. Rudolf Wiesner, für ihr Interesse an meiner Arbeit danken. Mit ihrem Rat und fachlicher Expertiese geholfen haben auch meine beiden Tutorinnen Prof. Dr. med. Christine Kurschat und Prof. Dr. med. Nicole Skoetz, wofür ich Ihnen sehr dankbar bin. Jedes Ergebnis, das in dieser Thesis beschrieben wird, wurde durch die Hilfe und Unterstützung von Eurotransplant, vertreten durch Dr. Ineke Tieken, und den Nephrologen, Chirurgen und Labormitarbeitern der beteiligten Studienzentren, ermöglicht. Insbesondere danke ich Prof. Sandrine Florquin vom Amsterdam Medical Center, Prof. Dr. med. Dirk Ludger Stippel vom Universitätsklinikum Köln, Dr. med. Wolfgang Arns vom Krankenhaus Köln-Merheim, PhD Dr. med. Nika Kojc der Universität von Ljubljana, PhD Dr. med. Marian Clahsen-van Groningen vom Erasmus Medical Center in Rotterdam, Prof. Dr. med. Heinz Regele vom Allgemeinen Krankenhaus Wien, Prof. Danica Galesic Ljubanovic von der Universität Zagreb, Dr. med. Christos Chatzikyrkou vom PHV-Dialysezentrum Halberstadt, Dr. med. Florian Scurt vom Universitätsklinikum Magdeburg und Francesco Vasuri und Deborah Malvi vom IRCCS Azienda Ospedaliero-Universitaria di Bologna. Nicht zu vergessen war die Unterstützung meiner Kolleginnen und Kollegen vom IMSB die geduldig auch Vorträgen gelauscht und bei zusätzlichen Arbeiten geholfen haben.

Einen besonderen, herzlichen Dank gilt meiner ehemaligen Kollegin und Unterstützerin, PhD-Absolventin und Diplom-Statistikerin Christina Samel, die einen scharfen, wissenschaftlichen Blick auf die Dissertationsschrift hatte. Zuletzt danke ich meiner Familie die, unersetztbar, in jeder Lebenslage, mit offenen Herz mir Rückhalt und ein Zuhause schenkt; ohne sie wäre all die Arbeit gewiss nicht möglich gewesen.

Zusammenfassung

Der Mangel an Spendernieren in Eurotransplant (ET) Ländern ist unumstritten, weshalb die Entscheidung zur Verwerfung eines Organs nicht ohne driftigen Grund getroffen werden sollte. Eine Evidenzbasis, insbesondere hisichtlich der Rolle von nephropathologischen Einschätzungen, fehlt bislang. Mit zusätzlicher pathologischer Evaluation verlängert sich die kalte Ischämiezeit, zudem ist sie nicht in allen Transpalantationskliniken bzw. -zentren verfügbar.

Ziel dieser Arbeit war es daher, flexible klinisch-pathologische Scores zur Vorhersage der verzögerten Transplantatfunktion (DGF) und des einjährigen, todeszensierten Transplantatverlusts (1y-tl) nach Nierentransplantation zu entwickeln, zu validieren und zu vergleichen. Da das Modellierungsverfahren von der Menge vorausgewählter, potenzieller Prädiktoren abhängt, bestand ein zweites Ziel darin, fünf statistische und auf maschinellem Lernen basierende Variablenauswahlverfahren anzuwenden, und die Ergebnisse hinsichtlich der Vorhersagekraft der abgeleiteten Modelle zu vergleichen.

Im Hinblick auf die Erstellung eines guten Vorhersagemodells waren die Methoden des maschinellen Lernens den klassischen statistischen Modellen bei der Auswahl geeigneter Prädiktoren nicht überlegen. Es wurden zwei zweistufige Scores entwickelt, welche, mit Einschränkungen, die derzeitige Praxis in sechs Eurotransplant Ländern (Österreich, Belgien, Kroatien, Deutschland, Ungarn und die Niederlande) widerspiegeln. Um auf die verlängerte kalte Ischämiezeit bei der nephropathologischen Evaluation und einer mangelnden Verfügbarkeit dieser einzugehen, wurden Scores mit optionaler Nephropathologie entwickelt. Bezüglich der Fähigkeit, Fälle von DGF zu dikriminieren war der 2-Stufen Score den Modellen von Irish, Balaz und Chapal nicht unterlegen. Ebenso wenig wie der 2-Stufen Score für 1y-tl den von Snoeijs, Port, De Vusser und Miller.

Die entwickelten, flexiblen 2-Stufen Scores sollten gute Ergebnisse für die klinische Praxis in Eurotransplant liefern und komplexeren Scores nicht unterlegen sein. Sie können und sollten jedoch für den zunehmenden Einsatz von Perfusionspumpen und die Spende nach Herztod, sowie bei der Risikogruppe von Spendern mit Nieren von marginaler Qualität, angepasst und validiert werden.

Abstract

The shortage of donor kidneys in Eurotransplant (ET) countries is undeniable, which is why the decision to discard an organ should not be made without solid reasoning. An evidence base, particularly with regard to the role of nephropathogical assessments, is still lacking. With additional pathological evaluation the cold ischaemia time is prolonged, and it is not available in all transplantation clinics and centers. The aim of this study was therefore to develop flexible clinico-pathological scores for the predict delayed graft function (DGF) and one-year death-censored graft loss (1y-tl) after kidney transplantation, to validate and compare them.

As the modelling procedure depends on the set of pre-selected potential predictors, a second aim was to apply five statistical and machine-learning variable selection methods and compare results in terms of predictive discriminative power of the derived models. In terms of generating a good prediction model, machine learning methods were not superior to classical statistical models in selecting adequate predictors.

Two two-step algorithms were developed which, with limitations, reflect the current practice in six ET countries (Austria, Belgium Austria, Belgium, Croatia, Germany, Hungary and the Netherlands). In order to address the prolonged cold ischaemia time in nephropathological evaluation and a lack of availability of these, scores with optional nephropathology were developed. With regard to the ability to discriminate cases of DGF the 2-Step scores was not inferior to the models of Irish, Balaz and Chapal. Neither was the 2-Step scores for 1y-tl inferior to that of Snoeijs, Port, De Vusser and Miller.

The flexible 2-Step scores developed should provide good results for clinical practice in Eurotransplant and should not be inferior to more complex scores. However, they can and should be adapted for the increasing use of perfusion pumps and donation after cardiac death, as well as the risk group of donors with marginal quality kidneys.

Contents

Acronyms

000 MM zero HLA-A, -B, and -DR mismatches [acronym by Eurotransplant]

1y-tl Death-censored transplant loss within one year (3y-tl, 5y-tl, 10y-tl within 3, 5 and 10 years, respectively)

AB0 AB0 blood group rules

ABMR antibody-mediated rejection

AIC Akaike Information Criterion

aka also known as

AKI acute kidney injury

AM Acceptable Mismatch

ANN Artificial Neural Network

ANZDATA the Australian and New Zealand Dialysis and Transplant Registry

ATI acute tubular injury

ATN acute tubular necrosis

AUC area under the receiver operating curve

BIC Bayesian Information Criterion

Cal calibration

CART Classificationd And Regression Trees

CC complete case

CHARMS Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies

CIT cold ischaemia time

CMV cytomealovirus

CRISP-DM Cross Industry Standard Process for Data Mining

DBCD donation after brain death followed by circulatory death

DBD donation after brain death

DCD donation after circular death

DDKT deceased donor kidney transplant

DFG Deutsche Forschungsgemeinschaft

DGF Delayed graft function

Disc discrimination

DIVAT French: Données Informatisées et VAlidées en Transplantation; English: computerized and validated data in transplantation

DSO Deutsche Stiftung Organtransplantation

ECI estimated calibration index

eGFR estimated glomerular filtration rate

EN Elastic Net

EPP events per predictor

EPV events per variable

ESOT European Society for Organ Transplantation

ESP Eurotransplant Senior Program

ESRD end stage renal disease

ET Eurotransplant

ETKAS Eurotransplant Kidney Allocation System

Ext external

Fail-set Training data set for outcome 1y-tl

Fail-val Validation data set for outcome 1y-tl

FSGS focal segmental glomerulosclerosis

GB Gradient Boosting

GDPR General Data Protection Regulation

HCVAB hepatitis C virus antibody-positive

HL test Hosmer-Lemeshow test

HLA human leucocyte antigens

HMP hypothermic machine perfusion

hr hour

HU high urgency

Int internal

KAOO kidney after other organ

krt kidney replacement therapy

LASSO Least Absolute Shrinkage and Selection Operator

LR Logistic Regression

MAR missing at random

MCAR missing completely at random

MI multiple imputation

mice Multivariate Imputation by Chained Equations

min minute

ML machine-learning

MNAR missing not at random

MPR median P rule

No number of

NPV negative predictive value

NRI net reclassification improvement

NRP normothermic regional perfusion

OLS ordinary least squares

OPTN Organ Procurement and Transplantation Network

Ov overall

PICO Participants, Intervention, Comparison, Outcome

PPV positive predictive value

PROBAST Prediction model Risk Of Bias Assessment Tool

RefMod Reference regression model

RF Random Forest

RI regression imputation

RoB Risc of Bias

ROC receiver operating curve

SI single imputation

SVM Support Vector Machine

TCMR T cell-mediated rejection

TMA thrombotic microangiopathy

UKRR United Kingdom Renal Registry

UNOS United Network for Organ Sharing

VI variable importance

VIF variance inflation factor

VSURF Variable Selection Using Random Forests

WSI whole slide image

XGB eXtreeme Gradient Boosting

y year

List of figures

List of tables

Chapter 1

Introduction

1.1 Background

According to the European Renal Association, the prevalence among European citizens (excluding Germany) in 2021 requiring kidney replacement therapy (krt) due to chronic kidney disease was 1 per 1000 [1]. Based on data from statutory health insurers, prevalence was in Germany with 1.2 per 1000 even higher [2]. For patients with end-stage renal disease (ESRD), of whom most are in demand of dialysis, a successful transplantation can significantly improve quality of live and the probability of survival [3, 4].

Within the European Union, Eurotransplant is responsible for the national and transnational allocation of deceased donor kidneys, livers, hearts, lungs, pancreas, and intestine in a consortium of eight hereafter referred to as "ET countries" (Germany, Austria, Belgium, Croatia, Hungary, Luxembourg, The Netherlands, and Slovenia). As a a non-profit service organization, its objective is to “ensure optimal use of available donor organs” by providing a common allocation system [5]. In 2021, Eurotransplant reported that 1897 deceased donors were eligible for organ allocation [1] of which 1573 (83%) were kidney donors; about 50% (N=794) of all donors were from Germany [6, page 10]. However, by the end of 2021, there were still 10269 (Germany: 6593; 64%) patients on the Eurotransplant kidney waiting list [6, page 23].

The demand for kidneys from both living and deceased donors is a major public health challenge, especially as it is expected to increase in the future due to demographic changes. Furthermore, organs from older donors are associated with more marginal functionality and availability of donor kidneys varies at national and also regional levels, posing an additional challenge to allocation systems [7, 8, 9]. Allocation itself can follow several principles which were categorized into four fundamental val-

[1] "Eurotransplant does not report on the number of Actual donors, where an organ has been recovered for the purposes of transplantation, but not necessarily transplanted. The number of Actual donors is slightly higher than the number of Utilized donors"[6, page 10]

ues by medical ethicists [10]. The utility on balance includes, among others, recipient survival and potential harmful consequences such as delayed graft function (DGF), a common early complication, and premature transplant loss.

Although its definition varies amongst users, "most define DGF as an acute kidney injury (AKI) that occurs in the first week after transplantation and requires a dialytic treatment" [11]. As a consequence, patients often require longer hospital stays with higher healthcare resource utilization and an increased risk of further complications like hospital-acquired infections and acute rejection [12, 13]. In long-term, DGF was found to be associated with decreased graft survival and graft function [14], also with recipient death [15, 16]. Hence, DGF has been an important endpoint to predict during the allocation process.

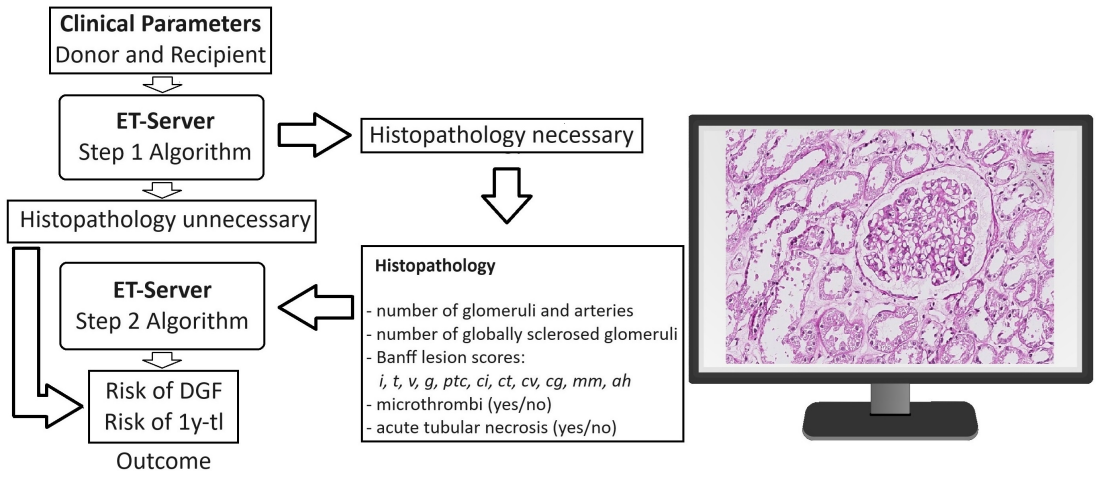
Several research groups have examined predictors of DGF and developed predictive models. So far, outside the Eurotransplant consortium predictive scores were developed in North America (Irish, 2003 and 2010 [17]; Schold, 2005 [18]; Jeldres, 2009 [19]; Balaz, 2013 [20]; Kawakita, 2020 [21]), France (Chapal, 2014 [22]), Italy (Zaza, 2015 [23]), The United Kingdom (Moore, 2007 [24]) and China (Ding, 2018 [25]; Chen, 2020 [26]; Xue, 2021 [27]; Wang, 2022 [28]; Zhao, 2022 [29]) [30]. Decruyenaere used data from the Ghent University Hospital, Belgium, to develop and compare 9 distinct predictive regression and machine-learning models in 2015 and published a second paper in the same year, combining existing scores (Irish, 2010, and Zaza, 2015) in a meta-model [31, 32]. In the same year, Chaumont et al. published a regression-based model derived from Eurotransplant data aiming to identify factors that led to a decrease in the incidence of DGF in Belgium over the last three decades [33]. In 1997, the European Multicenter Study Group used data from the Eurotransplant registry to explore risk factors for DGF in cadaveric kidney transplantations comparing single and multi-organ donors [34].

Apart from DGF, 1-year death-censored transplant loss (1y-tl) is another hitherto less well-regarded short-term outcome discussed in the literature. Two popular scores based on data from the USA (Port, 2002 [35]; Rao, 2009 [36]) have already considered this endpoint. Trained on data from Eurotransplant, Miller developed and published online available models in 2023 based on Cox regression to predict risk of transplant loss and recipient survival at a user-selected time point after transplantation [37].

The scores already developed differ not only in terms of the outcomes studied and the cohorts used for training, but also in terms of the influencing variables considered. Scores can base on clinical, composite clinical-histological or histological variables available before transplantation. The mentioned scores by Jeldres [19] and Zaza [23] are clinical; Wang [28] and Balaz [20], however, also include histological information from procurement "harvest" biopsies performed before the period of cold ischaemia. Miller's models are also based exclusively on clinical donor data, but

with optional recipient data. Histological findings are not taken into account because these are not regularly recorded by Eurotransplant. To be useful in a prediction model, histological assessment needs to be done in a standardised way. In practice, this is not necessarily the case. Mayor differences are related to the timing (procurement vs. preimplantation), type (wedge, vs. needle vs. skin punch) and procession method (frozen vs. paraffin-embedded) [38]. Further, the experience and availability of the on-call nephropathologist evaluating and scoring the biopsy is another source of variability one usually can not account for [38, 39].

Reproducibility of procurement biopsy findings was shown to be poor and the benefit in the ability to predict the transplant outcome is controversial, although inter-rater reliability among specialised nephropathologist was better on procurement biopsies [40, 41, 42, 43]. Nevertheless, renal pathology offered as a "round-the-clock service" is a scarce, costly resource in ET countries. Since clinical data relating to donor, recipient and transplant procedure alone allow at least for a rough prediction of both endpoints of interest, DGF and death-censored transplant loss within one year, two flexible scores with optional histology are proposed that can be set up at the ET-server as shown schematically in figure 1.1 [44]. Such a 2-step approach, allowing the allocating physician to decide, whether to perform a histological evaluation to improve accuracy of prediction while taking into account an possible increase in cold ischaemia time, is new.



Abbreviations: DGF, delayed graft function; 1y-tl, death-censored transplant loss within the first year; ET, Eurotransplant

Figure 1.1: 2-Step algorithm with optional histology 1.1[44]

The steps to generate a predictive, medical score can be adapted from the **CR**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (CRISP-DM) reference model, which represents a six-point circle: business understanding (what do we need?), data understanding (collection, description, exploration, quality assessment), data preparation (selection, cleaning, integration, formatting), modelling, evaluation (success criteria,

review) and deployment [45]. The data understanding and preparation phases are in general most time consuming. Given the available data while taking into account applicable data protection guidelines like the EU's General Data Protection Regulation (GDPR, see: https://gdpr-info.eu), following these steps should provide the most meaningful predictors to be included into the modelling step.

Statistical and machine-learning (ML) methods can support the data preparation by reducing the number of candidate predictors without loss of information by providing information on variable importance [46, 47, 48]. However, there's no guideline on which method is best with regard to sample size, number of predictors and missing values, but there are recommendations on variable selection [49]. Therefore, five methods for variable selection prior to actual score generation are applied and the results are compared with regard to the selected predictors and the models' performance. Additionally, the ability of the final 2-Step scores to discriminate recipients at increased risk was assessed on both the training and an independent validation dataset and compared with established, more and less simple scores. This provides additional external validation of the other scores.

In a population with an increasing prevalence of ESRD, the expectation of a shrinking pool of kidneys from deceased donors exerts increasing pressure on the health systems and in particular on those who make the decision on the acceptance or rejection of a donor organ. In order to make the best use of this valuable resource and support clinical decision makers such as transplant surgeons and nephrologists, the first aim of this project was to develop a regression-based algorithm that estimates the individual probabilities for the transplant complication DGF. The second endpoint to be predicted before transplantation, 1y-tl, was chosen to avoid the increasing influence of factors beyond the quality of the donor organ, such as recipients post-transplantation behaviour and the transplantation procedure itself which cumulate over the years after transplantation. As a histological evaluation in addition to a macroscopic or donor clinical history based evaluation of the donors kidney quality is neither available nor necessary in every case, the algorithms should be designed to incorporate clinical variables with optional histology. Ultimately, evidence-based scores should be created that can be integrated into the Eurotransplant database and updated with new data in real time, as suggested by Donna Ankerst et al.[50], among others, to assist in the decision to accept or reject an organ.

1.2 Structure

Before providing detail about how the scores were created, I will first give an overview of existing, established scores for predicting DGF and, respectively, transplant loss within the first year after transplantation together with the allocation system as it is currently used by Eurotransplant (Chapter 2). Next, the methodological background section describes the statistical and machine learning methods for a) candidate predictor preselection and b) two-step modelling in Chapter 3. In Chapter 4, the assessment of the training and validation datasets is described. The results of the variable selection and modelling are presented, supported by tables and figures, in the results Chapter 5. A comprehensive discussion of the derived scores, the strengths and limitations of the study and the importance of scores as a tool to support informed decision-making are then presented in the discussions Chapter 6. To conclude, future prospects are presented.

1.3 Sponsors

The thesis is based on the project "Entwicklung eines Scores für die Qualitätsbeurteilung von Verstorbenenspendernieren in Eurotransplant" (BE 3801/2-1), which was funded by the Deutsche Forschungsgemeinschaft (DFG) [Engl.: German Research Foundation] and sponsored by the University Hospital of Cologne in collaboration with Eurotransplant.

1.4 Software used

The statistical data analysis of the Eurotransplant, as well as the supplementary histological data, which were both provided as Excel files, was performed exclusively with the software R of the Comprehensive R Archive Network (cran.r-project.org) in version 4.4.2 for Windows. The online Latex editor from the Digital Science UK Limited "Overleaf" (de.overleaf.com) was used to create the dissertation. During the writing process, the AI-based translator DeepL Pro Write (www.deepl.com) was used for translation and spelling correction in English.

Chapter 2

Published scores, evidence on nephropathological findings and the Eurotransplant allocation system

Scores and algorithms for predicting outcomes after deceased donor transplantation have been modeled for more than 20 years. Various short- and long-term outcomes are considered. Short-term outcomes include delayed graft function within 7-8 days after transplantation, transplant loss, reduced graft function, and death within the first year. In the long term, treatment adherence, transplant loss and reduced kidney function, measured by estimated glomerular filtration rate (eGFR), and patient death are of interest. As mentioned above, the scores have been developed and adapted in different *regions* like US-America, Brazil, Europe (Eurotransplant, The UK, France), China, and Oceania. In addition, the independent variables considered may vary with regard to donor-, recipient-, transplant-specific predictors with and without histology. There are other ways to categorize existing scores or the target group for which the score is to be applied:

- ***Modelling method***: classical statistical, based on binary logistic regression (LR), Cox Regression, elastic net regression (EN) versus machine and deep learning like support vector machines (SVM), random forests (RF), gradient boosting (GB), extreme gradient boosting (XGB) and artificial neural networks (ANN) versus meta-regression modelling.
- ***Data source***: (nationwide) registries like Eurotransplant, the US United Network for Organ Sharing (UNOS) and the Organ Procurement and Transplantation Network (OPTN), which is operated under contract with the U.S. Department of Health and Human Services, the Australia and New Zealand Dialysis and Transplant registry (ANZDATA) and the UK Renal Registry (UKRR), which was established by the UK Renal Association, CRISTAL, which is the French

organ transplant registry [51], Scandiatransplant [52] vs. clinic networks like the French Données Informatisées et VAlidées en Transplantation (DIVAT) [English translation: computerized and validated data in transplantation] vs. single or multiple transplant centres.

- ***Donor and recipient characteristics***: brain vs. cardiovascular deceased vs. living donors, adult vs. pediatric donors or recipients, high vs. marginal risk recipients, good vs. marginal quality kidneys.

- ***Model presentation***: regression estimates including intercept or baseline hazard function, nomograms, point-scores, trees and networks. The more complex the associations between outcomes and predictors, as well as within predictors are, the harder to interpretate are they. Algorithms derived by deep learning and support vector machines, as well as regression trees are therefore usually not published in papers but can be applied via online tools they're linked to.

The following is an overview of existing scores for the outcomes DGF and one-year (death-censored) transplant loss. Only those that were modelled with but not solely for adult recipients with deceased donors are considered. Results are partly from a systematic literature review conducted as part of a Cologne Fortune project "Development of a Meta Score for the Quality Assessment of Kidneys from Deceased Donors" [Project Nr. 268/2022 (account: 7103-9713-0005-01)].

Independent reviewers were A. Ernst and PD Dr. med. J. U. Becker. The search and PICO (Participants, Intervention, Comparison, Outcome) terms used are found in the supplemental Chapter 7.

2.1 Scores and algorithms predicting delayed graft function

A systematic literature search identified 30 publications published since 1998 on predictive models for the outcome DGF that were not explicitly generated for minor recipients or living donors. As a template for data extraction and risk of bias assessment of prediction models, the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and Prediction model Risk Of Bias Assessment Tool (PROBAST) were used [53]. The derived CHARMS for DGF are described in tables 2.1, 2.2 and 2.3. Registries and networks were used as the *data source* for 6 (UNOS/OPTN, N=5; DIVAT, N=1) publications, while the remaining data came from single (N=18) or multiple clinics or transplant centres. The relatively free access to registry data such as UNOS for research purposes allows regular recalculations and recalibrations of the prediction models, which is reflected in the large proportion of models from the USA (N=8). Ten models came from China, five from Europe (France, Italy, Belgium, Czech Republic), and one each from Canada and Brazil. In two cases (Konieczny, 2021 [54] and Moore, 2007 [24]), it was not clear where the data came from. Although regression analysis is considered to be the 'classic' *modelling method* for deriving predictive models, artificial neural networks were already used to predict DGF in the early work by Shoskes, 1998 [55], and Brier, 2003 [56]. Logistic regression was the main method of analysis in 21 papers, but was also used in five others to be compared with machine learning algorithms. In addition, the PROBAST tables on variable selection and modelling are presented in Tables 2.4, 2.5 and 2.6. As discussed in more detail in Section 3.3.10, *validation* is an integral part of assessing the quality of a predictive model. Internal validation by cross-validation, bootstrapping or a randomly selected subset of the data was performed in 15 publications, and the performance of nine other models was measured on the full training dataset. Nine models were externally validated, although other authors or groups of authors may have validated the models after publication. Validation information is missing completely in three publications (Santos, 2023[57]; Wang, 2022 [28]; Losappio, 2014 [58]). The most commonly used measure to assess discrimination, with 25 publications, was the area under the receiver operating curve (ROC), AUC, with graphical presentation; information on calibration and overall performance was significantly less common (N=12 and N=6 publications, respectively).

Table 2.1: Study characteristics of published models and scores predicting DGF [part 1]

Author, Year	Study design	Enrolment period	Study setting	Study region	Participant characteristics				
					DGF = yes N (%)	CIT (hr or min)	Age, years (donor)	No. of previous transplants >= 1	Age, years (recipient)
Brier, 2003 [56]	Retrospective cohort	09.1984 to 03.1995	We tested the hypothesis that artificial neural networks can predict the development of DGF following kidney transplantation. (...) We also employed traditional logistical regression analysis to construct a model to predict ARF that also identified individual clinical factors that correlated with ARF.	Jewish Hospital, Louisville, KY (USA)	45	White: 1350.15 Black: 1385.39	NA	NA	White: 40.11 Black: 39.67
Konieczny, 2021 [54]	Retrospective cohort	2012 to 2018	In our work, we verified the factors that may be responsible for the occurrence of DGF with the use of information technology, particularly the branch of artificial intelligence called machine learning.	NA	60 (38.2)	20.29 (6.63)	46.35 (13.68)	12 (26.8)	51.26 (12.74)
Bae, 2020 [59]	Existing registry	01.01.2005 to 31.12.2017	To better understand the possible role of ML in transplantation, we aimed to evaluate the performance of ML algorithms in a "common" study setup relevant to a wide gamut of transplantation research.	USA (OPTN)	26794 (27.4)	17 (11.7-23)	41 (25, 52)	9.9	54 (44, 63)
Irish, 2003 [60]	Existing registry	1995 to 1998	In the study presented here, we used the data available from more than 13846 adult cadaveric renal transplant recipients recorded in the United States Renal Data System (USRDS) to develop an index, or nomogram, that quantifies the likelihood of DGF after renal transplantation by using both donor and recipient factors known before transplantation.	United States Renal Data System (USRDS)	3780 (23.7)	20.4 (8.5)	34.1 (16.9)	1926 (13.9)	45.7 (12.8)
Moore, 2007 [24]	Prospective cohort	2003 to 2006	Retrospective data extraction. Donated kidneys were perfused with Marshall's solution and stored on ice until transplantation. No machine perfusion was used. Frusemide 200 mg intravenously was administered upon release of the vascular clamps. No mannitol was administered.	A multi-ethnic European urban population.	75 (35.7)	NA	NA	23 (15.2)	NA
Jeldres, 2009 [61]	Retrospective cohort	1979 to 2004	We decided to revisit the topic of DGF risk prediction and to test whether the concept developed by Irish and colleagues can be used in a different transplant population	Montréal Health Centre, Canada	103 (19.4)	16.8 (17)	34 (33)	NA	43 (43)
Irish, 2010 [62]	Existing registry	01.01.2003 to 31.12.2006	Data obtained from United Network for Organ Sharing / Organ Procurement and Transplantation (UNOS/OPTN) on adult ($\geq$ 16 years) recipients of a solitary, nonpreemptive, nonmachineperfused, deceased donor kidney	USA	around 6255 (25.7)	17.8 $\pm$ 7.8	36.6 $\pm$ 16.7	3001 (12.3)	49.8 $\pm$ 13.7
Balaz, 2013 [63]	Retrospective cohort	01.2005 to 12.2009	The aim of this study was to determine the association between a histopathologic score and clinical variables with the risk for DGF and to identify factors that may better predict the risk of DGF in ECD-positive donors.	Czech Republic	120 (35)	18.02 (3.73)	53 (17-76)	NA	50 (20-79)
Chapal, 2014 [64]	Existing registry	2007 onwards	Adult (>=18 years) recipients of isolated, non preemptive, non machineperfused, deceased donor kidneys, who were prospectively collected since January 2007 and computerized in the DIVAT (www.divat.fr, N1 CNIL 891735 version 2, August 2004) databank were included in the study.	France	468 (25.4)	19.21 (7.02)	51.91 (15.6)	311 (25.1)	51.92 (13.19)
Zaza, 2014 [39]	Retrospective cohort	1984 to 2012	To achieve (...), we retrospectively analyzed main demographic and clinical features, follow-up events and outcomes registered in a large dedicated dataset including 2,755 patients compiled collaboratively by four Italian renal/transplant units.	Italy	774 (28)	NA	NA	125 (4.5)	48 (12)
Decruyenaere, 2015 [31]	Retrospective cohort	01.01.2005 to 31.12.2011	In this study, the goal is therefore to analyze and discuss the performance of different modeling techniques in the prediction of DGF and to identify which method is most suited to the task at hand.	Ghent University Hospital, Belgium	62 (12.5)	14.19 (4.328)	42.6 (14.77)	NA	52.8 (11.68)
Decruyenaere, 2015 [32]	Retrospective cohort	01.01.2005 to 31.12.2011	In this study, the goal is therefore to analyze and discuss the performance of different modeling techniques in the prediction of DGF and to identify which method is most suited to the task at hand.	Ghent University Hospital, Belgium	62 (12.5)	14.19 (4.328)	42.6 (14.77)	NA	52.8 (11.68)

Table 2.2: Study characteristics of published models and scores predicting DGF [part 2]

Author, Year	Study design	Enrolment period	Study setting	Study region	Participant characteristics				
					DGF = yes N (%)	CIT (hr or min)	Age, years (donor)	No. of previous transplants >= 1	Age, years (recipient)
Ding, 2017 [65]	Retrospective cohort	01.12.2011 to 31.10.2016	In this study, the aim was to develop a simple donor risk score model that could be readily applied by physicians to evaluate the quality of DCD kidneys and identifies kidneys at the highest risk of DGF before DCD.	The First Affiliated Hospital of Xi'an Jiaotong University, China	55 (15.2)	8.1 (4.5)	41.2 (13.8)	0 (0)	36.2 (10.3)
Sun, 2018 [66]	Retrospective cohort	02.2012 to 12.2015	A multi-center, retrospective, observational cohort study was carried out involving 383 kidney transplants from DBCD	Affiliated Hospitals of three Universities, China	74 (19.3)	5.5 (2.2)	40.4 (13)	9 (2.3)	42.5 (11.1)
Chen, 2020 [67]	Retrospective cohort	01.01.2018 to 31.12.2019	Considering the contradiction between marginal kidneys and DGF, this retrospective study was conducted based on the medical records of our center to investigate pretransplant risk factors for DGF.	Second Xiangya Hospital of Central South University, China	78 (16.9)	12.35 (3.86)	40.91 (19.77)	0 (0)	37.83 (10.50)
Zheng, 2020 [68]	Retrospective cohort	01.2018 to 09.2019	In this study, the correlation of donor patient parameters, kidney pre-implant pathology Remuzzi scores and HMP parameters were analyzed collectively, instead of individually, to enrich the comprehensive evaluation of donor kidney quality.	First Affiliated Hospital of Xi'an Jiaotong University, China	46 (13.8%)	8.7 $\pm$ 3.4	50.8 $\pm$ 12.5	0 (0)	36.2 $\pm$ 9.3
Kawakita, 2020 [69]	Existing registry	01.01.2007 to 31.05.2012	The patient cohort consisted of all DDKT patients, including single organ and simultaneous multiple organ transplants, pre-emptive and non-preemptive transplants, and machine-perfused and non-machine perfused kidneys.	USA	13792 (25.1)	17.73 (9.65)	38.77 (16.45)	NA	52.76 (13.02)
Luo, 2022 [70]	Retrospective cohort	01.2015 to 12.2019	The use of artificial intelligence (AI) is increasing explosively in the medical field. Deep learning methods, such as convolutional neural networks (CNNs), are very practical for image and audio analysis. Inspired by these ideas, we aimed to explore whether the combination of kidney biopsy whole-slide images (WSIs) and clinical features can provide more information for transplant outcome prediction.	Third Affiliated Hospital of Sun Yat-sen University, China	25 (11.4) RGF: 92 (42)	8.2 (6.9, 10.2)	44 (35, 51.5)	6 (3)	40 (33.5, 48.5)
Xue, 2021 [71]	Retrospective cohort	11.2011 to 03.2018	The present study attempted to collect and analyze data from different parts of China to preliminarily establish a quantitative and representative model for DGF prediction after renal transplantation and guidance for evaluation of deceased donors.	China	479 (13.3)	6.2 (3.2)	38.9 (16.8)	NA	41.7 (13.3)
Wang, 2022 [28]	Retrospective cohort	01.2012 to 06.2017	After combining and screening the 16 pathological changes, we finally developed a simple and easy-touse dichotomous (mild-to-severe) grading criterion for donor allograft ATI. We intend to clarify the discriminative and predictive ability of this grading criterion for the short-term and long-term prognoses by retrospective analysis.	First Affiliated Hospital of Sun-Yat Sen University, China	28 (20)	11 [7-14]	32.5 (15.1)	NA	42.8 (12.9)
Zhao, 2022 [72]	Retrospective cohort	10.2020 to 10.2021	Here, we discuss the involvement of serum markers in renal function impairment, construct a novel, reliable and highly accurate prediction model combine with serum markers, aiming to provide new insights for the DGF prediction after kidney transplantation.	First Affiliated Hospital of Nanchang University, China	23 (32)	9 (7-11)	44 (17 - 48)	0 (0)	39.97 (11.49)
Pan, 2021 [73]	Retrospective cohort	01.01.2017 to 31.12.2019	Our purpose is to help clinicians induce defense and treatment according to the patient's DGF risk level.	First Affiliated Hospital of Anhui Medical University	29 (13.7)	NA	45.1 (16.3)	0 (0)	39.9 (9.9)
Shoskes, 1998 [55]	Retrospective cohort	01.09.1994 to 31.12.1996	We wished, therefore, to construct a neural network to predict the occurrence of DGF in our cadaveric renal transplant population and test it with prospective data to test its validity and applicability.	Harbor-UCLA Medical Center and St. Mary Medical Center	24 (24)	NA	NA	NA	NA

Table 2.3: Study characteristics of published models and scores predicting DGF [part 3]

Author, Year	Study design	Enrolment period	Study setting	Study region	Participant characteristics				
					DGF = yes N (%)	CIT (hr or min)	Age, years (donor)	No. of previous transplants >= 1	Age, years (recipient)
Costa, 2020 [74]	Retrospective cohort	01.01.2015 to 31.12.2017	This study aimed to evaluate the risk factors for DGF, including in the analysis donor maintenance-related (DMR) variables, which were thoroughly investigated from multidisciplinary records.	Brazil	235 (53)	NA	NA	36 (8.1)	44.2 (14.7)
Losappio, 2014 [58]	Retrospective cohort	between 1997 and 2007	Although some studies support the policy that no kidneys should be discarded, providing that pre-implantation kidney biopsy is adequate, Cecka et al. suggested that pretransplant kidney biopsy increases the risk to discard acceptable kidneys. (...) In the attempt to address this issue, the aim of the present study was to investigate the influence of pretransplant biopsy scoring on long-term SKT graft function and survival in a large, single-center cohort of transplant recipients receiving both standard and ECD grafts.	University of Bari Transplant Center, Italy	about 74 (20)	NA	NA	32 (8.6)	45.3
Santos, 2023 [57]	Existing registry	12.2007 to 06.2015	Our objective was to identify consistent predictors of multiple adverse outcomes of adult deceased donor (DD) kidney transplant recipients (KTRs) of varying sensitization status.	USA	NA	NA	NA	NA	NA
Qiao, 2021 [75]	Retrospective cohort	01.01.2019 to 31.08.2019	The purpose of this study was to establish a prediction model combing HMP parameters with perfusate biomarkers, as well as to evaluate its efficacy in predicting DGF and renal function recovery after renal transplantation.	First Affiliated Hospital of Xi'an Jiaotong University, China	20 (17.7)	NA	NA	0 (0)	NA
Chaumont, 2015 [33]	Retrospective cohort and registry	25.08.1983 to 30.06.2014	Our aim was to answer four questions: (1) What are the factors behind the decrease in DGF incidence in our cohort across the three past decades? (2) What role does rejection play in DGF occurrence? (3) What are the risk factors for DGF taking rejection into account or not? (4) Does DGF impact patient and transplant outcome differently with and without rejection?	Erasmus Hospital Brussels, Belgium and Eurotransplant	382 (21.4)	NA	NA	NA	NA
Jen, 2021 [76]	Retrospective cohort	01.01.2010 to 31.12.2018	All deceased donor kidney transplants with known recipient DGF status at our institution from January 1, 2010, to December 31, 2018, were initially included in this study.	California, USA	562 (27.2)	27.1 (11.1)	34 (19.3)	NA	NA
Ding, 2018 [77]	Retrospective cohort	01.09.2012 to 31.08.2016	The objectives of this study were to use a readily available HMP variable to design a scoring model that could identify the highest risk of DGF and provide guidance and advice for organ allocation and DCD kidney assessment.	First Affiliated Hospital of Xi'an Jiaotong University, China	38 (15.6)	7.8 (4.5)	41 (13.7)	0(0)	36.5 (10.5)

Abbreviations [part 1 to part 3]: CIT, cold ischaemia time; hr, hour; min, minute; No., number of; NA, not applicable; DDKT, deceased donor kidney transplant; DCD, donation after circulatory death; DBCD, donation after brain death followed by circulatory death; HMP, hypothermic machine perfusion

The observed AUC values vary less with the method of analysis than with the ratio of the number of predictors used to the number of observations in the cohort. Internally validated, but with a high (N=38, N=18, N=24) number of predictors in the model, Konieczny's random forests, Costa's neural network and Decruyenaere's linear SVM from Decruyenaere achieved values of up to .92, .886 and .84, while the neural network from Brier with 10 predictors only achieved .668. Fewer than 500 transplantations were performed in all four studies. Bae and Kawakita, who used UNOS datasets with 97787 and 55044 observations respectively, provided regional external validation of the results. Regression and machine learning methods with 39 and 26 predictors, respectively, yielded similar AUC values (Bae: LR .721, GB .723, RF .717; Kawakita: LR .728, RF .735, EN .728, XGB .742, ANN .737). Results between regression and machine learning models were also comparable in the paper by Brier and Decruyenaere. Neural networks performed significantly better in a direct comparison in Costa, although the final models contained a different number of predictors. Four publications that used only logistic regression were externally validated. Here, a similar picture emerges: models based on homogeneous cohorts and a small number of centres resulted in AUCs ranging from .846 to .89, while the models based on register data from Irish in 2003 and 2010 yielded AUCs of .703 and .704. The AUCs of the models that were exclusively internally validated and trained on data from a few centres varied considerably, although here, too, the ratio of the number of predictors to events seems to correlate positively with the AUCs.

Table 2.4: Modelling methods of published models and scores predicting DGF [part 1]

Author, Year	Modelling method	Sample size	Events n (%)	No predictors Cand.	No predictors Final	EPV or EPP	Selection of candidate predictors	Selection of final predictors	Number (%) and handling of missing data	Type of validation	Performance measures
Brier, 2003	Regression and Machine Learning	198	45 (22.7)	10	3	4.5	Stepwise selection	Forward selection	n (%): 0 (0.0) Method: Complete-case analysis	Int: Random split data Ext : None	Cal: Not evaluated Disc : AUC graph Ov: Not evaluated
Costa, 2020	Regression and Machine Learning	443	235 (53.0)	27	18	7.7	prior knowledge	No information	n (%): Unkown Method: Median imputation	Int: Random split data Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph Ov: Not evaluated
Bae, 2020	Regression and Machine Learning	97 787	26794 (27.4)	39	38	687.0	All available predictors	None	n (%): 0 (0.0) Method: Multiple imputation	Int: None Ext : Geographical	Cal: Calibration plot Disc : C-Statistic Ov: Brier score
Irish, 2003	Generalized additive models	13 846	3780 (27.3)	19	17	198.9	Unclear	Unclear	n (%): 0 (0.0) Method: Complete-case analysis	Int: None (Apparent performance) Ext : Temporal	Cal: Not evaluated Disc : C-Statistic / AUC graph / Sensitivity and speificity for different AUROC thresholds Ov: Not evaluated
Moore, 2007	Logistic regression	210	75 (35.7)	31	6	2.4	All available predictors	Backward elimination	n (%): Unkown Method: No information	Int: None (Apparent performance) Ext : None	Cal: Not evaluated Disc : Not evaluated Ov: R-squared
Jeldres, 2009	Logistic regression	532	103 (19.4)	11	6	9.4	No information	Backward elimination	n (%): 0 (0.0) Method: Complete-case analysis	Int: Bootstrap Ext : None	Cal: Calibration plot Disc : C-Statistic Ov: Not evaluated
Irish, 2010	Logistic regression	24 337	6255 (25.7)	20	19	312.8	prior knowledge	Generalized additive models (GAMs)	n (%): 0 (0.0) Method: Missing data of continuous variables (CVs) were accounted for in the model by creating two indicator variables (IV1 and IV2): If CV = missing then IV1 = 1or 0 otherwise; and If CV = missing then IV2 = 0 otherwise IV2 = CV. For categorical data, missing or unknown were grouped into the 'no' or 'absence' category.	Int: None (Apparent performance) Ext : Temporal	Cal: Calibration plot Disc : C-Statistic / AUC graph Ov: Likelihood ration chi-square
Balaz, 2013	Generalized linear model with logit link	344	126 (36.6)	22	7	5.7	univariable associations	None	n (%): Unkown Method: Unclear	Int: Random split data Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph Ov: Not evaluated
Chapal, 2014	Logistic regression	1 238	468 (37.8)	23	5	20.3	prior knowledge	In a multivariate model variables were retained if their corresponding P-value was lower than 0.05.	n (%): 0 (0.0) Method: Complete-case analysis	Int: Random split data Ext : None	Cal: Calibration plot / HL test Disc : C-Statistic / AUC graph / Positive and negative predictive valueas according to different threshold values Ov: Not evaluated

Table 2.5: Modelling methods of published models and scores predicting DGF [part 2]

Author, Year	Modelling method	Sample size	Events n (%)	No predictors Cand.	No predictors Final	EPV or EPP	Selection of candidate predictors	Selection of final predictors	Number (%) and handling of missing data	Type of validation	Performance measures
Zaza, 2014	Logistic regression	2 755	774 (28.1)	8	5	96.8	prior knowledge	Pre-specified model (not selection)	n (%): Unkown Method: Single imputation	Int: None (Apparent performance) Ext : None	Cal: The incidence of DGF was plotted against deciles of the score, and a non-parametric trend test was used to analyze the significance of the association between deciles of the prediction score and incidence of DGF Disc : C-Statistic / AUC graph Ov: Not evaluated
Decruyenaere, 2015	Logistic regression	497	62 (12.5)	55	20	1.1	The full feature set is iteratively pruned by removing the feature with the lowest importance until the 10-fold stratified cross-validation score decreases significantly	Pre-specified model (not selection)	n (%): 0 (0.0) Method: Single imputation	Int: Cross-validation Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph / Sensitivity, PPV Ov: Not evaluated
Decruyenaere, 2015	Logistic regression	497	62 (12.5)	55	20	1.1	The full feature set is iteratively pruned by removing the feature with the lowest importance until the 10-fold stratified cross-validation score decreases significantly	Pre-specified model (not selection)	n (%): 0 (0.0) Method: Single imputation	Int: Cross-validation Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph / Sensitivity, PPV Ov: Not evaluated
Ding, 2017	Logistic regression	543	55 (10.1)	22	8	2.5	univariable associations	Stepwise selection	n (%): Unkown Method: No information	Int: Random split data Ext : None	Cal: HL test Disc : C-Statistic / AUC graph Ov: Not evaluated
Sun, 2018	Logistic regression	383	74 (19.3)	12	4	6.2	Unclear	Unclear	n (%): Unkown Method: Complete-case analysis	Int: None Ext : Temporal	Cal: Not evaluated Disc : C-Statistic / AUC graph Ov: Not evaluated
Chen, 2020	Logistic regression	461	78 (16.9)	11	6	7.1	univariable associations	Pre-specified model (not selection)	n (%): 0 (0.0) Method: Complete-case analysis	Int: Cross-validation Ext : Temporal	Cal: Calibration plot / HL test Disc : C-Statistic / AUC graph Ov: Not evaluated
Zheng, 2020	Logistic regression	333	46 (13.8)	8	4	5.8	prior knowledge	Unclear	n (%): 0 (0.0) Method: Complete-case analysis	Int: None (Apparent performance) Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph / Sensitivity (0.804); Specificity (0.805) Ov: Not evaluated
Kawakita, 2020	Logistic regression	55 044	13792 (25.1)	126	19	109.5	prior knowledge	Pre-specified model (not selection)	n (%): Unkown Method: Complete-case analysis	Int: Cross-validation Ext : Temporal	Cal: Calibration plot / HL test Disc : C-Statistic Ov: Brier score
Teixeira, 2020	Logistic regression	159	85 (53.5)	19	NA	3.9	univariable associations	None	n (%): 0 (0.0) Method: Complete-case analysis	Int: None Ext : None	Cal: Not evaluated Disc : Not evaluated Ov: Not evaluated
Xue, 2021	Logistic regression	3 599	479 (13.3)	9	6	53.2	Prior knowledge and univariable associations	Pre-specified model (not selection)	n (%): 0 (0.0) Method: Complete-case analysis	Int: None (Apparent performance) Ext : Temporal	Cal: Calibration plot / HL test Disc : C-Statistic / AUC graph Ov: Not evaluated
Wang, 2022	Logistic regression	140	28 (20.0)	8	2	3.5	prior knowledge	No information	n (%): Unkown Method: Complete-case analysis	Int: None Ext : None	Cal: Not evaluated Disc : Not evaluated Ov: Not evaluated

Table 2.6: Modelling methods of published models and scores predicting DGF [part 3]

Author, Year	Modelling method	Sample size	Events n (%)	No predictors Cand.	No predictors Final	EPV or EPP	Selection of candidate predictors	Selection of final predictors	Number (%) and handling of missing data	Type of validation	Performance measures
Zhao, 2022	Logistic regression	71	23 (32.4)	32	4	0.7	prior knowledge	No information	n (%): Unkown Method: Complete-case analysis	Int: Bootstrap Ext : Temporal	Cal: Calibration plot / Slope / HL test / Calibration diagram Disc : C-Statistic / AUC graph Ov: R-squared / Brier score
Losappio, 2014	Logistic regression	372	74 (19.9)	8	4	9.3	prior knowledge	No information	n (%): Unkown Method: Complete-case analysis	Int: None Ext : None	Cal: Not evaluated Disc : Not evaluated Ov: Not evaluated
Santos, 2023	Logistic regression	62037 [41310/ 11093/9634]	NA	16	16	NA	prior knowledge	Pre-specified model (not selection)	n (%): Unkown Method: Complete-case analysis	Int: None Ext : None	Cal: Not evaluated Disc : Not evaluated Ov: Not evaluated
Qiao, 2021	Logistic regression	113	20 (17.7)	9	3	2.2	prior knowledge	Forward selection	n (%): 0 (0.0) Method: Complete-case analysis	Int: None (Apparent performance) Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph / Sensitivity (83.3%); Specifity (79.5%) Ov: Not evaluated
Chaumont, 2015	Logistic regression	1 784	382 (21.4)	24	7	15.9	All available predictors	Backward elimination	n (%): Unkown Method: Complete-case analysis	Int: None Ext : None	Cal: HL test Disc : C-Statistic / AUC graph Ov: Not evaluated
Jen, 2021	Logistic regression	2 117	562 (26.5)	21	21	26.8	All available predictors	None	n (%): 28 (1.3) Method: Complete-case analysis	Int: Unclear Ext : None	Cal: Not evaluated Disc : C-Statistic / Accuracy, Sensitivity, Specificity, PPV, NPV Ov: Not evaluated
Ding, 2018	Logistic regression	244	38 (15.6)	6	3	6.3	All available predictors	Stepwise selection	n (%): 0 (0.0) Method: No information	Int: Random split data Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph Ov: Not evaluated
Konieczny, 2021	Machine learning techniques	125	46 (36.8)	38	NA for NN, RF varying	1.2	prior knowledge	Other	n (%): Unkown Method: No information	Int: Random split data Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph Ov: Not evaluated
Pan, 2021	Machine learning techniques	211	28 (13.3)	40	15	0.7	univariable associations	LASSO selection	n (%): 0 (0.0) Method: Complete-case analysis	Int: Cross-validation Ext : None	Cal: Calibration plot / HL test Disc : C-Statistic / AUC graph Ov: Not evaluated
Shoskes, 1998	Machine learning techniques	100	24 (24.0)	12	NA	2.0	Other	Analyzed using a back-propagating neural network software package, Neuralyst (Cheshire Engineering, Pasadena, Calif). The number of hidden layers and nodes per layer were optimized using an automatic genetic supervisor. The model was then applied prospectively on the next 20 cadaveric renal transplants performed field)	n (%): 0 (0.0) Method: Complete-case analysis	Int: None (Apparent performance) Ext : Temporal	Cal: Not evaluated Disc : Sensitivity: DGF-prediction 80% correct Ov: Not evaluated

Abbreviations [part 1 to part 3]: EPV, events per variable; EPP, events per predictor; n, number; Int, internal;Ext, External;Cal, calibration; Disc, discrimination; Ov, overall;
HL test, Hosmer-Lemeshow test; AUC, area under the receiver operating curve; PPV, positive predictive value; NPV, negative predictive value; LASSO, least absolute shrinkage and selection operator.

Of particular interest for this work is the comparison of the performance of existing scores that include *histological predictors* in addition to clinical and immunological predictors. Histological scores come from Balaz, Zheng, Losappio, Wang and Luo. While Balaz uses Banff lesion scores when assessing histology, Losappio, Wang and Zheng used Remuzzi scores. The authors around Luo, on the other hand, were the first to use deep learning algorithms in 2022 to extract features from whole-slide images of pre-transplant biopsies using EfficientNet-B5, and to update the predictive model created with clinical data. The combination of histological and clinical parameters increased the AUC from .708 to .799. According to Balaz, BANFF ci and cv combined in one score were significantly associated with DGF; Losappio generated a histological Remuzzi score based on the assessment of four compartments glomeruli, tubules, interstitium and vessels, supplemented by clinical parameters. Wang focused on acute tubular injury (ATI) and compared outcomes between donor kidneys with low vs. high ATI. Zheng also found that the AUC could be increased to .89 by including histopathological Remuzzi scores combined with a donor score, ATI and terminal resistance, compared to AUCs of .65 - .75, the values of the individual predictors.

Differences can be partly explained by different *definitions* of DGF. In most of the cases, this was equivalent to the return to dialysis within the first week after transplantation, which is the definition used here. There was a discrepancy in the early publication by Shoskes (1998) and additionally in Sun (2018) and Quiao (2021). In all three publications, conspicuous postoperative serum creatinine values were also counted as DGF. Quiao's study also extended the time to DGF to 14 days after transplantation; Chaumont defined DGF as the need for dialysis after transplantation, regardless of its duration, excluding recipients without interruption of dialysis.

Another difference lies in the *applicability and quality* of the scores and algorithms, as measured by the PROBAST criteria. Here, models were considered to be applicable with high concern due to the selection of the predictors, if they included variables collected during or after transplantation, such as warm ischaemia time. This means that they can no longer be used to support the decision for or against transplantation. This was the case in three papers by Pan, Chen and Decruyenaere [73, 67, 31].

When selecting the cohort, it should be clearly defined. In particular underage and living donors and underage recipients should be excluded. Uncertainty about their exclusion prevailed in twelve cases [55, 24, 61, 58, 33, 77, 66, 67, 69, 73, 72, 76]. Applicability was also judged to be unclear in studies that used small cohorts for training. A widely discussed rule of thumb in logistic regression is that there should be approximately ten observed events in the cohort for each predictor in the model [78]. This ratio was frequently undercut, especially in studies using single-center cohorts, as shown in Tables 2.4, 2.5 and 2.6. The modelling methods were generally appropriate. However, there were limitations when the selection of variables for the final models

Table 2.7: Risk of bias of models predicting DGF (PROBAST)

Author, Year	Risk of Bias (RoB)				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	RoB	Applicability
Brier, 2003	?	+	+	-	-	+	+	-	-
Konieczny, 2021	?	+	+	+	-	+	+	?	-
Bae, 2020	+	+	+	+	+	+	+	+	+
Irish, 2003	+	+	+	+	+	+	+	+	+
Moore, 2007	-	+	+	?	?	+	+	-	?
Jeldres, 2009	?	+	+	+	?	+	+	?	?
Irish, 2010	+	+	+	+	+	+	+	+	+
Balaz, 2013	?	+	+	?	?	+	+	?	?
Chapal, 2014	+	+	+	+	+	+	+	+	+
Zaza, 2014	-	+	+	?	-	+	+	-	-
Decruyenaere, 2015	+	+	+	+	+	-	+	+	-
Decruyenaere (meta-model), 2015	+	+	+	+	+	+	+	+	+
Shoskes, 1998	?	+	-	+	?	+	?	-	?
Ding, 2017	-	+	+	?	-	+	+	-	-
Sun, 2018	+	+	+	-	-	+	+	-	-
Chen, 2020	+	+	+	?	+	-	+	?	-
Zheng, 2020	+	+	+	+	?	+	+	+	?
Kawakita, 2020	+	+	+	+	-	+	+	+	-
Luo, 2022	?	+	-	?	-	+	-	-	-
Xue, 2021	+	+	+	-	-	+	+	-	-
Wang, 2022	-	+	+	?	+	+	+	-	+
Zhao, 2022	+	+	+	+	?	+	+	+	?
Pan, 2021	?	+	+	?	-	-	+	?	-
Costa, 2020	+	+	+	+	?	+	+	+	?
Losappio, 2014	+	+	+	-	?	+	+	-	?
Santos, 2023	+	+	+	-	?	+	+	-	?
Qiao, 2021	+	+	+	+	?	+	+	+	?
Chaumont, 2015	?	+	+	+	+	-	+	?	-
Jen, 2021	+	+	+	-	+	+	+	-	+
Ding, 2018	-	+	+	+	?	+	+	-	?

Abbreviations: +, high; -, low; ?, unclear

was based on univariate analyses or the method used to select the predictors was not described. In some cases, model-based variable selection was omitted, resulting in a large number of predictors. In practice, this has the disadvantage that predictive accuracy also suffers when the probability of missing values is higher. A summary of the assessment of applicability and risk of bias is given in Table 2.7.

2.2 Scores and algorithms predicting transplant loss within one year after transplantation

For the outcome 1y-tl, nine publications of prediction models published since 2002 were identified that did not explicitly make predictions for underaged recipients or living donors. In seven publications, registries (UNOS and the National Scientific Transplant Registry database) were the *data source*, while the cohorts of the other two publications came from individual clinics or transplant centres. Finally, eight publications were from the USA and one from Iran. Study characteristics of the published models predicting 1y-tl are shown in Table 2.8, including enrolment period, study setting and some baseline characteristics. The latter were more likely to be

missing in publications with the outcome 1y-tl compared to DGF.

As summarised in Tables 2.9 and 2.10, Cox regression was the main *modelling method* in a publication by Port, 2002 [79] and was also used by Lin, 2008 [80] and Paquette, 2022 [81] for comparison with machine learning. Deep learning (Bayesian belief networks and Bayes net classifiers) and machine learning were used for the other models by Schold, 2005 [82], Krikov, 2007 [83], Brown, 2012 [84], Naqvi, 2021 [85], Shahmoradi, 2016 [86] and Li, 2010 [87].

Internal *validation* using cross-validation or a randomly selected subset of the data was performed in seven papers, while the performance of one model was measured on the entire training dataset. Brown was the only author to perform an external validation with a temporally independent cohort. No information on any validation procedure was provided in the paper by Port.

The *ability to discriminate* was indicated by means of AUC and graphical presentation in the *internal validation* five times: in Krikov's tree-based models (AUC=.63), Paquette's Cox model (AUC=.646), random survival forest (AUC=.644) and artificial neural networks (DeepSurv, AUC=.65; DeepHit, AUC=.661; RNN, AUC =. 659), Brown's Bayesian Belief Network (AUC=.59), Naqvi's logistic regression model (AUC=.62), random forest (AUC=.70), SVM (AUC=.82) and neural networks (AdaBoost, AUC=.78; ANN, AUC=.61) as well as by Lin's logistic regression (AUC=.71), single-output ANN (AUC=.73), multiple-output ANN (AUC=.61), Cox model without time-varying effect (AUC=.65) and with time-varying effect (AUC=.72) and by Shahmoradi, where the accuracy of three data-mining methods was given (classification and regression tree =.89; ANN=.87; C5.0=.915).

Information on *calibration and overall performance* was, as with DGF, much less common (N=2 and N=3 publications respectively). The AUCs and accuracy data given are based on registry data from 5144 to 180141 observations, and a retrospective cohort of 513 transplants in Shahmoradi. Schold stated that he had worked with a training and test dataset, but no information was provided on the possibility of discrimination and calibration. In the external validation cohort, with 55 out of 138 transplants lost in the first year after transplantation, the AUC for Brown's model was =.63.

Apart from Naqvi's models, the performance between regression and machine learning models was comparable. There, neural networks and SVMs performed significantly better in a direct comparison, although it was not clear how large the numbers of predictors in the final models were. Li's cohort consisted of 1228 transplants and was validated internally on the entire training dataset.

Table 2.8: Study characteristics of published models and scores predicting 1y-tl

Author, Year	Study design	Enrolment period	Study setting	Study region	Participant characteristics:				
					Transplant lost within 1st year = yes	CIT (hr)	Age, years (donor)	Number of previous transplants >= 1	Age, years (recipient)
Port, 2002 [79]	Existing registry	06.03.1995 to 30.11.2000	This current study was undertaken to provide information from relevant national data to refine the definition of expanded donor kidneys, based on recent practice in the United States. This study also sought to provide a factual basis for possible modifications of national allocation policy with the broader goal of achieving greater utilization of retrieved cadaveric kidneys.	United States of America	NA	20.3	36 (17)	NA	47 (14)
Schold, 2005 [82]	Existing registry	between 1996 and 2002	We intended to demonstrate an objective measure relating to the quality of deceased donor organs in the most comprehensive, accurate and adaptable manner.We further investigated several ramifications of the more granulated donor risk stratification in the context of other clinical outcomes.	United States of America	NA	NA	0 (0)	NA	NA
Krikov, 2007 [83]	Existing registry	01.01.1990 to 31.12.1999; follow-up to 31.12.2000	The attempt herein described is undertaken to develop a tree-based model predicting the probability of transplant survival at posttransplant year 1, 3, 5, 7, and 10.	United States of America	NA	15.5 (8.7)	NA	1.2 (0.4)	34.4 (15.5)
Paquette, 2022 [81]	Existing registry	01.01.2000 to 31.12.2019	The objective of this project is to develop an innovative solution of technology readiness level 4 (TRL-4; component and validation in a laboratory environment) that would use ML to support medical decisions about accepting kidney transplants for particular donor-recipient pairs, with specific attention to DCD donations.	United States of America	NA	NA	NA	NA	NA
Brown, 2012 [84]	Existing registry	between 2000 and 2001	We explored the principle of Bayesian Belief Network (BBN) to determine whether a predictive model of transplant survival can be derived using pretransplant variables. Our hypothesis was that pretransplant donor and recipient variables, when considered together as a network, add incremental value to the classification of transplant survival	United States of America	6.6	NA	NA	0 (100)	NA
Naqvi, 2021 [85]	Existing registry	between 2000 and 2017	In this study, the intent is to investigate kidney transplant allograft survival, that is, estimating the time-to-event and the evolving influence of clinical features leading to an event—within three temporal cohorts of 1 year, >1-5 years, and >5 years of a kidney transplant.	United States of Amerika	7554 (14.3)	NA	NA	NA	NA
Shahmoradi, 2016 [86]	Retrospective cohort	09.2007 to 09.2013	The present study aims at comparing the three predictive classification models of neural networks, C5.0 and C&R Tree in predicting kidney transplant survival before transplantation.	Sina Hospital Urology Research Center, Iran	NA	NA	17 to 58	NA	9 to 68
Li, 2010 [87]	Retrospective cohort	1987 to 2009	Prediction of transplant survival rate for a given pool of potential recipients for an organ transplant such as a kidney is helpful for decision makers. It is the aim of this paper to investigate the feasibility of such a tool in the form of a Bayesian belief network classifier designed to predict the renal transplant status and survival period for a given patient.	University of Toledo Medical Center Hospital, USA	71 (6)	NA	NA	NA	NA
Lin, 2008 [80]	Existing registry	01.01.1995 to 31.12.2002	This paper compared single and multiple time-point models in the prediction of graft and recipient survival in kidney transplantation. This study investigated two regression modeling techniques and two types of ANNs: logistic regression and single-output ANNs (as single time-point models) versus Cox models and multiple-output ANNs (as multiple time-point models).	United States of America	3531 (7)	NA	NA	NA	NA

Abbreviations: hr, hour; CIT, cold ischaemia time; No, number of; NA, not applicable; DCD, donation after circular death; ANN, artificial neural network

Table 2.9: Modelling of published models and scores predicting 1y-tl [part 1]

Author, Year	Modelling method	Sample size	Events n (%)	No predictors Cand.	No predictors Final	EPV or EPP	Selection of candidate predictors	Selection of final predictors	Number (%) and handling of missing data	Type of validation	Performance measures
Port, 2002	Cox regression	29068	NA	15	13	NA	prior knowledge	Pre-specified model (not selection)	n (%): 0 (0.0) Method: Complete-case analysis	Int: No information Ext : None	Cal: Not evaluated Disc : Not evaluated Ov: Not evaluated
Schold, 2005	Unclear	NA	NA	13	13	NA	Unclear	Unclear	n (%): Unkown Method: Missing variables for the variable cold ischemia time were categorized as an additional level and utilized in the model construction, missing values of pretransplant dialysis time were considered preemptive transplants, for the variable of recipient weight in the Cockgroft-Gualt clearance estimation, we utilized the average weight observed in the immediate follow-up periods if it was missing, or in several cases, when the straddling weights were also not available,we utilized the transplant weight indication alone.	Int: Random split data Ext : None	Cal: Not evaluated Disc : Not evaluated Ov: Not evaluated
Krikov, 2007	Machine learning techniques	92844	NA	32	31	NA	Only variables that had significant association ($p < 0.05$) with the outcome in all (..) models were included in the final tree-based analysis. In addition (...) we also included several variables that were originally excluded.	(...) we tested the tree-based model for convergence and demonstrated poor performance, which were thought to be due to potential collinearity in the data. To make the model more practical and parsimonious we evaluated the performance of the model with the shorter list of variables, excluding the variables that were considered nonessential.We also excluded the variable describing dialysis network because the model did not converge in its presence.	n (%): 0 (0.0) Method: Complete-case analysis	Int: Random split data Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph Ov: R-squared / Correlation r = .94
Paquette, 2022	Machine learning and regression models	180141	NA	170	NA	NA	prior knowledge	No information	n (%): Unkown Method: Unclear	Int: Cross-validation Ext : None	Cal: Calibration plot / Integrated Calibration Index (0.00942 - 0.01171) Disc : C-Statistic Ov: Brier score
Brown, 2012	The Bayesian Belief Networks (BBNs)	5144	NA	793	49	NA	prior knowledge	The current network model was constructed using a minimum description length (MDL) gain (a weighting of the MDL or the bayesian information criterion that trades off goodness of fit for model complexity) of 0.5	n (%): Unkown Method: An additional bin was included for missing data where appropriate.	Int: Cross-validation Ext : Temporal	Cal: Not evaluated Disc : C-Statistic / AUC graph Ov: Not evaluated

Table 2.10: Modelling of published models and scores predicting 1y-tl [part 2]

Author, Year	Modelling method	Sample size	Events n (%)	No predictors Cand.	No predictors Final	EPV or EPP	Selection of candidate predictors	Selection of final predictors	Number (%) and handling of missing data	Type of validation	Performance measures
Naqvi, 2021	Machine learning techniques	52827	7554 (14.3)	37	37	204.2	Algorithm based (ANN, RF, SVM, Log. Reg., Adaptive boosting)	Other	n (%): Unkown Method: Complete-case analysis	Int: Cross-validation Ext : None	Cal: Not evaluated Disc : C-Statistic / AUC graph / Log-rank test / F1 measure (SVM .61; AdaBoost .56; RF .45; ANN .05; LR .39) Ov: Not evaluated
Shahmoradi, 2016	Machine learning techniques	360	NA	11	11	Unknown	prior knowledge	Pre-specified model (not selection)	n (%): Unkown Method: Single imputation of age	Int: None (Apparent performance) Ext : None	Cal: Not evaluated Disc : C-Statistic / Sensitivity and specificity Ov: Not evaluated
Li, 2010	Bayes Net Classifiers (Local K2, Kill Climber, Tabu Search, CI Search, Genetic Algorithm, Simulated Annealing)	1228	71 (5.8)	70	70	1.0	prior knowledge	Pre-specified model (not selection)	n (%): Unkown Method: Unclear	Int: None (Apparent performance) Ext : None	Cal: Not evaluated Disc : C-Statistic / Sensitivity and specitivity Ov: Not evaluated
Lin, 2008	Machine learning and regression models	57389	3531 (6.2)	71	71	49.7	prior knowledge	Pre-specified model (not selection)	n (%): 0 (0.0) Method: Complete-case analysis	Int: Cross-validation Ext : None	Cal: HL test Disc : C-Statistic / Global goodness-of-fit for proportional hazards assumption Ov: Not evaluated

Abbreviations [part 1 to part 3]: EPV, events per variable; EPP, events per predictor; n, number; Int, internal; Ext, External; Cal, calibration; Disc, discrimination; Ov, overall; RF, random forest; HL test, Hosmer-Lemeshow test; AUC, area under the receiver operating curve; PPV, positive predictive value; NPV, negative predictive value; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; LR, logistic regression; NA, not applicable.

Table 2.11: Risk of bias of models predicting 1-year transplant loss (PROBAST)

Author, Year	Risk of Bias (RoB)				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	RoB	Applicability
Port, 2002	+	+	+	-	+	+	+	-	+
Schold, 2005	+	+	?	?	+	+	?	?	?
Krikov, 2007	+	+	+	+	-	+	+	+	-
Paquette, 2022	+	+	+	+	-	+	+	+	-
Brown, 2012	+	+	+	+	+	-	+	+	-
Naqvi, 2021	+	+	+	+	+	+	+	+	+
Shahmoradi, 2016	?	+	?	+	-	+	+	?	-
Li, 2010	+	+	+	+	?	-	-	+	-
Lin, 2008	+	+	+	+	?	+	+	+	?

Abbreviations: +, high; -, low; ?, unclear

The comparison of scores or algorithms with and without consideration of histological factors is not applicable, as all publications only included clinical and demographic variables.

Differences in the *definition of the outcome* were mainly related to the treatment of deceased recipients. Schold and Li did not report this, while Krikov, Paquette and Lin censored deceased recipients and Shahmoradi, Brown, Port and Naqvi considered them as surrogate markers of transplant loss. In none of the publications was death independent of graft failure considered a competing risk and analysed accordingly.

As for the outcome DGF, models including predictors measured after or during transplantation were considered to be of high concern regarding their *applicability*. This applied to the model by Krikov who included immunosuppressive therapy at time of hospital discharche, and Brown and Li who both included warm ischaemia time. There were major concerns about applicability due to the inclusion of living or paediatric donors or paediatric recipients in the cohorts by Krikov, Paquette and Shahmoradi. Li and Lin did not report any such exclusion criteria. As described above, the outcome definition by Schold and Li was not clear, which is why applicability is of high concern. A summary of the assessment of applicability and risk of bias is given in Table 2.11. In the absence of validation, the risk of bias was unclear with regard to the analysis as it was the case in the Port paper, where follow-up duration was also unclear.

2.3 Nephropathological findings derived from pre-transplantation biopsies: evidence so far

Including 47 studies published between 1994 and July 1, 2014 in their systematic review on the utility of both procurement and implantation biopsies for predicting posttransplant outcomes, Wang et. al. reported that "the most salient finding (...) is that there were no consistent associations between donor biopsy findings and posttransplant outcomes" [88]. However, according to the authors, the studies were characterised by inconsistencies in histological evaluation and interpretation and other limitations. A subsequent review by Moeckli from 2019 on the evaluation of donor kidneys prior to transplantation also concluded that "the current literature fails to demonstrate the clinical utility of pretransplantation histological assessment of grafts" [89]. Authors describe concerns regarding pretransplant histology:

1 lack of evidence of the clinical utility

2 missing consensus on the relative importance of each histological factor

3 increased material and personnel costs

4 difficulties in scoring of the biopsies by (on call) pathologists, especially in centres without experienced nephropathologists which might lead to low reproducibility with regard to the inter- and intrarater correlation of findings [90, 91]

5 an increase in cold ischaemia time, which itself is associated with worse transplant outcomes

This may partly explain the limited availability of such an assessment and, consequently, the irregular collection of this type of data. The scores quoted in Section 2.1 for predicting DGF, for example, all used data from single centres where histological evaluation is performed. Registries such as UNOS/OPTN, Eurotransplant, ANZDATA, the UKRR and DIVAT collect data from a large number of transplant centres and hospitals, not all of which are able to provide this information. However, the need to provide a platform for a systematic, standardised assessment has been met: the OPTN Deceased Donor Registration Worksheet (available online at `https://unos.org/wp-content/uploads/DDR.pdf`), last updated on September 14, 2023, now includes right and left kidney biopsy evaluation as a mandatory section. A similar quality assessment is included in the quality-forms by the DSO.isys portal, a system used by German transplant centres, hospitals and laboratories to coordinate

transplantations (homepage: https://isysweb.dso.de). Section 2.4 lists the minimum data that must be reported to Eurotransplant prior to donor matching or the allocation. Histological evaluation, however, is only required and evaluated in the case of unexpected findings during the surgical examination of the donor [92]. According to their dataset Specification, version 2023.3, the ANZDATA also collects histological parameters including BANFF lesion scores [93]. Yet, histological parameters are not included in the dataset by the UKRR, also known as CKD/AKI clinical dataset [94].

Tackling the first concern, standardisation how the biopsy is performed, the kidney after transplantation stored and reporting of these is suggested. The use of paraffin sections (PS) or frozen sections (FS) differs not only in the time and resources they take, over 3 hours vs. half an our, but also in the possible histological assessments that can be made with them. Also, one should describe which type of biopsy is used: needle, punch or wedge, as the functional compartments e.g. number of glomeruli, differ regarding this. After explantation, machine perfusion or static cold storage are used before transplantation for organ storage and preservation. Machine perfusion (hypothermic or normothermic) itself can be used for assessment of organ quality as perfusion parameters like renovascular resistence are shown to be associated with transplant outcome [89]. The European Society for Organ Transplantation (ESOT) has also set itself the goal of standardising the performance of pre-transplantation biopsies by formulating PICOS and, based on this, a guideline. In 2023, it published the results in 'European Society for Organ Transplantation (ESOT)-TLJ 3.0 Consensus on Histopathological Analysis of Pre-Implantation Donor Kidney Biopsy: Redefining the Role in the Process of Graft Assessment' [39]. Also, prospective studies addressing the first concern are required. The National Health Service has started a registry-based trial, "PITHIA", that aims to evaluate pre-implantation kidney biopsies from donors aged over 60 years to increase the number and quality of kidneys transplanted (trial homepage https://www.nhsbt.nhs.uk/clinical-trials-unit/current-trials-and-studies/pithia/) [95]. During the trial, 22 participating UK centres will have access to biopsy service. The results are still pending, but should provide more evidence in the future on the question of whether a histopathological evaluation before transplantation is even useful, as called for by Wang and Moeckli in the conclusion of their reviews. With regard to reporting, tools like the "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies" [96] were developed to ensure reporting of information relevant to assess quality of the study.

The second concern could be tackled by evaluating well defined, composite histopathological scoring systems rather than single parameters. Thereby, reported inconsistency in assessment could be reduced, when scored in a standardised manner, and more or less important histological factors be identified. Known composite scoring

systems are the Maryland Aggregate Pathology Index [97], the Charonic Allograft Damage Index (CADI) [98], the Remuzzi score [99] and the Banff criteria [100]. Although there is evidence of a significant association between these composite scoring systems and transplant outcome, there are only few predictive models for the outcomes DGF (Sections 2.1) and transplant and patient survival that include them [101, 102, 103, 57, 104, 105].

Lately, advances in digital pathology using deep learning showed promising progress in the detection of pathological (glomerular or interstitial) abnormalities. Supportive systems like this could, when implemented as a routine in the transplant centres or clinics, augment the diagnostic process [106, 107, 108, 70]. This could improve consistency of nephropathological findings and accelerate the evaluation process, thereby reducing concerns 3 to 5 mentioned by Moeckli, among others. In wedge biopsies, for example, the number of glomeruli counted might easily be more than one hundred. Assessment of Banff Lesion scores of the vascular-glomerular compartment and glomerular sclerosis in this case are very time-consuming and error-prone. Automated segmentation and classification could assisst by marking the glomeruli and extraction of counts. Also, digital pathology is locally not bounded: biopsies could be preprocessed and scanned at one clinic, and evaluated and scored in another.

Regardless of the results by the PITHIA trial or preferred method of taking a biopsy and its evaluation, the optional assessment of histological parameters should be reflected in the prediction models derived from and used by transplant registries and centres to make best use of the available information.

The 2-step approach with optional histopathology, including assessment of the Banff Lesion scores as described in section 3.3.8, is intended to address this issue.

2.4 Eurotransplant Kidney Allocation System

Eurotransplant not only provides a register in which patients waiting for an organ donation (in addition to the kidney, this also includes the liver, pancreas, heart, lungs and intestines) are registered, but also offers guidelines according to which donor organs are allocated to them 24/7. A follow-up register is also kept for monitoring purposes to evaluate post-transplant results.

Standards of quality and safety are followed according to the Directive 2010/45/EU of the European Parliament and of the Council of 7 July 2010 [109]. This states, among other things, that before a potential donor is registered in and allocated by Eurotransplant, a minimal amount of information must be reported before Eurotransplant duty desk. This includes the registration date, center, ABO, HR, donor type, sex, height, weight, date of birth and virology (HbsAg, HbcAb, HCV, HIV) [92].

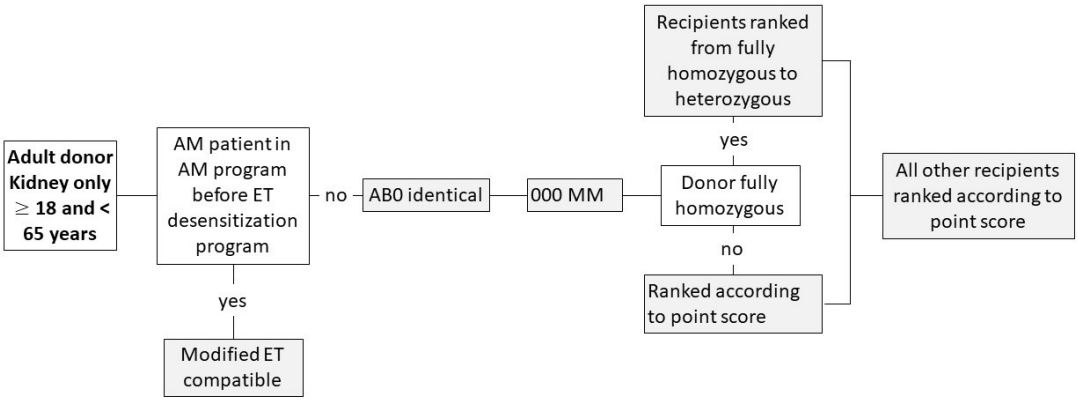
According to chapter 4 of the Eurotransplant manual [92], kidneys of donors $<$65 years are allocated according to the Eurotransplant Kidney Allocation System (ETKAS, algorithm Figure 2.1) and to the Eurotransplant Senior Program (ESP, algorithm Figure 2.2), otherwise. Before being allocated to ETKAS, special consideration is given to recipients who are, due to a current and/or historical HLA-sensitization, immunologically compromised by including them into the Acceptable Mismatch (AM) program.

After 3 years in the AM program, patients not having received a donation have the possibility to enter the ET desensitization program. In both the AM and ET desensitization program, potential recipients that have an high urgency (HU) status will be prioritised above non-high urgent patients. Also, these patients have priority above ETKAS-selected patients.

In ETKAS, potential recipients of donors $\geq$ 18 years and $<$ 65 yeas of age are first identified using the AB0 blood group rule (A with A, B with B, 0 with 0 and AB with AB). In case of deceased donors, AB0-incompatible kidney transplantations are not allowed. This is followed by a point-score based ranking according to:

- age [age $<$ 18 years: bonus of 100 points $+$ double points given for HLA-antigen mismatches, outside Germany additional points until the age of 30 are given in a gradual system]
- waiting time [33.3 points/year]
- medical urgency [500 points in case of high urgency]
- HLA-A, -B and -DR mismatches (000 MM) [0 - 400 points]
- mismatch probability [0 - 100]

- distance between donor and transplant center [0, 100 or 200 points]
- national [(highest import balance - recipient country balance)*30 points] and regional (Austria only [(Austrian National Balance - Regional Balance)*0.25 points]) exchange balance



Eurotransplant manual – version 2024.1 January 23, 2024 – subject to change, page 19

Figure 2.1: ETKAS allocation algorithm [92]

The mismatch probability, taking into account the ABO blood group rules and vPRA, calculates the probability of receiving a kidney offer with 0 and 1 broad HLA-A, -B or split DR mismatches based on 1000 kidneys offered. With regard to the distance between donor and transplant center, national regulations are considered. Bonus points are possible in case of kidney after other organ transplant (KAOO).

The ESP allocation algorithms first perform a patient-oriented kidney allocation to local, regional, or national patients. Within the region/nation, highly urgent, followed by KAOO patients, and patients with a longer waiting time (=dialysis time) are at the top of the waiting list. In case of no match according to these criteria, rescue allocation rules apply: first extended allocation followed by second- and third-line rescue.

Patients can be included in the ESP or the ETKAS program. In Germany, patients $\geq$ 65 have to choose to be included in either program.

Special donor characteristics that must be taken into account before decision making are, in addition to AB0-blood group, HLA-mismatches and age: virology (-HBsAg, HBcAb, HCVAb), domino donor, sepsis, meningitis, malignant tumor, IV-drug abuse, kidney en bloc, donation after circulatory vs. brain death and euthanasia.

Further information on the background of the allocation algorithms and the point scores used can be found on the Eurotransplant homepage (`https://www.eurotransplant.org/allocation/`).

Eurotransplant manual – version 2024.1 January 23, 2024 – subject to change, page 20

Figure 2.2: ESP allocation algorithm [92]

Chapter 3

Methodological Background

3.1 Adaptation of CRISP-DM reference model to the medical score development

The CRISP-DM reference model was created at the end of 1996. It provides an overview of the life cycle of a data mining project. The left side of Figure 3.1 displays the life cycle of a data mining project as proposed by the CRISP-DM Consortium [45]. Starting with business understanding, the focus is on "understanding the project objectives and requirements". Next, data collection and first exploration should lead to a good data understanding. Hypotheses regarding hidden information are generated in this step. To test the set up hypotheses, data preparation, covering all activities needed to construct the analysis dataset, is required. However, the sequence of the steps is not rigid and it might be required to move one step back. Once the analysis is prepared, the modelling can start (techniques are selected, parameters calibrated). Developed models are then evaluated, also with regard to business issues that have not been considered. The final model then needs to be presented and organised in a way the desired customers can use it. As emphasised by the authors "even if the analyst will carry out the deployment effort, it is important for the customer to understand up front what actions need to be carried out in order to actually make use of the created models" [45]. Although developed for industrial use, CRISP-DM is also applied in medical decision making [110] and adapted to machine-learning [111].

In the field of medical statistics, Steyerberg and Vergouwe proposed a similar seven step instruction for development of a predictive model [112]. As shown in the right side of Figure 3.1, their first step "Consideration of research question and initial data inspection", combines the first CRISP-DM steps. The research question comes first, accompanied by a review on what is already know about predictors. In this preparation step, the authors highlight the importance of a close interaction between clinical researchers and statisticians. The second step describes the coding of

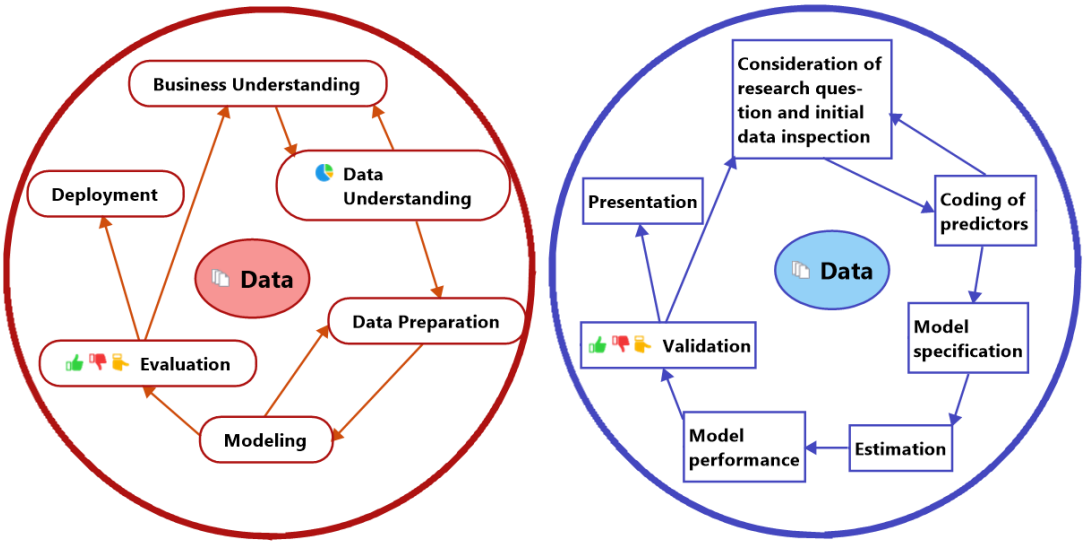


Figure 3.1: Prediction modeling cycles [45]

predictors, aka data preparation. However, dichotomization of continuous predictors in this step is not advised [113]. Before the final model is estimated in the fourth step, "Estimation", predictors for inclusion in the prediction model are explored by model specification. This step is challenging, particular in case of datasets with small sample sizes, as the selection might be unstable. Given several different models, compatible with a set of observations, Occam's razor advises us to choose the simplest. Bayesian inference automatically embodies Occam's razor as "simple models tend to make precise predictions. Complex models, by their nature, are capable of making a greater variety of predictions [114]. Steyerberg and Vergouwe, as well, propose ,without mentioning Bayesian inference, that "a simple, robust model may not fit the data perfectly, but should be preferred to an overly fine-tuned model for the specific data under study". After estimation of the model the quality need to be determined by assessing discrimination and calibration in the "Model performance" step. The last step before presenting and publishing the result is the validation. Internal validation refers to the assessment of the model performance based on a validation that comes from the same cohort the training was derived. External validation, an in general stronger test for prediction models, refers to the generalizability of the model [112]. In the "Presentation" step, one needs to decide on the proper format, paper- or electronic-based, appropriate for the user. However, despite comprehensive planning and proper application of proposed steps, there's no guarantee that the results are practical and applicable and should therefore be presented. Validation can lead to a revision of the initial hypothesis and return to the first step.

In this thesis, the development and validation of the DGF- and, respectively, 1y-

tl-score, follow the statistical prediction modelling cycle.

3.2 Selection and exclusion of clinical predictors based on availability and correlation

Preliminary, clinical variables were taken into consideration if correlation with either DGF or 1y-tl was shown in the relevant literature and if they were contained in the Eurotransplant database. Before applying variable selection via statistical and machine-learning methods, variables were excluded from the initial selection in case of:

- $\geq 40\%$ ***missing values*** in the training dataset. Although missing values can be imputed for analytical purposes, this method is generally not available to practising clinicians who have to make decisions based on the information available. Therefore, in order to be applicable, only variables collected on a regular basis were considered eligible for score building.

- ***(Near-) zero variance***: relative to the number of samples, they have very few unique values ($< 20\%$) and the number of observations of the most common value divided thru the number of the second most common value is large ($>$ 20) [115].

- ***High correlation*** (correlation coefficient r > 0.8 [116]) with other variables. This criterion should reduce the occurrence of multicollinearity. Multicollinearity can lead to "ambiguity in estimation of regression coefficients and selection of variables" [117]. As a result, regression coefficients may not be reasonably interpretable. Also, highly correlated variables provide little independent information [118, 119]. To deal with multicollinearity, the variance inflation factor (VIF), which excludes highly correlated variables through a stepwise procedure, can be used [120]. In a multivariable model, the VIF_j for the j^{th} predictor is the factor by which the variance of its regression coefficient β_j is inflated by the existence of correlation among other independent variables in the model.

$$VIF_j = \frac{1}{1 - R_j^2} \tag{3.1}$$

  where the R^2-value is obtained by regressing the j^{th} predictor on the remaining independent variables. In R, the function *vif()* from the package *usdm* ("Uncertainty Analysis for Species Distribution Models") [121] can be used to obtain the VIF based on the different correlation methods 'pearson', 'spearman' and

'kendall' and from the package *car* ("Companion to Applied Regression") [122] for regression-based and (generalized) linear models.

NOTE: each study centre (aka country) represents one cluster of individuals within the Eurotransplant region. Baseline outcome probabilities may vary between centres, independent of available donor and recipient related covariables, due to systematic differences in treatment policies. This can arise "confounding by cluster" (CBC) caused by confounding of the exposure-outcome relationship [123]. Accounting for these center-effects by including them as fixed or random covariables may be reasonable to improve future predictions, however, due to sample size limitations and deviations in the distributions from both independent and dependent variables, doing so may induce bias. Study center was therefore not included in the analysis despite its significant association with DGF.

3.3 Reduction of clinical candidate predictors based on statistical and machine-learning methods

Throughout the thesis, the following definitions apply: the outcome vectors for DGF and, respectively, 1y-tl are $\vec{Y}_{DGF} \in \{\text{DGF} = 1, \text{no DGF} = 0\}$ and $\vec{Y}_{1y-tl} \in \{\text{TX-loss} = 1, \text{no TX-loss} = 0\}$. The matrix of the *k* candidate predictors with N_{DGF} and, respectively, N_{1y-tl} number of observations is X=$\{\vec{X}_1, ..., \vec{X}_k\}$, with corresponding regression coefficients $\beta^T_{X,Y_{DGF}} = \{\beta_{1,Y_{DGF}}, ..., \beta_{k,Y_{DGF}}\}$ and $\beta^T_{X,Y_{1y-tl}} = \{\beta_{1,Y_{1y-tl}}, ..., \beta_{k,Y_{1y-tl}}\}$ and constant baseline odds, aka intercepts, $\beta_{0,Y_{1y-tl}}$ and $\beta_{0,Y_{DGF}}$.

3.3.1 Reference model: Logistic regression on multiple imputed dataset

Missing outcome or predictor data are a common phenomenon in medical research. The prediction model building on such datasets can be challenging when complete data is required for outcome prediction. The type of missing values can be subdivided according to the underlying mechanisms or pattern into completely at random (MCAR) i.e. independently of other, possibly unobserved values; at random (MAR) and not at random (MNAR)[124, 125]. Several methods are proposed to account for missing data in both a model building, but also validation setting. Complete case (CC) analysis or multiple imputation (MI) approaches are most commonly used [124, 125, 116]. Although not advised as the uncertainty of the imputed values is not fully taken into account in the estimation of the final model, single imputation (SI) such as regression imputation (RI) or average imputation could also be applied [126,

124]. Application of stochastic single imputation may additionally lead to less stable point estimates [116]. However, in case of big datasets with more than 100 events or datasets with relatively few missing values, the disadvantages of SI may be regarded as less relevant and the advantage of a simplified analysis after imputation may dominate [116]. With regard to the missing value pattern, preliminary, both stochastic (regression) imputation and multiple imputation follow the assumption that data are missing at random.

When working with data in the area of transplantation, missing information is more likely to be observed with regard to the donor rather than the recipient or transplant procedure information. This applies in particular to the donors medical history. A donors diagnosis of diabetes or high blood pressure are two examples for conditions which are associated with a decreased kidney function and are therefore collected in the Eurotransplant database [127, 128]. As it was an aim to include these and other variables despite their expected increased proportion of missing values in a prediction model in this first step of the algorithm, multiple imputation, excluding imputation of outcomes, was chosen assuming missing values to be MAR.

On each of the imputed datasets, the method chosen to reduce the number of candidate variables was forward selection (p-value for inclusion α_{in}=0.25) followed by backward elimination (p-value for exclusion α_{ex}=0.1) using "classical" logistic regression which is defined as follows. Given a random variable x the *probability density function* $\pi(x)$ of the logistic distribution is continuous with:

$$\pi(x) = \frac{e^x}{(1+e^x)^2} \tag{3.2}$$

The *cumulative density function* (CDF) of this is:

$$\Pi(x) = \int_{-\infty}^{\infty} \pi(x)dx = \frac{e^x}{1+e^x} \tag{3.3}$$

With *inverse-logit function* $\Pi^{-1}(x) = 1/(1+e^{-x})$. According to Hosmer and Lemeshow, "in any regression problem, the key quantity is the mean value of the outcome variable, given the value of the independent variable" [129], called the conditional mean $E(Y|x)$. In case of the binary outcome DGF ($\vec{Y}_{DGF}$), for example, the probability that an individual j has DGF, given its observed independent variables

$x_{j1}, ..., x_{jk}$ is expressed by

$$\begin{aligned}\Pi(x_j) &= E(y_{j,DGF}|x_{j1}, ..., x_{jk}) \\ &= P(y_{j,DGF} = 1|x_{j1}, ..., x_{jk}) \\ &= \frac{e^{\beta_0+\beta_1 x_{j1}+...+\beta_k x_{jk}}}{1 + e^{\beta_0+\beta_1 x_{j1}+...+\beta_k x_{jk}}} \\ &= \frac{1}{1 + e^{-(\beta_0+\beta_1 x_{j1}+...+\beta_k x_{jk})}} \\ &= (1 - P(y_{j,DGF} = 0|x_{j1}, ..., x_{jk}))\end{aligned} \tag{3.4}$$

Via logit transformation $g(x)$ in terms of $\Pi(x)$, properties of a linear regression model can be derived:

$$g(x) = ln\left[\frac{\Pi(x)}{1 - \Pi(x)}\right] = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \tag{3.5}$$

To derive the estimates $\hat{\beta}_{X,Y}$ of the unknown regression coefficients $\beta_{X,Y}$, the sum of differences of the observed y_i and predicted $\hat{y_i}$ are minimised (equation 3.6) by maximising the likelihood of the observation.

$$\hat{\beta} = minimum_{\beta_0,\beta}\left[\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2\right] \tag{3.6}$$

The contribution of each individual observation (x_i, y_i) to the likelihood is the expression:

$$\Pi(x_i)^{y_i}[1 - \Pi(x_i)]^{(1-y_i)} \tag{3.7}$$

Assuming independence between individual observations, the *likelihood function* is expressed as:

$$l(\beta|Y) = \prod_{i=1}^{N} \Pi(x_i)^{y_i}[1 - \Pi(x_i)]^{1-y_i} \tag{3.8}$$

Taking the natural logarithm of this likelihood function, the *log likelihood function* is defined as:

$$L(\beta) = ln(l(\beta|Y)) = \sum_{i=1}^{N}\{y_i ln(\Pi(x_i)) + (1 - y_i)ln(1 - \Pi(x_i))\} \tag{3.9}$$

The values of $\hat{\beta}$ derived by maximising $L(\beta)$ are called *maximum likelihood estimates*.

Forward selection begins with fitting an "intercept only model" ignoring covariable effects [129]. Next, univariable analyses are performed with all possible predictors. The one predictor with the smallest p-value that is smaller than the defined inclusion α_{in}, if there's any, is chosen as X_{e1}. Models including X_{e1} and one of the remaining

k-1 predictors X_j, $j \neq e1$ are then fitted and the pair minimising the p-value p_j^1 for the likelihood ratio chi-square statistic with regard to the log-likelihood of solely considering X_{e1}, selected ($X_j = X_{e2}$). If p_j^1 is greater than α_{in}, we stop. Otherwise, all models including both X_{e1} and X_{e2} and one of the remaining X_j, $j \neq \{e1, e2\}$ are fitted. This procedure continues until step m if no more variables are left to be included ($m = k$) or the likelihood ratio chi-square statistic $p_m^{(m-1)}$ is greater than α_{in} for all variables X_j with $j \notin \{e1, .., e(m-1)\}$. Backward elimination, in contrast, starts with the full model including all predictors and iteratively eliminates predictors with p-values $> \alpha_{ex}$ [129]. Here, the variable with the smallest test statistic, inter alia, highest p-value greater than the selected α_{ex} is removed first and the procedure stopped, once all remaining variables are significantly associated with the outcome of interest. For each of the variables X_i selected on each individual imputation set, regression coefficients and their standard errors can be pooled into a single model applying Rubin's Rule, which accounts for both within and between-imputation variability, as follows [130, 131, 132]: given *l* imputed datasets, the overall point estimate of each regression coefficient *i*, $\beta_{i,Y}$, is the average of the $m \leq l$ estimates of $\hat{\beta}_{i,Y}$ from the imputed datasets

$$\overline{\beta}_{i,Y} = \frac{1}{m} \sum_{j=1}^{m} \hat{\beta}_{ij,Y} \tag{3.10}$$

The associated total variance T_i is derived by assessing the within W_i and between B_i imputation variance by

$$T_i = \overline{W_i} + \left(1 + \frac{1}{m}\right) B_i \tag{3.11}$$

with

$$\overline{W_i} = \frac{1}{m} \sum_{j=1}^{m} \hat{W}_{ij} \tag{3.12}$$

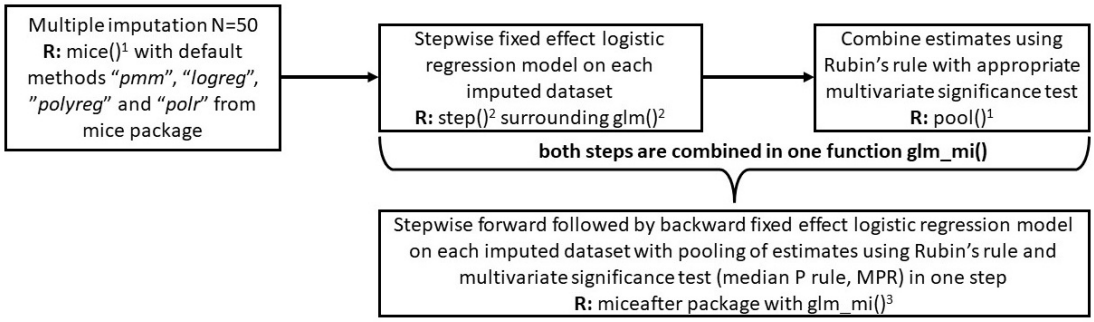
and

$$B_i = \frac{1}{m-1} \sum_{j=1}^{m} \left(\hat{\beta}_{ij} - \overline{\beta}_i\right)^2 \tag{3.13}$$

Depending on the variable types (numeric, binary, categorical with $\geq$ 3 groups), different multivariate significance tests for combining *k* different estimates $\left\{\hat{\beta}_{1,Y}, ..., \hat{\beta}_{k,Y}\right\}$ from *m* imputed datasets are available and described elsewhere [132, 133]. As a mix of all mentioned variable types was assumed to get selected, the *median P rule* (MPR) method, which was originally developed for comparing predictive performances of two methods in a cross-validation setting, was applied [134, 132]. Accordingly, reject the null hypothesis $H_{0,i} : \beta_i = 0$ on a given significance level α if:

$$\tilde{p}_i = med(p_{1i}, ..., p_{mi}) \leq \alpha \tag{3.14}$$

For imputation of 50 datasets for each DGF and, respectively, 1y-tl, the R-package with the function of the same name *mice* [135] "Multivariate Imputation by Chained Equations" was used; Rubin's rule to pool the logistic regression estimates was used as implemented in the *miceafter* R-package [136](see Figure 3.2).



[1] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate Imputation by Chained Equations in R. In: Journal of Statistical Software 45.3 (2011), pp. 1–67. DOI:10.118637/jss.v045.i03
[2] R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
[3] Martijn W Heymans (2021). miceafter: Data Analysis and Pooling after Multiple Imputation. R package version 0.5.0. https://mwheymans.github.io/miceafter/

Figure 3.2: Steps: variable selection on multiple imputed datasets

3.3.2 Method 1: Logistic regression on unimputed, complete case dataset

Analyses on multiple imputed datasets are significantly more complex and computationally intensive than those performed on a non-imputated "complete case" dataset. The number of imputations, the method and variables chosen for replacing missing values as well as the procedure for pooling the results could themselves lead to unnecessary distortions and bias. For this reason, the logistic regression as described in the previous section was also performed on the unimputed DGF- and 1y-tl-datasets. As in regression analyses only complete cases can be considered, the number of observations used to derive different models can vary depending on the included variables. For variable selection, a stepwise forward and backward selection was conducted. To avoid the selection of too complex models with many predictors, the *Akaike Information Criterion* (AIC), penalising the logarithm of the likelihood function with the number of variables *k* included in the model, was used [137]. This means that the decision to include or exclude a variable is no longer based on a p-value or significance level (α_{in} and α_{ex}). The AIC with *k* independent variables is defined as [137]:

$$AIC(k) = -2ln(l(\beta|Y)) + 2k \tag{3.15}$$

The AIC allows the comparison of unnested models with different numbers of independent variables by considering the model with the smallest AIC as better. To account for differences in the number of observations $n \leq N_{DGF}$ (respectively N_{1y-tl}), the *Bayesian Information Criterion* (BIC) can be used instead. This no longer penalises complex models with the fixed factor of two, but with the logarithm of the number of observations, which is stronger than the AIC from eight cases ($\ln(8) > 2$).

$$BIC(k) = -2ln(l(\beta|Y)) + ln(n)k \tag{3.16}$$

Without multiple imputation, fixed effect logistic regression and variable selection can be done using the *glm* function "**G**eneralized **L**inear **M**odels" from the *glmnet* package [138] followed by *step* from the *stats* package in R.

3.3.3 Method 2: Univariable p-value < 0.25

This method, for sake of simplicity from now on called "p-method", was described by Hosmer & Lemeshow, [129, chapter 4] and is fairly simple. It bases on the idea that all variables showing an univariable measured p-value < 0.25, along with those of clinical importance, are candidates for a multivariable model. The proposed cutoff of 0.25 bases on the work on linear regression by Bendel & Afifi [139] and logistic regression by Mickey & Greenland [140]. The outcomes of interest are both binary, non-time-dependent. For p-values to be comparable amongst candidate predictors measured on different scales (nominal, ordinal, continuous), the following univariable tests of significance are proposed: in the nominal or ordinal case with few (say m) integer values, the likelihood ratio chi-square test with m-1 degrees-of-freedom or, since it is asymptotically equivalent, the Pearson chi-square test are applicable. For continuous, within each of the two outcome groups approximately normally distributed predictors, the two-sample t-test and the univariable logistic regression are equivalent at the univariable level and therefore applicable [141]. However, the most desirable univariable analysis is the univariable logistic regression model. The p-method was chosen in this thesis as it seems, from personal experience, to be a common practice amongst clinicians.

3.3.4 Method 3: LASSO

So far, only standard "traditional" logistic regression-based variable reduction methods were described. Challenges using logistic regression were shown to arise in various settings like the analysis of outcomes with low prevalences [142], datasets with highly correlated variables (multicollinearity) [143] and small samples with more predictors than observations $k \gg N$ [144]. To deal with the problem of $k \gg N$ or the difficulty

in interpretation rising with a large number of predictors, Tibshirani R. proposed to penalize regression coefficients by "shrinking some coefficients and setting others to 0" [145]. Compared to equation 3.6, the Least Absolute Shrinkage and Selection Operator (LASSO) or l_1 penalization solves the problem [144]:

$$\hat{\beta} = minimum_{\beta_0,\beta} \left[\frac{1}{2} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j}^{k} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \right] \tag{3.17}$$

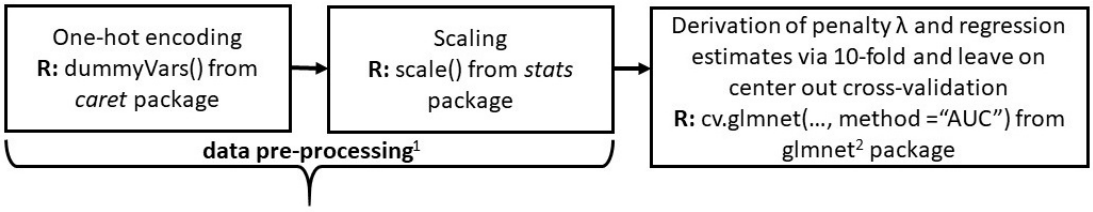
In equation 3.17, λ is an additional tuning parameter balancing the tradeoff between the complexity, inter alia number of predictors, of the model and its goodness of fit and can be derived by cross validation [144]. Once λ is fixed, equation 3.17 has nonzero $\hat{\beta}$ values on a subset of predictors. To select the most informative predictors with regard to the regression coefficients β, data needs to get scaled in advance, particularly in case of variables measured on different scales. For a continuous or one-hot encoded, categorical random variable X_i, scaling is applied with standard deviation (std) σ utilising the following equation:

$$X_{i,scaled} = \frac{X_i - \overline{X_i}}{\sigma_i} = \frac{X_i - mean(X_i)}{std(X_i)} \tag{3.18}$$

One-hot encoding is a common technique applied before running a machine learning algorithm for handling categorical data [146], transforming each categorical value of a variable into a new column to a digital form with value 1 if the value was observed and 0 otherwise. A categorical variable colour with three characteristics "green", "red", "yellow", for example, would get transformed into three variables green, red and yellow. When colour equals "green", green is assigned "1" and "0", otherwise. Shrinkage leads to a less-extreme distribution of predictions, it addresses overfitting and the derived model can be used for prediction [147, 148]. However, LASSO (or penalized regression in general) does not produce as causal interpretable coefficients and, as shown by Belloni and Chernozhukov, a model derived by applying ordinary least squares (OLS) after a first-step LASSO, "performs at least as well as LASSO in terms of the rate of convergence, and has the advantage of a smaller bias" [148, 149]. LASSO, therefore, was used as an alternative method to reduce the number of optional predictors before step 1.

Figure 3.3 describes the steps to perform a LASSO regression with R, beginning with the *dummyVars* function from the *caret* package [150] for encoding, followed by scaling with the *scale* function and in the *cv.glmnet* function implemented algorithms described by Tibshirani et. al. [151][138]. For cross-validation, two folds were used: 10 randomly selected subsets and, for comparison a center-based "leave one center

out" cross-validation. As measure of loss to use in cross-validation the AUC was used.



[1] Kiran Maharana, Surajit Mondal, Bhushankumar Nemade, A review: Data pre-processing and data augmentation techniques, Global Transitions Proceedings, Volume 3, Issue 1, 2022, Pages 91-99, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2022.04.020

[2] https://glmnet.stanford.edu/articles/glmnet.html

Figure 3.3: Steps: LASSO regression

NOTE: in STATA "**lassopack** implements lasso, square-root lasso, elastic net, ridge regression, adaptive lasso, and postestimation ordinary least squares" [148].

3.3.5 Method 4: CART

Tree-based models are said to be conceptually simple and easy to interpret, which facilitates medical decision-making, and they have a sixty-year history of development. The first regression tree algorithm was published by Morgan and Sunquist in 1963, Classification And Regression Trees (CART was developed by Breiman et al. in 1984 [152, 153, 154]. Yet, so far, Decruyenaere and colleagues were the only group applying CART for predicting DGF and this only to compare performance to other methods (tree not published) [31]. While logistic regression and LASSO output regression coefficients of an prediction equation, CART builds a tree with binary splits where the outcome predictions, given the selected variables included in the root and nodes of the tree, are displayed in the lowest level, called leaves (schematic presentation in Figure 3.4). The outcome of interest can be categorical (classification trees [153]) but also continuous (regression trees [153]), counts (Poisson regression trees [154]) or estimate time-to event (survival trees [155, 156]). Additional scaling as before LASSO is not necessary, however, several parameters can be set in advance allowing a high degree of flexibility in modelling.

As the outcomes of interest, DGF and 1y-tl, are both binary, the focus in this section is on binary classification.

At each node A (including the root) of the tree, the set S of observations $\vec{Y}$ can be split at a candidate cutoff s into two subsets S_L and "not S_L" $= S_R$ with n_L and n_R observations and proportions p_L and p_R, respectively ($S = S_L \cup S_R$, $n = n_L + n_R$, $p_L = n_L/n$, $p_R = n_R/n$). While in logistic regression one aims to maximise the likelihood function, the goal of CART is to minimise the impurity of each node A in way that the partaining left split S_L and right split S_R include only observations of

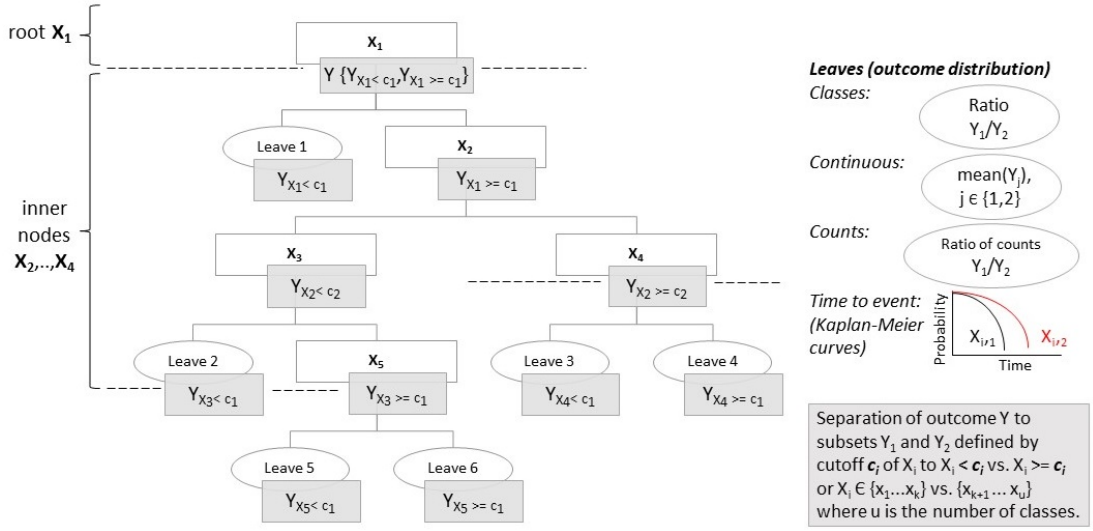


Figure 3.4: CART: example

the same class k and $\tilde{k}$ defining its label ($y_i = k \; \forall \; y_i \in S_L$ and $y_i \in \tilde{k} \; \forall \; y_i \in S_R$, respectively). If S_L, for example, only includes patients without an event of interest and S_R only patients with event, the impurity function $I(S)$ is minimised and no further splitting steps are required. Out of a set of candidates s, splits are therefore selected in a way that the *information gain* aka decrease in impurity $\Delta I(S_L, S_R)$ is maximised where[153]:

$$\Delta I(s, S) = I(S) - p_L * I(S_L) - p_R * I(S_R) \tag{3.19}$$

To measure this information gain for each optional split s, several functions depending on the definition of the *impurity function* $I(S)$ are proposed [157]. Given a subset S with l observations and a splitting variable $X_j = \{x_{1,j}, ..., x_{l,j}\}$ the fraction p_k of data points of class k (respectively $\tilde{k}$) is:

$$\begin{aligned} p(k|S) &= \frac{1}{n_S} \sum_{x_{i,j} \in S} I(y_i = k) \\ &\xrightarrow[\text{classification}]{\text{binary}} \frac{1}{n_S} \sum_{x_{i,j} \in S} I(y_i = k) = 1 - p(\tilde{k}|S) \end{aligned} \tag{3.20}$$

Accordingly, impurity functions include:

$$\begin{aligned} \text{Misclassification error: } I_M(S) &= 1 - max\{p(k|S), p(\tilde{k}|S)\} \\ &= 1 - max\{p(k|S), (1 - p(k|S))\} \\ &= min\{p(k|S), (1 - p(k|S))\} \end{aligned} \tag{3.21}$$

$$\begin{aligned} \text{Gini index:} I_G(S) &= p(k|S)(1-p(k|S)) + p(\tilde{k}|S)(1-p(\tilde{k}|S)) \\ &= 2p(k|S)p(\tilde{k}|S) \end{aligned} \tag{3.22}$$

$$\text{Deviance: } I_D(S) = -p(k|S)log(p(k|S)) - p(\tilde{k}|S)log(p(\tilde{k}|S)) \tag{3.23}$$

As described by Breimann, both 3.22 and 3.23 give similar results [153]. However, the Gini index is easier to compute which is why, for sake of simplicity, one should prefer it over deviance. Each splitting step leads to at least equal or smaller impurity, which is why one additionally has to define a stopping criterion to avoid overly complex trees. Instead of just stopping to split up to a given minimum information gain, Breimann proposed to build a complete tree T_{max} and then apply *minimal cost-complexity pruning*, a method that minimises the risk of misclassification while accounting for the complexity (number of terminal nodes) of the tree.
Given a complexity parameter $\alpha \geq 0$, the number of terminal nodes $|\tilde{T}|$ of any subtree $T \prec T_{max}$, the overall cost of misclassification $R(T)$ for each subtree, assuming equal costs for misclassificating a class k object to $\tilde{k}$ and vice versa, one aims to minimise the cost-complexity measure $R_\alpha(T)$:

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}| \tag{3.24}$$

$$\begin{aligned} \text{Tree misclassification} = R(T) &= \sum_{S \in \tilde{T}} r(S)p(S) \\ &= \sum_{S \in \tilde{T}} R(S) = \text{Sum of node misclassifications} \end{aligned} \tag{3.25}$$

$$\begin{aligned} R(S) &= r(S)p(S) \\ r(S) &= 1 - max\{p(k|S), p(\tilde{k}|S)\} \\ p(S) &= \text{probability that any case falls into node (subset) } S \end{aligned} \tag{3.26}$$

The product $\alpha|\tilde{T}|$ can be regarded as penalty term comparable to LASSO equation 3.17, $|T|$ is analogous to the degree of freedom in regression [158]. The optimal α can be found via cross validation. Additionally to minimal cost-complexity pruning, one can choose a minimal number of observations in the final leaves to reduce complexity.
Besides being simple to interpret, CART, as implemented in the *rpart* routines in R, can handle missing data by estimating surrogate values once a splitting variable X_j and a split point s are selected [158]. A missing value of X_j, or more precisely the class $\{k, \tilde{k}\}$ the observation is to be assigned, is predicted using the other independent

variables by re-applying the partitioning algorithm.
What particularly characterises CART and makes it especially interesting for this 2-step approach is that interaction effects of order two and higher are naturally included in the tree [147]. Only few of the available scores consider two-way interaction effects as regression-based models usually include main effects unless interactions are specified. These include both scores by Irish et al. (2003 and 2010) for the outcome DGF [60, 17]. A classification tree created in advance can provide indications of possible interaction effects relevant in the first step of the logistic regression model. A frequently cited disadvantage of CART, however, is the dichotomisation of the independent variables, which leads to a loss of information [147] and the sensitivity to minor changes in the data [157].
Several packages are available in R, offering algorithms and functions for recursive partitioning, including *rpart*, *tree*, *Weka*, *party*, *C50* and visualization of trees like *maptree*, reflecting the wide range of applications (`https://cran.r-project.org/view=MachineLearning`). In this analysis, the *rpart*-package was used with the *information*-splitting index $I_M(S)$ for building the tree applying 10-fold cross-validation. The minimal number of observations per final leave was set to 8, the complexity parameter α derived was tested for a range of values between 0.01 and 0.1 for pruning off splits that are not worthwhile. Prior probabilities for the outcomes DGF and 1y-tl were chosen to be their prevalences in the analysis sets.

3.3.6 Method 5: VSURF

VSURF is actually the name of the R-package implementing an algorithm for Variable Selection Using Random Forests [159]. This algorithm was proposed by Genuer et al. [160, 161] in 2010 and bases on random forests (RF), a method introduced by Breiman (who also defined CART) in 2001 [161]. With respect to CART, random forests differ mainly in two steps [159]. First, a subset of candidate predictors is randomly chosen and a maximal tree, like in CART, is created. Second, pruning, which would now follow in CART, is not performed and all trees T of the forest are maximal trees. Trees are then aggregated into a single random forest. A more detailed explanation on random forests is given by Breiman [161].
The idea behind VSURF, as proposed by Genuer, is a two step approach: to first use random forests to identify the variables highly related to the outcome variable according to a statistical measure "preliminary elimination an ranking" and then, secondly, to find a small number of variables which is sufficient enough for prediction "variable selection".
The variable importance (VI) is defined for the random forest of the first step as follows [159]: be T a single tree based on a bootstrap sample of observations and OOB_T a

out of bag sample of data not included in the bootstrap sample. Then $errOOB_T$ is the *out of bag error* of a single tree T on OOB_T, equalling the misclassification rate (cf. 3.24) in case of a classification problem. $err\widetilde{OOB}_{T,j}$ is the error of the predictor T after randomly permuting the values of X_j in the sample OOB_T and n_T is the number of trees aggregated in the random forest. Then

$$VI(X_j) = \frac{1}{n_T} \sum_T (err\widetilde{OOB}_{T,j} - errOOB_T) \tag{3.27}$$

is the variable importance of X_j sum up over all trees T of the random forest. Having repeated this procedure for all candidate predictors X_j, variables of small importance with regard to a threshold given by a CART model are eliminated and, say, only m variables are left. For note: at the LASSO procedure, variables get scaled in advance to select the most informative predictors with regard to their regression estimates β, VSURF, however, doesn't require this step.
Next, VSURF includes a variable selection step for a) interpretation and b) prediction. In the interpretation step, nested random forests starting with k = 1 to the m previously selected variables are created and the variables of the RF with the smallest OOB error rate are selected.
For prediction, be $\tilde{m} \leq m$ the number of variables in the random forest with the smallest OOB error rate from the interpretation step, sorted in ascending order of importance. Then, in a stepwise way, random forests are constucted including predictors if the OOB error decrease as long as their inclusion leads to a decrease in the OOB error that is "significantly greater than the average variation obtained by adding noisy variables" [159]. As in the case of logistic regression (Section 3.3.1), the final model is based on the variables added up to the last step.

3.3.7 Score development, step 1: clinical data

Models for DGF and death-censored transplant loss within one year were trained with stepwise logistic regression using the likelihood-ratio test on the subsets of eligible predictors identified in previous Sections 3.3.3 to 3.3.6. As the reference model, Section 3.3.1, and the complete case model of Section 3.3.2 already based on logistic regression, no further modeling was necessary for step 1 of the score building.

Model performance of the derived multivariable logistic regression models was assessed statistically and graphically by discrimination (C statistic, equivalent to the area under the receiver operating characteristic curve, AUC), calibration (calibration intercept and slope) and overall (Scaled Brier score, which is the Brier score scaled by its maximum Brier score [$Brier_{scaled} = 1 - Brier/Brier_{max}$])[162] and estimated calibration index (ECI) for each model on the validation dataset. The decision on the final model was based on these statistics and clinical judgement.

Regression estimates of the independent variables $\beta^T_{X,DGF}$ and $\beta^T_{X,1y-tl}$ and baseline intercepts $\beta_{0,DGF}$ and $\beta_{0,1y-tl}$ of the "final" models with m selected predictors were then extracted and used to predict the individual probability of each outcome. The calculated probabilities were then transformed into an integer point scale using the "Framingham Study risk score function" [163, 164] as a template.

First of all, continuous variables X_i, $i = \{1, ..., l\}, l \leq m$ were transformed into $j = 1, ..., k_i$ meaningful categories C_{ij} and the mid-point of each category chosen as a reference value W_{ij}. One of the k_i categories with midpoint $W_{i,ref}$ was chosen a reference $C_{i,ref}$.

Factor variables were recoded numerically and the reference value set to 0, e.g. in case of the recipients CVM-IgG positivity, the "negative"-class was recoded to 0 and "positive" to 1. Next, the distance between each category and the reference category was calculated and weighted by the assigned regression coefficient: $\beta_i(W_{ij} - W_{i,ref})$.

The set constant B, corresponding to one score point, was chosen to reflect the increase in risk associated with a 4-hour increase of cold ischaemia time: $B = \beta_{CIT} * 4$. To determine the points associated with each risk factor, the weighted distance was divided by B: $Points_{ij} = \beta_i(W_{ij} - W_{i,ref})/B$.

Given the sum of the points for each predictor X_j and the intercept, β_0, risk estimates can be derived using formula 3.4. Adding the values which are considered the reference values for the continuous predictors $W_{i,ref}$, $i = 1, ..., l$ multiplied by the regression coefficients, the sum $\sum_{i=0}^{p} \beta_i X_i$ can be approximated by $\beta_0 + \sum_{i=1}^{m} \beta_i X_i \approx \beta_0 + (\sum_{i=1}^{l} \beta_{i,ref} * W_{i,ref}) + B*$(Points total).

3.3.8 Histological evaluation

All biopsies included in this study were scored on two level sections by an experienced nephropathologist (PD Dr. med. Jan Ulrich Becker, Institute of Pathology, University Hospital of Cologne) according to Banff 2018 [100]. At the timepoint of scoring, the nephropathologist was blinded for the outcomes after transplantation. The Banff lesion scores assess histopathological changes in the tree different functional compartments of kidney biopsies: the interstitial, vascular-glomerular and tubular compartiment. Each score has scales of four to six degrees of change, beginning with "no change" and consider interstitial inflammation "i", tubulitis "t", glomerulitis "g", intimal arteritis "v", peritubular capillaritis "ptc", total inflammation "ti", inflammation in area of interstitial fibrosis and tubular atrophy "i-IFTA", staining for C4d, "C4d"), double contour "cg", mesangial matrix expansion "mm", arteriolar hyalinosis "ah", hyaline arteriolar thickening "aah", vascular fibrous intimal thickening "cv", interstitial fibrosis "ci" and tubular atrophy "ct". Combined with additional diagnostic parameters, the Lesion Scores are extended to the Banff diagnostic

categories like acute T cell-mediated rejection (TCMR) grade, active and BK-virus nephropathy (ABMR). These are, however, not used here. Besides the Banff lesion scores i, t, v, g, ptc, ci, ct, cv, cg, mm and ah, the number of glomeruli, the number of globally (>50% of the tuft) sclerosed glomeruli, the ratio thereof, the presence of thrombotic microangiopathy with platelet, fibrin or mixed thrombi in preglomerular or glomerular location and acute tubular necrosis (ATN) with detached, cytoplasmic fragments devoid of nuclei are assessed.

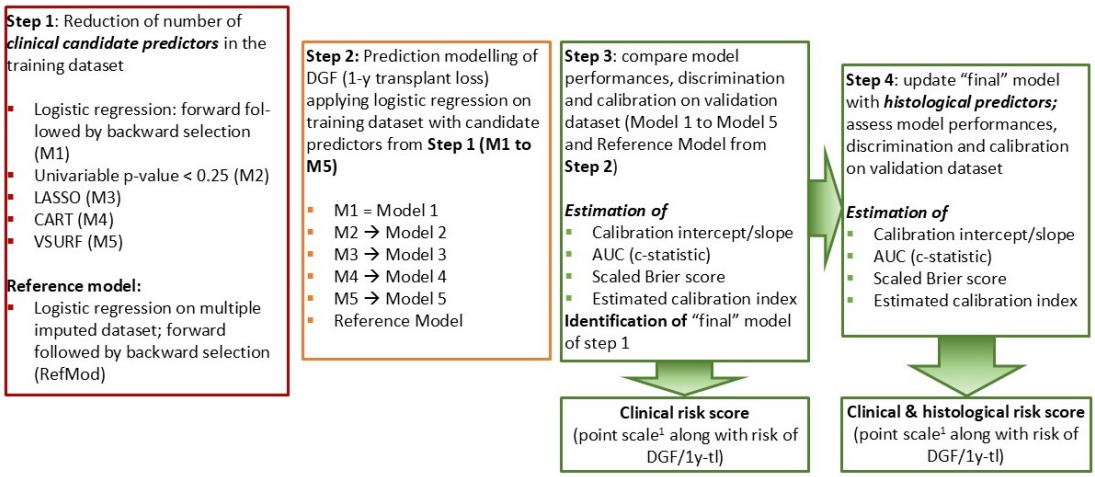
All scoring was done with whole slide images (WSIs) loaded into QuPath [165] on a 24 inch Eizo screen (Eizo Europe, Mönchengladbach, Germany). For the number of glomeruli, all sclerosed and non-sclerosed glomeruli and tufts dislodged from the biopsy cores were counted, empty Bowman's capsules (without any discernible cellular or matrix tuft elements) were disregarded [30].

3.3.9 Score development, step 2: update with histological findings

In the literature, several methods for model recalibration, updating and extending are described [166, 167, 168]. Here, the assumptions are that 1. the models derived after the first step are nested in the models from the second step and 2. the histological variables used for extension are independent of the predictors included in the models derived after step 1. In step 2, prior odds ($\beta_{0,DGF}+\beta_{X,DGF}^{T}$X and $\beta_{0,1y-tl}+\beta_{X,1y-tl}^{T}$X) of the "final" models derived in step 1 together with the eligible histological parameters were entered into a second multivariable stepwise logistic regression selection and modelling step, as described by Grill et al. [166]. How to actually quantify the value of a new biomarker, like the ones derived by histological findings, is a point of discussion and depends on the application as well as the measure of the outcome predicted. Performance measures include the already described changes in the AUC, ROC curves or c-statistics (Section 3.3.7) to evaluate discrimination, changes in the AIC or BIC (Section 3.3.2), which is related to changes in the likelihood function and the (scaled) Brier score (Section 3.3.7) as a category-free measure. As the outcomes DGF and 1y-tl are both binary, a measures of risk reclassification, the net reclassification improvement (NRI) was estimated as an alternative [169].

Additionally to an update on the entire training dataset, subsets to define an intermediate risk group were also considered as following: for the final models after step one, the probabilities of DGF and 1y-tl were estimated for each recipient according to equation 3.4. In the training dataset, the ordered predicted probabilities of each outcome were divided into 20 percentiles. Pairwise combinations of the derived cut-offs served as intervals defining a potential intermediate risk group on the training dataset. Modelling was then done first on the full training datasets and then

additionally on all subsets containing the training datasets' prevalence of DGF and 1y-tl, respectively. If any histological predictor showed a statistically significant association with the outcome of interest on either a subset or the full training dataset, the model from step 1 was updated. A graphical overview of the steps is given in Figure 3.5[30]. Results of the two steps are given as logistic regression models and simplified point-scores (Figures 5.24[30] and 5.25[30]).



Abbreviations: ***LASSO***, least absolute shrinkage and selection operator; ***CART***, classification and regression tree; ***VSURF***, variable selection using random forest; ***DGF***, delayed graft function; ***1y-tl***, one-year transplant loss; ***AUC***, area under the receiver operating characteristic curve; ***M1*** to ***M5***, subset of predictors for model 1 to model 5

[1] LM Sullivan, JM Massao, RB D'Agostino Sr., "TUTORIAL IN BIOSTATISTICS Presentation of multivariate data for clinical use: The Framingham Study risk score functions"; Statist. Med. 2004; 23:1631-1660 (DOI: 10.1002/sim.1742)

Figure 3.5: Flow of modelling steps [30]

3.3.10 Validation

With regard to generalisability, validation of a predictive model is crucial. In validation, a distinction is made between three scenarios: internal, internal-external and external validation [170].

Internal validation refers to methods that use the same data available for model development to predict performance on new observations from the same underlying cohort [171]. It can be achieved in different ways. One method is by applying *bootstrapping*, ideally in all steps (variable selection, modelling, assessment of discrimination and calibration) of the modelling process. Particularly in datasets with small sample sizes compared to number of candidate predictors, overfitting can be a problem. Bootstrapping quantifies the optimism of the derived model [172].

A second option would be the *split-sample validation* [170]. Here, the given is divided into a training and a test (hold out) dataset, e.g. by 50%/50% or 70%/30%. However, this method is not recommended as it can only be used on very large datasets, but then overfitting is not an issue and internal validation is not required.

Lastly, *apparent validation* on the entire training can be applied. The performance of a model, when being tested on the data the model was trained on, runs the risk to be overly optimistic and is therefore not recommended.

The resulting models for both outcomes after both steps were validated after derivation of the final models on an independent validation dataset. This was done as described in section 3.3.7 first graphically by plotting both the ROC curves of the predicted outcomes and the calibration plots of the predicted vs. observed probabilities of each outcome. Second, calibration and discrimination were assessed by estimation of the c-statistics (AUC), the scaled Brier scores and the estimated calibration index.

3.3.11 Comparison with established scores

As described in Chapter 2, several models to predict delayed graft function or, respectively, transplant loss within one year after transplantations, were already published. However, besides applicability, their prominence and acceptance varies. External validation and recalibration on new cohorts is one method to check applicability and gain confidence in the model. The scores by Irish (2003 [60] and updated 2010 [62]) for the outcome DGF were evaluated comparatively with the scores by Jeldres [61], Chapal [64] and Zaza [39] from Decruyenare [32] and Kers [173] on Belgian cohorts, and Zhang [174] in China. In each of the validation cohorts, generalisability was declared to be best for the score by Irish (AUCs: .78; .761 and .737, respectively). Further, the scores generalisability was explored by Grossberg in the area of Rhode Island (USA), who compared mean nomogram values amongst recipients with and without DGF (.45 vs. .40) [175]; Kaisar in Australia [176], Michalak in Belgium [177], Ietto [178] in Italy, and Rodrigo in Spain [179] (AUCs: .76, .69, .693 and .704 respectively). Again, the scores by Irish performed best with regard to the AUC values when other prediction models were also evaluated. These are just a few examples of attempts to externally validate the Irish nomogram, but the results indicate a good generalisability.

Here, the scores by Balaz [63], Irish (2010) [62], Chapal [64] and Jeldres [61] are externally validated by assessing their accuracy with regard to prediction of DGF on the training and validation datasets. Of note: with regard to missing variables in the external validation dataset, it is advisable not to calculate a 'reduced' model with the remaining variables and the coefficients of the complete model, as these were adjusted for all of the other variables unless adjusted estimands of such sub-models are provided [180]. This advice is usually neglected for the predictor 'race' in the model of Irish, when validated in studies performed outside the USA. Accordingly, in this study, the race for all recipients is set to 'non-black' for all recipients.

With regard to transplant loss within the first year, there is no such clear ac-

ceptance or preference for any of the published models in Section 2.2, and external validation is largely lacking. For this reason, existing regression-based models were externally validated on our training and validation data and compared with the results of the new 2-Step scores.

One prerequisite for an proper external validation is that the predictors used are present in the data and that the models are adequately described. For example, in a logistic regression model, the intercept should also be specified as the 'baseline probability' in order to be able to correctly determine the outcome probabilities using this and the regression coefficients. Although such a constant value does not have an influence on the calculation of the AUC, the interpretation of the prediction is ultimately only possible in relation to other observations and no longer absolutely. Derivation of correct AUC values is more difficult in case of a missing baseline hazard as a function over time in a Cox regression, if no parametric form is available. Even more complicated is the external validation of machine-learning or deep-learning based predictors, unless tools are provided by the developer for calculating risks given complete datasets.

Yet, besides the score by Port [79], the hazard ratios of the recently published KTOP score by Miller et. al. with optional recipient parameters [37], the Leuven score by De Vusser et. al. [102] and the clinico-pathological score by Snoeijs [103] predicting 5-year graft survival from old ($>$ 60 years) donors are used for comparison of the accuracy of 1y-tl prediction.

Chapter 4

datasets

4.1 Assessment of the training and external validation datasets

4.1.1 Sample size estimation

Based on the number of predictors used in published scores predicting DGF and, respectively, graft loss and the variables Eurotransplant collects on regular basis, 30-60 variables were expected to be included into the modelling step 1 [181, 182, 183, 184]. However, in order for the final model to be applicable in practice, and also to avoid overfitting, the final model after step 2 should be restricted to no more than 10 predictors and their interactions. We expected the incidence of DGF within the first seven days after transplantation to amount to 30%. Taking into account missing values, according to the "10 events per variable rule of thumb", the minimal sample size required to evaluate this number of variables in a multivariable binary logistic regression analysis was 1000 biopsies [185]. According to the proposal submitted to the DFG, the biopsies were to be provided in equal parts by the five study centres, i.e. 200 per centre.

No separate sample size estimation was carried out for the second outcome of interest, 1y-tl, or the external validation cohort.

4.1.2 Source of training and external validation dataset

The training based on retrospectively collected data of five Eurotransplant centres (Antwerp [Belgium], Cologne [Germany], Szeged [Hungary], Vienna [Austria], Zagreb [Croatia]). First, pariffin-embedded zero-hour biopsies were compiled and scanned with a NanoZoomer slide scanner of the manufacturer Hamamatsu Photonics Germany with a x40 objective to a resolution of 0.23 μm per pixel in either Rotterdam, the Netherlands, or Cologne, Germany [30]. Second, after a first screening for eligibility by an experienced nephropathologist, the corresponding clinical donor, recipient and transplant data were provided by Eurotransplant.

The validation was composed of two datasets. One collection of retrospectively collected procurement biopsies along with donor and recipient data from deceased donors with marginal kidney quality provided by the German Organ Procurement Organization [German: Deutsche Stiftung Organtransplantation] (DSO-set)[186]. In this sample, transplantations were performed from October 2008 onwards between 2008 and 2012 in several German centres. The second consists of prospectively collected zero-hour biopsies from Ljubljana (Slovenia) and Cologne with corresponding clinical data from Eurotransplant.

For both training and validation datasets, biopsies were eligible for screening if the transplantation was performed between January 2008 and December 2019 and if donors were deceased. Before screening, biopsies of tansplantations after cardiac death or en block transplantations were excluded. If both kidneys from a single donor have been transplanted into different recipients in one centre, the right kidney was excluded from analysis; if the identification as right or left was not documented, both were excluded. Also, only transplants from adult ($\geq$ 18 years) donors with adult recipients were accepted.

During screening, biopsies with insufficient tissue according to Banff 2018 [100], i.e. less than 1 artery and/or 12 glomeruli, were excluded.

Chapter 5

Results

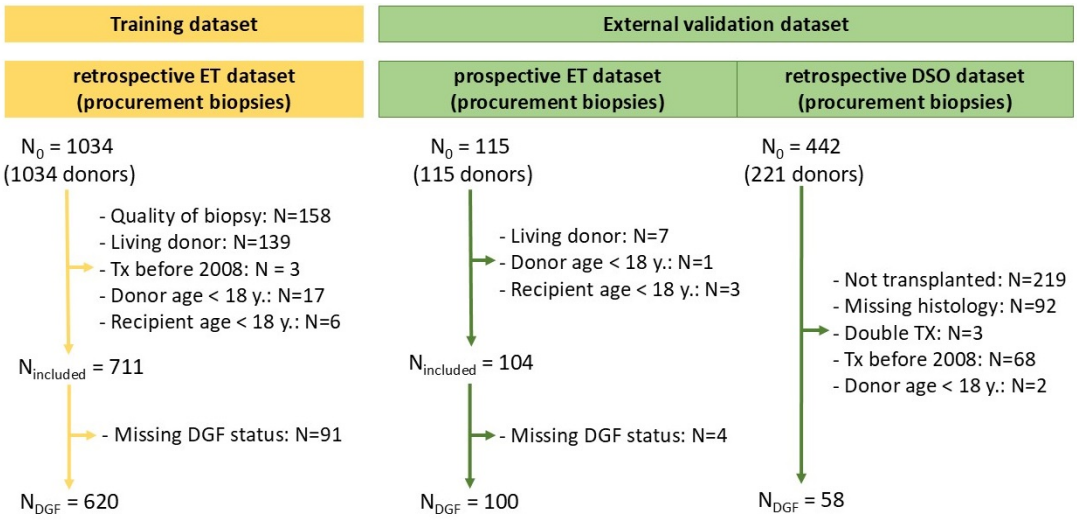
5.1 Assessment of training and validation dataset

5.1.1 Flow of training- and validation dataset definition

The training cohort based of 1034 offered zero-hour biopsies from 1034 donors. Of these, 869 were eligible for screening, 165 cases met the exclusion criteria. Due to deficiencies in the quality of the Periodic acid-Schiff (PAS) sections, 158 slides were excluded. The remaining 711 biopsies were eligible for the modelling of the 1y-tl-score (Fail-set). Delayed graft function status was available for 620 recipients defining the DGF-score training dataset (DGF-set).

The first subset defining the validation cohort consisted of 115 zero-hour biopsies prospectively collected in two study centres. Of these, 104 were eligible for screening and, after screening, the analysis of 1y-tl. Due to missing DGF-status, 100 recipients were included in the DGF validation dataset.

The second subset (DSO-set) based on 442 kidneys from 221 donors. Of these, 219 kidneys were not transplanted, in 92 cases no samples to assess histological data was available. Due to predefined exclusion criteria, 72 more kidneys were excluded before screening. As a result, the prospectively collected procurement biopsies were supplemented by procurement biopsies from 58 kidneys representing 47 different donors. Overall, the validation dataset for 1y-tl contains of 162 samples (Fail-val); the DGF-scores were validated on data of 158 recipients (DGF-val). Flow of donor kidney exclusion see Figure 5.1[30].



***Abbreviations**: **y**, years; **DGF**, delayed graft function; **TX**, transplantation; **N** number of observations; **ET**, Eurotransplant; **DSO**, Deutsche Stiftung Organtransplantation [German Organ Procurement Organization]*

Figure 5.1: Flow of dataset creation [30]

5.1.2 Description of cohorts by centre and outcome

As shown in Table 5.1[30], although equal distributions of recipients amongst study centres were planned, the final contributions varied between N=44 (6%) and N=442 (62%) in the seven centres contributing to the training dataset and N=6 (3.7%) and 98 (60.5%) in the three centres contributing to the validation dataset for the analysis of 1y-tl.

Distributions were not proportional to the population figures of the countries in which the centres are located. In the training dataset, prevalence of 1y-tl was 9% (N=61) whereas prevalence of DGF among the 620 recipients with available DGF-status was 27% (N=166). The notification of a delayed graft function to the Eurotransplant registry is obligatory. After contacting all transplant centres to update this information, DGF-status was unavailable for N=91 (13%) of the recipients. In the validation dataset, prevalences of DGF and 1y-tl did not differ significantly with 27% (46) and 7% (N=11), respectively. DGF-status was missing in 3% (N=4).

In the training dataset, not only the distribution of participants by study centre varied, but also the outcomes, as is sum up in Table 5.1[30]. Compared to centre no. 2, which was chosen as the reference with N=105 observations and 31 cases with DGF (odds = 31:74), the odds for DGF in the first centre was significantly lower with an odds ratio (OR) value of .1 and p = .002, while the OR of centre no. 5 was 1.91 compared to the reference. The odds for DGF, however, differed significantly only in the first centre; the values for centres 3 and 4 (OR = .46 and .98 respectively) did not differ significantly. Looking at the odds for transplant loss in the first year, there were no significant differences between the values of centre no. 2 (odds = 11:101)

and the other four centres.

Table 5.1: Outcomes and study centres by training and validation dataset for recipients with available 1y-tl (Fail-set and Fail-val)[30]

Variables	**Overall**	**Training** Fail-set	**Validation** Fail-val	**p**
N	873	711	162	
1y-tl [Yes] (N(%))	72 (8.2)	61 (8.6)	11 (6.8)	.556
DGF (N(%))[1]				
No	566 (64.8)	454 (63.9)	112 (69.1)	.624
Yes	212 (24.3)	166 (23.3)	46 (28.4)	
Unknown	95 (10.9)	91 (12.8)	4 (2.5)	
Study centre (N(%))				
Centre 1	50 (5.7)	50 (7.0)	0 (0.0)	NA
Centre 2	210 (24.1)	112 (15.8)	98 (60.5)	
Centre 3	63 (7.2)	63 (8.9)	0 (0.0)	
Centre 4	442 (50.6)	442 (62.2)	0 (0.0)	
Centre 5	44 (5.0)	44 (6.2)	0 (0.0)	
Centre 6	58 (6.6)	0 (0.0)	58 (35.8)	
Centre 7	6 (0.7)	0 (0.0)	6 (3.7)	

Abbreviations: DGF, delayed graft function; 1y-tl, death-censored transplant loss within one year

[1] P-value estimation excluding cases with unknown DGF-status

The ORs were between .30 (centre 3, p=.126) and .99 (centre 4, p=.976), whereby the absolute values of the observed cases were very low at < 5, particularly in centres 1, 3 and 5. As expected from the literature, both the odds for transplant loss within one year after transplantation and death within the same period were positively associated with DGF (ORs [95%-CI] = 4.2 [2.36-7.49], p<.001 and 2.08 [.91-4.59], p=.073, respectively). As defined in the introduction, recipients dying within the first year after transplantation are regarded as "censored at the time point of death" in the analysis of transplant loss within the first year. However, if death within the first year is seen as the outcome variable, the 30 observed cases were significantly associated with previous transplant loss OR [95%-CI] = 12.69 [5.87-27.46], p<.001.

Table 5.2: Odds ratios of transplant outcomes by recipient survival and transplant centre in the training dataset [30]

Variable	No DGF	DGF	Univariable OR (95% CI, p)	1y-tl = no	1y-tl = yes	Univariable OR (95% CI, p)
N	***454***	***166***		***650***	***61***	
Deceased within one year (N(%))	15 (3.3)	11 (6.6)	2.08 (0.91-4.59, p=.073)	15 (2.3)	15 (24.6)	12.69 (5.87-27.46, p<.001)
1y-tl = yes (N(%))	23 (5.1)	33 (19.9)	4.20 (2.36, 7.49, p<.001)			
Study centre (N(%))						
Centre 1	48 (10.6)	2 (1.2)	0.10 (0.02-0.35, p=.002)	48 (7.4)	2 (3.3)	0.38 (0.06-1.50, p=.223)
Centre 2	74 (16.3)	31 (18.7)	1	101 (15.5)	11 (18.0)	1
Centre 3	52 (11.5)	10 (6.0)	0.46 (0.20-0.99, p=.055)	61 (9.4)	2 (3.3)	0.30 (0.05-1.17, p=.126)
Centre 4	260 (57.3)	107 (64.5)	0.98 (0.61-1.60, p=.942)	399 (61.4)	43 (70.5)	0.99 (0.51-2.08, p=.976)
Centre 5	20 (4.4)	16 (9.6)	1.91 (0.87-4.17, p=.104)	41 (6.3)	3 (4.9)	0.67 (0.15-2.28, p=.557)

Abbreviations: OR, odds ratio; y, year; p, p-value; CI, confidence interval; DGF, delayed graft function; 1y-tl, death-censored transplant loss within the first year

NOTE: The odds ratio as a result of, for example, logistic regressions and case-control studies, and the relative risk aka. risk ratio, are not to be used interchangeably. As an example, Table 5.3 shows a simple 2x2 cross-table of an outcome (alive vs. deceased) by treatment vs. control. The odds for being deceased after treatment are d/c, after control b/a. Odds, therefore, can range from 0 to ∞. The risk for being deceased after treatment, however, is d/(c+d) and after control b/(a+b). Risks obviously can range from 0 to 1. If only a single predictor is analysed, the odds can be converted to the risk by "risk = odds/(1+odds)", which is the average relative risk. Given the odds ratio of dying in the treatment vs. control group, one further needs the baseline risk in the control group (probability p_0 of dying in the control group) to derive the risk ratio "relative risk = odds ratio/(1-p_0+(p_0*odds ratio))".

However, in multivariable adjusted logistic regression, with more than one independent variable of interest, averaging of the risk is more complicated. To calculate individual risks ("marginal effects"), given a logistic regression model, the R package "effects" [187] can be used.

Group	Alive	Deceased	**Sum**
Control	a	b	a+b
Treatment	c	d	c+d
Sum	a+c	b+d	a+b+c+d

Table 5.3: Example: 2x2 cross-table

5.2 Baseline characteristics of donors, recipients and nephropathological evaluation by training and validation data set

5.2.1 Donors

As sum up in Table 5.4[30], median donor age in the training cohort was 54 years (IQR: 44-64), 46% were female and the mean (SD) BMI before transplantation amounted to 26.6 (4.3) kg/m^2. Status of diabetes mellitus and hypertension were known for 72% and 79% of the donors. Amongst them, prevalences were 8% and 47%, respectively. Type of brain death (primary vs. secondary) was identifiable for 79% with a majority of them (N=555; 87%) showing primary brain death. The most common known causes of death were cerebrovascular accidents (69%), followed by traumatic brain injuries (N=93, 17%). At the last measurement before explantation, serum creatinine levels averaged to mean = 0.9 (SD=0.34) mg/dl, while mean (SD) of eGFR according to Cockroft-Gault amounted to 110.28 (51.6) ml/min.

Compared to the training cohort, donors of the validation cohort showed a significantly increased median age (60 years, IQR: 45-70) and mean BMI (28 kg/m^2). Before transplantation, average last serum creatinine levels significantly increased (1.5 mg/dl, SD=1.43), while mean Cockroft-Gault eGFR decreased (88.5 ml/min, SD=54.8). Grouping of cause of death in the DSO-set differed from how causes are sum up in Eurotransplant.

No cases of apnea, bacterial meningitis, drug suicide and primary brain tumor were observed in the validation dataset while there were no cases of subarachnoid haemorrhage in the training data set. About 50% of the donors cause of death (primary vs. secondary and causes of death) were missing. Most of the observed deaths were caused by cerebrovascular accidents (N=54, 68%).

Table 5.4: Donor characteristics of training and validation dataset for recipients with available 1y-tl (Fail-set and Fail-val)[30]

Variables	**Overall**	**Training** Fail-set	**Validation** Fail-val	**p**
N	873	711	162	
Donor age [y] (median [IQR])	55 [44, 65]	54 [44, 64]	60 [45, 70]	.003*
Donor sex [female] (N(%))	410 (47.0)	329 (46.3)	81 (50.0)	.441
Donor BMI [kg/m²] (mean (SD))	26.49 (4.81)	26.20 (4.34)	27.76 (6.36)	<.001
Donor diabetes mellitus (N(%))				
No	541 (62.0)	470 (66.1)	71 (43.8)	<.001
Yes	62 (7.1)	42 (5.9)	20 (12.3)	
Unknown	270 (30.9)	199 (28.0)	71 (43.8)	
Donor hypertension (N(%))				
No	331 (37.9)	298 (41.9)	33 (20.4)	<.001
Yes	339 (38.8)	262 (36.8)	77 (47.5)	
Unknown	203 (23.3)	151 (21.2)	52 (32.1)	
Donor creatinine, last [mg/dl] (mean (SD))	1.04 (0.79)	0.94 (0.49)	1.50 (1.43)	<.001
Donor Cockcroft-Gault eGFR, last [ml/min] (mean (SD))	106.23 (52.84)	110.28 (51.58)	88.51 (54.78)	<.001
Type of brain death (N(%))				
Primary	555 (63.6)	479 (67.4)	76 (46.9)	<.001
Secondary	86 (9.9)	80 (11.3)	6 (3.7)	
Unknown	232 (26.6)	152 (21.4)	80 (49.4)	
Cause of death (N(%))				
Apnea	17 (1.9)	17 (2.4)	0 (0.0)	
Bacterial meningitis†	4 (0.5)	4 (0.6)	0 (0.0)	NA
Cardiac arrest	33 (3.8)	31 (4.4)	2 (1.2)	.098
Cerebrovascular accident	433 (49.6)	379 (53.3)	54 (33.3)	<.001
Drug suicide†	2 (0.2)	2 (0.3)	0 (0.0)	NA
Primary brain tumor†	1 (0.1)	1 (0.1)	0 (0.0)	NA
Subarachnoid haemorrhage†	14 (1.6)	0 (0.0)	14 (8.6)	<.001
Trauma	32 (3.7)	23 (3.2)	9 (5.6)	.235
Traumatic brain injury	94 (10.8)	93 (13.1)	1 (0.6)	<.001
Unknown	243 (27.8)	161 (22.6)	82 (50.6)	<.001

Abbreviations: eGFR, estimated glomerular filtration rate; SD, standard deviation; BMI, body-mass-index; IQR, interquartile range; p, p-value

* Non-parametric Mann-Whitney U-test

† Excluded before variable selection

5.2.2 Recipients

In the training cohort, recipients' median age was 55 (IQR: 45-64) years, 36% (N=257) were female and the mean (SD) BMI was 26.2 (4.98) kg/m^2 (Table 5.5[30]). About 10% (N=68) of the recipients had previous kidney transplantations, 39% (N=277) of the recipients were positive for CMV antibodies IgG (CMV IgG). However, this status was unknown in 41% (39% in recipients with known DGF status). The median dialysis vintage was 3.1 years (IQR 1.9-4.9). Most recipients were diagnosed with "glomerulonephritis (GN) or acquired glomerulopathy" (N=175, 25%) as primary disease responsible for ESRD. Due to their small number of observed cases, "Congenital dysplasia/ hypoplasia/ malformation with/ without urinary tract malformation", "Drug-induced/toxic", "Nephrocalcinosis/ nephrolithiasis", "Subarachnoid haemorrhage" and "Tumour/trauma/ surgery" were excluded from the analyses after dummy encoding ESRD. The number of unknown or missing information about an ESRD was high, with 22% and 23%, each.

In the validation cohort, recipient median age (58 years, IQR: 50-67), median dialysis vintage (8.8 years) and prevalence of CMV IgG positivity (60.5%) differed significantly. 37% of the recipients were female, slightly more than in the training data set. The mean BMI was almost identical (25.7 kg/m^2). With 37% and, respectively, 51%, the percentage of unknown and missing information on ESRD was high. In both the training and validation cohort, most diagnoses were related to the glomerular compartiment. The percentage of recipients with "genetic nephropathy/ glomerulopathy non-focal segmental glomerulosclerosis (FSGS)" was 15.6% and 14.2%, for "glomerulonephritis (GN) /acquired glomerulopathy" 25% and 24%, respectively.

Table 5.5: Recipient characteristics of training and validation dataset for recipients with available 1y-tl (Fail-set and Fail-val)[30]

Variables	**Overall**	**Training** Fail-set	**Validation** Fail-val	**p**
Recipient age [y] (median [IQR])	55 [46, 65]	55 [45, 64]	58 [50, 67]	.004*
Recipient sex [female] (N(%))	317 (36.3)	257 (36.1)	60 (37.0)	.903
Recipient BMI [kg/m²] (mean (SD))	26.10 (4.99)	26.21 (4.98)	25.69 (5.03)	.238
Dialysis vintage [y] (median [IQR])	3.61 [2.02, 5.94]	3.13 [1.86, 4.85]	8.77 [4.61, 13.32]	<.001*
Previous transplantations [yes] (N(%))	78 (8.9)	68 (9.6)	10 (6.2)	.225
Recipient CMV IgG (N(%))				
Negative	197 (22.6)	142 (20.0)	55 (34.0)	<.001
Positive	375 (43.0)	277 (39.0)	98 (60.5)	
Unknown	301 (34.5)	292 (41.1)	9 (5.6)	
Recipient PRA peak historic (median [IQR])	0 [0, 4]	0 [0, 5]	0 [0, 3]	.093*
ESRD (N(%))				
Congenital dysplasia /hypoplasia/malformation with/without urinary tract malformation†	8 (0.9)	7 (1.0)	1 (0.6)	NA
Diabetic nephropathy	115 (13.2)	98 (13.8)	17 (10.5)	.323
Drug-induced/toxic†	7 (0.8)	7 (1.0)	0 (0.0)	0.435
FSGS	31 (3.6)	25 (3.5)	6 (3.7)	>.900
Genetic nephropathy/ Glomerulopathy non-FSGS	134 (15.3)	111 (15.6)	23 (14.2)	.741
GN/Acquired Glomerulopathy	213 (24.4)	175 (24.6)	38 (23.5)	.835
Hypertensive nephropathy	54 (6.2)	50 (7.0)	4 (2.5)	.046
Nephrocalcinosis	2 (0.2)	1 (0.1)	1 (0.6)	.814
Pyelonephritis/Interstitial nephritis/obstructive uropathy/ reflux uropathy	47 (5.4)	41 (5.8)	6 (3.7)	.237
Renal vascular disease	28 (3.2)	24 (3.4)	4 (2.5)	.731
TMA/Systemic sclerosis	11 (1.3)	9 (1.3)	2 (1.2)	>.900
Tumour/trauma/surgery†	5 (0.6)	5 (0.7)	0 (0.0)	.622
Unknown cause†	218 (25.0)	158 (22.2)	60 (37.0)	<.001
Subarachnoid haemorrhage†	14 (1.6)	0 (0.0)	14 (8.6)	<.001
Trauma	32 (3.7)	23 (3.2)	9 (5.6)	.235
Traumatic brain injury	94 (10.8)	93 (13.1)	1 (0.6)	<.001
Unknown	243 (27.8)	161 (22.6)	82 (50.6)	<.001

Abbreviations: CMV IgG, cytomegalovirus antibody IgG; PRA, panel reactive antibody; ESRD, end-stage renal disease; SD, standard deviation; BMI, body-mass-index; IQR, interquartile range; FSGS, focal segmental glomerulosclerosis; TMA, thrombotic microangiopathy; p, p-value

* Non-parametric Mann-Whitney U-test

† Excluded before variable selection

5.2.3 Transplantation

As already mentioned in Section 2.4 on the Eurotransplant allocation system, in addition to the blood group, the HLA-A, -B and DR mismatches in particular are taken into account when assigning donor organs to potential recipients. Accordingly, the number of missing values (N=1) is low among these influencing factors. Most frequently observed in both the training and validation data sets was a single mismatch among the HLA-A, -B, and -DR typings. These accounted for 53% and 49% and 55% in the training dataset and 54% and 57% and 52% in the validation cohort, respectively. The sum of the HLA-A, -B and -DR mismatches ranged from 0 to 6 and was approximately symmetrical around the three, which was observed in about 30% of the donor-recipient combinations. Significant differences in distribution between the training and validation cohorts were only observed for HLA-A and -DR, where a significantly higher proportion of two mismatches was observed in the validation data set. The median cold ischaemia time was comparable between the cohorts at approximately 11 and 10 hours, respectively while the warm ischaemia time in the validation dataset was significantly longer, at 36 to 32 min. As the warm ischaemia time is not available at the time of allocation, this variable was excluded before variable selection based on statistical methods began. Results are sum up in Tables 5.6 and 5.7 [30].

Table 5.6: Transplant characteristics of training (N=711) and validation (N=162) dataset (Fail-set and Fail-val, part 1)[30]

Variables	**Overall**	**Training** Fail-set	**Validation** Fail-val	**p**
N	873	711	162	
HLA mismatches A (N(%))				
None	271 (31.0)	229 (32.2)	42 (25.9)	.247
One	460 (52.7)	373 (52.5)	87 (53.7)	
Two	141 (16.2)	108 (15.2)	33 (20.4)	
Unknown	1 (0.1)	1 (0.1)	0 (0.0)	
HLA mismatches B (N(%))				
None	156 (17.9)	137 (19.3)	19 (11.7)	.106
One	444 (50.9)	351 (49.4)	93 (57.4)	
Two	272 (31.2)	222 (31.2)	50 (30.9)	
Unknown	1 (0.1)	1 (0.1)	0 (0.0)	

Abbreviations: HLA, humane leucozyte antibody; p, p-value.

Table 5.7: Transplant characteristics of training (N=711) and validation (N=162) dataset (Fail-set and Fail-val, part 2)[30]

Variables	Overall	Training Fail-set	Validation Fail-val	p
HLA mismatches DR (N(%))				
None	244 (27.9)	207 (29.1)	37 (22.8)	.021
One	477 (54.6)	393 (55.3)	84 (51.9)	
Two	151 (17.3)	110 (15.5)	41 (25.3)	
Unknown	1 (0.1)	1 (0.1)	0 (0.0)	
HLA mismatches [sum] (N(%))				
0	82 (9.4)	73 (10.3)	9 (5.6)	.08
1	67 (7.7)	60 (8.4)	7 (4.3)	
2	156 (17.9)	125 (17.6)	31 (19.1)	
3	270 (30.9)	216 (30.4)	54 (33.3)	
4	192 (22.0)	159 (22.4)	33 (20.4)	
5	78 (8.9)	59 (8.3)	19 (11.7)	
6	27 (3.1)	18 (2.5)	9 (5.6)	
Unknown	1 (0.1)	1 (0.1)	0 (0.0)	
Cold ischaemia time [h] (median [IQR])	11 [0, 15.78]	11.27 [0, 16.03]	10.37 [6.25, 14.07]	.245*
Warm ischaemia time [min] (median [IQR])†	33 [0, 45]	32 [0, 45]	36 [22, 45.75]	.012*

Abbreviations: HLA, humane leucozyte antibody; h, hours; IQR, interquartile range; p, p-value; min, minutes.

* Non-parametric Mann-Whitney U-test

† Excluded before variable selection

5.2.4 Nephropathological evaluation

As described in section 2.3, nephropathological evaluation and scoring was done according to Banff 2018 [100] by a single, experienced nephropathologist. Results of the scoring and evaluation is sum up in Tables 5.9 and 5.8 [30].

The number of missing values was zero for all Banff lesion scores as availability of the biopsies was one of the inclusion criteria. In case of peritubular capillaritis (ptc) and intimal arteritis (v), no lesions were observed, so all biopsies were scored as pct0 and v0. Only eight cases with tubulitis, scored as t1 "foci with one to four leukocytes per tubular cross-section (or 10 tubular cells)" were observed in the validation data set, t0 otherwise. This was similar for Banff cg, double contour, with tree cases of cg1b "double contours affecting 26–50% of peripheral capillary loops in the most affected glomerulus" in each the training and validation dta set. Inflammation was observed solely in the validation dataset with one case of inflammation in 26–50% of unscarred cortical parenchyma (i2) and seven cases with inflammation in 10-25% (i1). Variation of percentage of lesions was higher in Banff g, ci, ct, cv, mm and ah where the whole range of scores was observed. Distributions in the training and

validation cohort were significantly different for vascular fibrous intimal thickening (cv), glomerulitis (g) and mesangial matrix expansion (mm) with more lesions in the validation data set. The percentage of biopsies with microthrombi was also higher in the validation dataset 4.9% vs. 2.3% as compared to the training data set.

Due to the small absolute number, this difference was not significant. Acute tubular necrosis, however, was significantly more often observed in the training dataset (48.4% vs. 38.3%, p = .049). The median [IQR] number of arteries and glomeruli varied significantly, with more arteries is the validation dataset 3 [2, 6] vs. 3 [2, 4], p = .005, and more glomeruli in the training dataset 29 [21, 36] vs. 27 [13, 29], p < .001. The mean ratio of globally sclerosed glomeruli was compareable with 3% in the training and 4% in the validation data set. Overall, one can say that the quality of kidneys donated was, according to the histological evaluation, worse in the validation dataset as compared to the training cohort.

This might be a result of including the DSO data into the validation cohort, as the sampling was based on donors with marginal kidneys. [186].

Table 5.8: Histological characteristics of training (N=711) and validation (N=162) dataset (Fail-set and Fail-val, part 1)[30]

Variables	**Overall**	**Training** Fail-set	**Validation** Fail-val	**p**
Banff ah (N(%))				
ah0	342 (39.2)	276 (38.8)	66 (40.7)	.053
ah1	376 (43.1)	314 (44.2)	62 (38.3)	
ah2	120 (13.7)	89 (12.5)	31 (19.1)	
ah3	35 (4.0)	32 (4.5)	3 (1.9)	
No. of arteries [median (IQR)]	3 [2, 4]	3 [2, 4]	3 [2, 6]	.005*
No. of glomeruli [median (IQR)]	26 [20, 35]	29 [21, 36]	27 [13, 29]	<.001*
No. of glomeruli with GSG [mean (SD)]	1.96 (3.99)	1.93 (3.57)	2.10 (5.48)	.627
Ratio number of glomeruli with GSG to all [median (IQR)]*	0.04 [0, 0.09]	0.03 [0, 0.09]	0.04 [0, 0.11]	.426*
Microthrombus (N(%))				
No	848 (97.1)	694 (97.6)	154 (95.1)	.151
Yes	24 (2.7)	16 (2.3)	8 (4.9)	
Unknown	1 (0.2)	1 (0.1)	0 (0)	
Acute tubular necrosis (N(%))				
No	465 (53.3)	365 (51.3)	100 (61.7)	.049
Yes	406 (46.5)	344 (48.4)	62 (38.3)	
Unknown	2 (0.2)	2 (0.3)	0 (0)	

Abbreviations: SD, standard deviation; GSG, globally sclerosed glomeruli; no., number; IQR interquartile range; p, p-value

* Non-parametric Mann-Whitney U-test

Table 5.9: Histological characteristics of training (N=711) and validation (N=162) dataset (Fail-set and Fail-val, part 1) [30]

Variables	Overall	Training Fail-set	Validation Fail-val	p
N	873	711	162	
Banff i (N(%))†				
i0	865 (99.1)	711 (100.0)	154 (95.1)	<.001
i1	7 (0.8)	0 (0.0)	7 (4.3)	
i2	1 (0.1)	0 (0.0)	1 (0.6)	
Banff t (N(%))†				
t0	865 (99.1)	711 (100.0)	154 (95.1)	<.001
t1	8 (0.9)	0 (0.0)	8 (4.9)	
Banff v [v0] (N(%))†	873 (100.0)	711 (100.0)	162 (100.0)	NA
Banff g (N(%))†				
g0	850 (97.4)	699 (98.3)	151 (93.2)	<.001
g1	13 (1.5)	4 (0.6)	9 (5.6)	
g2	8 (0.9)	7 (1.0)	1 (0.6)	
g3	2 (0.2)	1 (0.1)	1 (0.6)	
Banff ptc [ptc0] (N(%))†	873 (100.0)	711 (100.0)	162 (100.0)	NA
Banff ci (N(%))†				
ci0	768 (88.0)	628 (88.3)	140 (86.4)	.169
ci1	84 (9.6)	63 (8.9)	21 (13.0)	
ci2	19 (2.2)	18 (2.5)	1 (0.6)	
ci3	2 (0.2)	2 (0.3)	0 (0.0)	
Banff ct (N(%))				
ct0	540 (61.9)	430 (60.5)	110 (67.9)	.195
ct1	313 (35.9)	262 (36.8)	51 (31.5)	
ct2	18 (2.1)	17 (2.4)	1 (0.6)	
ct3	2 (0.2)	2 (0.3)	0 (0.0)	
Banff cv (N(%))				
cv0	321 (36.8)	272 (38.3)	49 (30.2)	.031
cv1	166 (19.0)	129 (18.1)	37 (22.8)	
cv2	180 (20.6)	135 (19.0)	45 (27.8)	
cv3	205 (23.5)	174 (24.5)	31 (19.1)	
Unknown	1 (0.1)	1 (0.1)	0 (0.0)	
Banff cg (N(%))†				
cg0	867 (99.3)	708 (99.6)	159 (98.1)	.069
cg1b	6 (0.7)	3 (0.4)	3 (1.9)	
Banff mm (N(%))				
mm0	858 (98.3)	707 (99.4)	151 (93.2)	<.001
mm1	10 (1.1)	2 (0.3)	8 (4.9)	
mm2	2 (0.2)	1 (0.1)	1 (0.6)	
mm3	3 (0.3)	1 (0.1)	2 (1.2)	

Abbreviations: p, p-value; N, number.

† Excluded before step 1 of the variable selection (Sections 3.3.1 to 3.3.6)

5.3 Pre-selection of potential clinical predictors

5.3.1 Selection based on completeness, correlation (VIF) and variance

Comparing predictors correlating with either DGF or 1y-tl with variables listed in the Eurotransplant database, thirty-five donor-, recipient- and transplant procedure-related variables were considered based on prior knowledge. Dummy encoding of the recipient's primary diagnosis (13 categories) and the donor's cause of death (9 categories) increased this number to fifty-five potential predictors.

Pre-transplantation donor and recipient variables included age in years (y), sex (male vs. female), body mass index (BMI) [kg/m^2] and IgG and IgM anti-cytomegalovirus (CMV) antibody. Donor history of diabetes mellitus and hypertension, serum creatinine [mg/dl], Cockroft-Gault eGFR [ml/min] and cause of death, as well as recipient dialysis vintage [years], previous transplantations, panel reactive antibody (in %) and end-stage renal disease along with their number of human leucocyte antigens HLA-A, -B and -DR mismatches were extracted from the Eurotransplant database.

Variables were excluded according to the criteria described in Section 3.2 and additionally, if they were not available before the decision to transplant was made. Both cold and warm ischaemia time were available. Despite its strong association with transplant outcome, warm ischaemia time was not included because this information is usually not known at the time of allocation. The variables donor and recipient CMV IgM - positivity and donor CMV IgG - positivity were excluded due to a high number of missing values ($>$40%). There was too little variance or observed cases for the dummy-encoded donor causes of death "bacterial meningitis", "primary brain tumour" and "subarachnoid haemorrhage", recipient end-stage renal disease "congenital dysplasia/hypoplasia/malformation with/without urinary tract malformation", "drug-induced/toxic", "tumour/trauma/surgery" and "subarachnoid haemorrhage".

As mentioned in Section 5.2, the Banff lesion scores i, t, v, ptc and cg showed (almost) zero variance. Therefore, the variables and additionally, due to the high correlation between this and Banff ct (ct0 and ct1 vs. ct2 and ct3) in the dichotomised versions, the variable Banff ci (ci0 and ci1 vs. ci2 and ci3), were excluded.

Both of donor serum createnine and donor eGFR, the first, last and highest measures were available. As these repeated measures were correlated according to their VIFs listed in Table 5.10, only the variable of the last measurement was retained, each. The same was true for the donors and recipients weight and height, which correlated with BMI and type of brain death (primary, secondary, unknown) when the donors cause of death was due to a cerebrovascular accident or traumatic brain injury.

Table 5.10: Variables with generalized variance inflation factors (VIFs) above 5 in multivariable generalized linear models with outcomes DGF and 1y-tl

	Outcome: DGF		Outcome: 1y-tl	
Predictors included in multivariable modelling (selection of VIFs > 5)	**Generalized VIFs with all predictors***	**Generalized VIFs after exclusion of corr. predictors**	**Generalized VIFs with all predictors***	**Generalized VIFs after exclusion of corr. predictors**
Type of brain death	32225810	excluded	30175310	excluded
Cause of death: cerebrovascular accident	5707932	3.29	13991820	2.61
Cause of death: traumatic brain injury	2438771	2.17	3945182	1.59
Recipient weight	30.54	excluded	198.99	excluded
Donor weight	124.13	excluded	137.64	excluded
Recipient BMI	22.26	1.23	112.12	1.25
Donor BMI	92.55	1.65	108.33	1.52
Recipient height	12.35	excluded	69.65	excluded
Donor height	41.84	excluded	48.04	excluded
Donor eGFR, max	9.76	excluded	30.07	excluded
Donor serum creatinine, highest	16.72	excluded	19.60	excluded
Donor eGFR, first	6.40	excluded	18.66	excluded
Donor serum creatinine, first	13.27	excluded	15.38	excluded
Donor eGFR, last	7.39	3.73	12.87	2.50
Donor serum creatinine, last	7.18	3.60	8.35	2.52
Recipient PRA, max	3.22	3.07	6.92	4.65

Abbreviations: eGFR, estimated glomerular filtration rate; PRA, panel reactive antibody; corr, correlating; VIF, variance inflation factor; max, maximum; DGF, delayed graft function; 1y-tl, death-censored transplant loss within one year.

* NOTE: variables excluded for other reasons than correlation, as described in Section 3.2, were not included in the multivariable models.

5.3.2 Logistic regression on multiple imputed dataset

After step 1, the imputation, 50 complete datasets with N=620 observations for the outcome DGF and N=711 observations for the outcome 1y-tl were available. In the subsequent modeling using logistic regression, the minimal model with cold ischaemia time as the only influencing variable for each of the outcomes was defined. The full models with all possible influencing variables could now also include those that were excluded before imputation due to their missing values (tables describing donor, recipient and transplant variables and Section 5.3.1).

Applying Rubin's rule to combine the regression models generated on each imputed dataset, the resulting model for DGF after forward selection included besides cold ischaemia time: donor and recipient BMI, donor and recipient CMV IgG-status, recipients PRA at time of transplantation, end-stage renal disease caused by hypertensive nephropathy or FSGS or "pyelonephritis/interstitial nephritis/obstructive uropathy/reflux uropathy" or "drug-induced/toxic" or "genetic nephropathy/glomerulopathy non-FSGS", dialysis vintage, previous transplantations, donors death caused by bacterial meningitis and the number of HLA-DR mismatches.

Respectively, variables identified to be associated with the outcome 1y-tl by for-

Table 5.11: Multivariable logistic regression model for outcome DGF generated on multiple imputed dataset [RefMod for DGF, Figure 3.5]

Predictor for DGF	beta	OR (95%-CI)	p-value
Intercept	-4.901	0.007 (0.001-0.04)	<.001
Donor BMI (kg/m²)	0.064	1.07 (1.02-1.11)	0.003
Recipient BMI (kg/m²)	0.046	1.05 (1.01-1.09)	0.017
Dialysis vintage (years)	0.092	1.1 (1.02-1.18)	0.012
HLA-DR mismatches: one	0.318	1.37 (0.87-2.16)	0.168
HLA-DR mismatches: two	0.867	2.38 (1.36-4.16)	0.002
Cold ischaemia time (hours)	0.031	1.03 (1.01-1.06)	0.014
ESRD: hypertensive nephropathy	-0.769	0.46 (0.2-1.06)	0.069

Abbreviations: CI, confidence interval; OR, odds ratio; ESRD, end-stage renal disease; HLA, humane leucozyte antibody; DGF, delayed graft function.

ward selection were: donor age, donor history of hypertension, recipients with previous transplantations or one of the end-stage renal diseases "drug-induced/toxic", "Congenital dysplasia/hypoplasia/malformation with/without urinary tract malformation", "TMA/systematic sclerosis" or "genetic nephropathy/glomerulopathy non-FSGS" and the sum of HLA-A, -B and -DR mismatches. The final models after backward selection are displayed in Table 5.11 and Table 5.12. In addition to cold ischaemia time, donor and recipient BMI, dialysis vintage and two HLA-DR mismatches were significantly associated with DGF, when beiing adjusted for hypertensive nephropathy as the recipients end-stage renal disease. Validation on the unimputed validation dataset revealed an AUC of .688 for prediction of DGF.

Table 5.12: Multivariable logistic regression model for outcome 1y-tl generated on multiple imputed dataset [RefMod for 1y-tl, Figure 3.5]

Predictor for 1y-tl	beta	OR (95%-CI)	p-value
Intercept	-4.662	0.009 (0.003-0.033)	<0.001
Donor age (years)	0.03	1.031 (1.011-1.051)	0.002
Cold ischaemia time (hours)	0.051	1.053 (1.017-1.089)	0.003

Abbreviations: CI, confidence interval; OR, odds ratio; 1y-tl, death-censored transplant loss within the first year.

Surprisingly, donor age is the only other variable besides cold ischaemia time that is sufficient to achieve an AUC of .724 when predicting 1y-tl on the validation dataset. Receiver operating characteristic curves are displayed in Figure 5.2. The calibration plot with the corresponding intercept and slope of the predicted DGF-values indicate a general underestimation of the risk (intercept = -.64, slope =.77) on the validation dataset 5.3. Consequently, the scaled Brier score was with 0.01 very poor. The estimated calibration index (ECI) for DGF was 1.79, a value to be compared to other ECIs of models predicting DGF on the validation dataset. Looking at the calibration plot for 1y-tl, Figure 5.4, one will notice the better calibration as the curve is much closer to the diagonal (intercept = -.37, slope = 1.54). However, the predicted probabilities only assume values between 0% and 35%, which again indicates an underestimation of the risk. Taking the scaled Brier score of .05 into account, the overall calibration was very poor while the ECI is rather good.

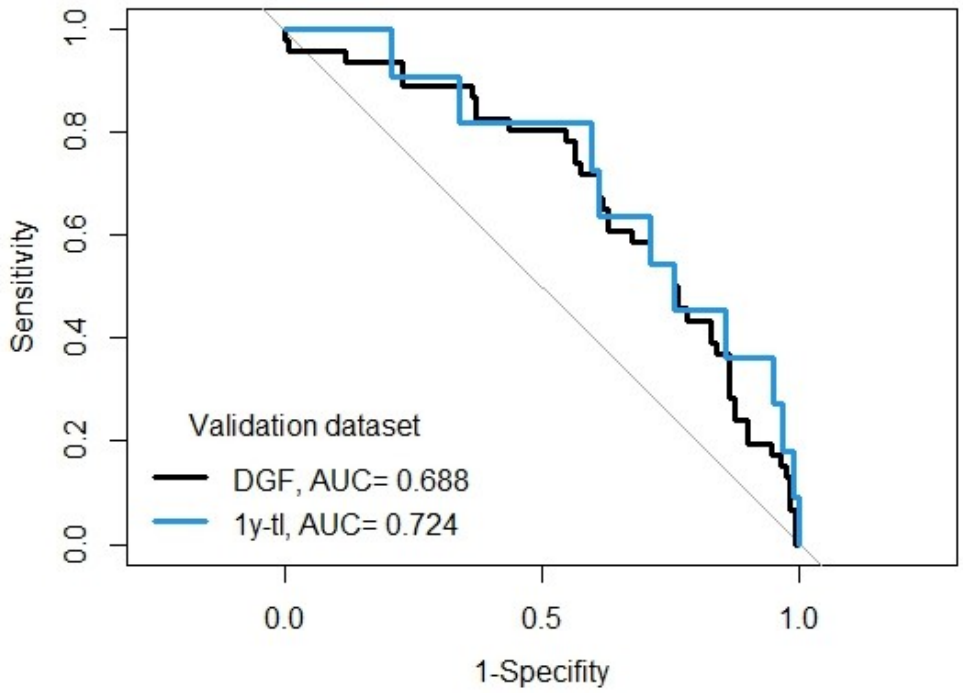


Figure 5.2: ROC of models trained on multiple imputed datasets for prediction of DGF and 1y-tl on the corresponding validation datasets

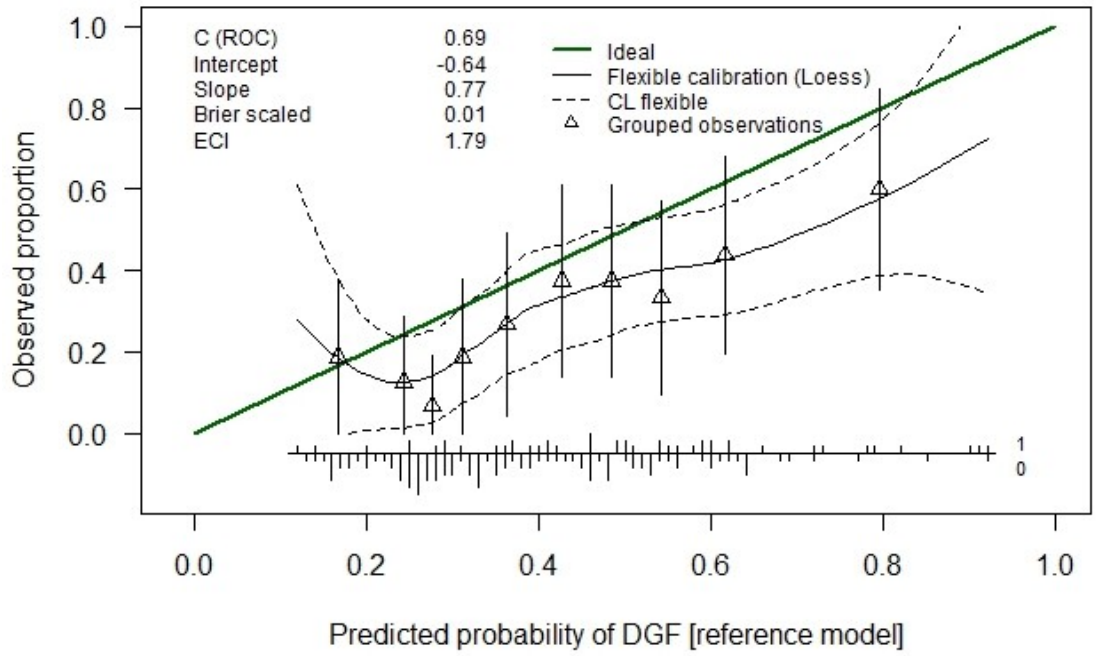


Figure 5.3: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model trained on the multiple imputed dataset for prediction of DGF on the validation dataset.

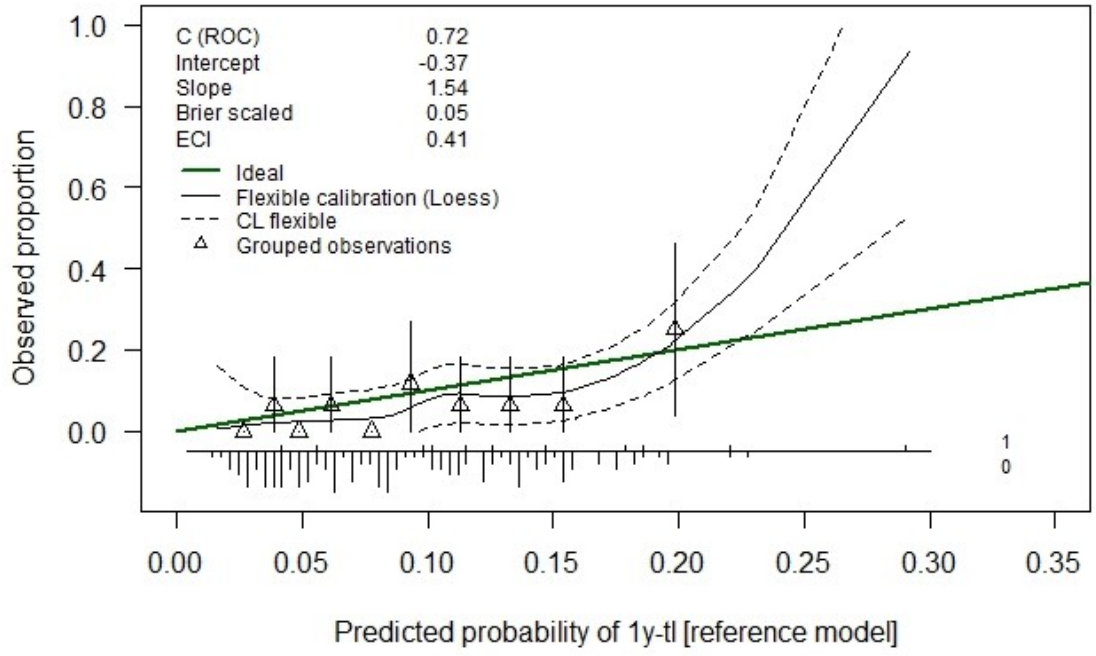


Figure 5.4: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model trained on the multiple imputed dataset for prediction of 1y-tl on the validation dataset

5.3.3 Logistic regression on unimputed, complete case dataset

Without prior variable selection and imputation of missing values, the number of potential influencing factors can be reduced to the most relevant ones using algorithms such as forward, backward and stepwise selection. Using a stepwise procedure with a smallest model including only the cold ischaemia time to predict DGF, the donor and recipient variables BMI and CMV-IgG positivity, as well as dialysis vintage, ESRD ("Pyelonephritis/Interstitial nephritis/Obstructive uropathy/Reflux uropathy" or "Hypertensive nephropathy" or "Genetic nephropathy/Glomerulopathy non FSGS") and the sum of HLA-DR mismatches were selected. However, on a significance level of 5%, donor CMV-IgG and ESRD ("Pyelonephritis/Interstitial nephritis/Obstructive uropathy/Reflux uropathy" and "Genetic nephropathy/Glomerulopathy non FSGS") were not statistically significant (Table 5.13). Also, in the presence of a quadratic term, an increased recipients BMI in kg/m^2 was associated with a decreased odds of DGF (OR [95%-CI]= .688 [.472-1.000]). Compared to other ESRDs, the presence of a hypertensive nephropathy was also associated with a decreased odds of DGF. The inclusion of statistically insignificant variables can be explained by the AICs close relationship to the likelihood ratio test. Under the χ^2 distribution with one degree of freedom, variable selection by the Akaike information criterion corresponds to a significance level of p = .157 [188]. Compared to the reference model for DGF (Table 5.11), four additional variables, related to the ERSD and the CMV IgG status, were selected despite a lower number of observations.

As a result, an overfitting of the regression model to the training data cannot be ruled out. The validation of this extensive model on the validation dataset has shown a significantly poorer ability to discriminate between cases with and without DGF (AUC = .509, Figure 5.5), which also suggests overfitting. The calibration plot with the corresponding intercept and slope of the predicted DGF- values indicate a stronger underestimation of the risk (intercept = -1.53, slope =.00) on the validation dataset as compared to the reference model (Figure 5.6). Consequently, the scaled Brier score was with -.78 very poor, just like the estimated calibration index which was with 7.88 much higher. However, the ECI value must be compared with the values of other models for predicting the DGF using the validation dataset for interpretation. A similar result was found when comparing the model for outcome 1y-tl (Table 5.14) with the corresponding reference model (Table 5.12). Besides cold ischaemia time and donor age, which were selected before, the recipients sex, the donors history of hypertension and age as quadratic term (age in $years^2$), and the HLA-B and -DR mismatches were identified as potential predictors of 1y-tl.

Table 5.13: Multivariable logistic regression model for outcome DGF generated on complete case dataset [M1 = Model 1 for DGF, Figure 3.5]

Predictor for DGF	beta	OR (95%-CI)	p-value
Intercept	-0.197	–	0.94
Cold ischaemia time (h)	0.042	1.043 (1.015, 1.072)	0.002
Donor BMI (kg/m²)	0.079	1.082 (1.034, 1.134)	<0.001
(Recipient BMI (kg/m²))^2	0.007	1.008 (1.001, 1.015)	0.03
Recipient BMI (kg/m²)	-0.374	0.688 (0.472, 1.000)	0.05
Dialysis vintage (y)	0.082	1.086 (1.001, 1.178)	0.05
Donor CMV IgG			
Negative (ref)	–	–	–
Positive	0.354	1.425 (0.931, 2.205)	0.11
Recipient CMV IgG			
Negative (ref)	–	–	–
Positive	0.737	2.090 (1.194, 3.768)	0.01
Unknown	0.346	1.413 (0.788, 2.590)	0.25
ESRD: Genetic nephropathy/ Glomerulopathy non FSGS	-0.489	0.613 (0.330, 1.094)	0.11
ESRD: Pyelonephritis/Interstitial nephritis/ Obstructive uropathy/ Reflux uropathy	-0.745	0.475 (0.152, 1.228)	0.15
ESRD: Hypertensive nephropathy	-1.127	0.324 (0.107, 0.801)	0.03
HLA-DR mismatches			
None (ref)	–	–	–
One	0.309	1.362 (0.839, 2.249)	0.22
Two	0.670	1.954 (1.055, 3.633)	0.03

Abbreviations: CI, confidence interval; OR, odds ratio; ESRD, end-stage renal disease; y, years; h, hours; CMV IgG, cytomegalovirus antibody IgG; FSGS, focal segmental glomerulosclerosis; HLA, humane leucozyte antibody; DGF, delayed graft function.

However, on a significance level of 5%, none of these additional variables were significantly associated with the odds of 1y-tl. Also, in the presence of the quadratic term, an increasing donor age in years was associated with a decreased odds of 1y-tl.

In contrast to the outcome DGF, the addition of further influencing variables led to an improvement in the AUC on the validation dataset (AUC = .762, Figure 5.5) compared to the reference model. Looking at the calibration plot for 1y-tl, Figure 5.7, one will notice a worse calibration compared to the reference model as the curve is less close to the diagonal (intercept = -.52, slope = .78). However, the predicted probabilities now also include values between 0% and > 60%, which indicates a lower underestimation of the potential risks. Taking the scaled Brier score of -.02 into account, the overall calibration was very poor while an ECI of .50 is again rather good.

Table 5.14: Multivariable logistic regression model for outcome 1y-tl generated on complete case dataset [M1 = Model 1 for 1y-tl, Figure 3.5]

Predictor for 1y-tl	beta	OR (95% CI)	p-value
Intercept	-2.022	–	0.17
Cold ischaemia time (hours)	0.052	1.053 (1.017, 1.091)	0.004
Donor history of hypertension			
No (ref)	–	–	–
Yes	0.609	1.838 (0.971, 3.584)	0.07
Unknown	-0.775	0.461 (0.149, 1.190)	0.14
(Donor age (years))^2	0.001	1.001 (1.000, 1.002)	0.05
Donor age (years)	-0.087	0.916 (0.828, 1.025)	0.11
HLA-B mismatches			
None (ref)	–	–	–
One	-0.161	0.851 (0.361, 2.141)	0.72
Two	0.517	1.677 (0.722, 4.186)	0.24
HLA-DR mismatches			
None (ref)	–	–	–
One	0.349	1.417 (0.707, 2.988)	0.34
Two	0.025	1.025 (0.402, 2.567)	0.96
Recipient sex			
Male (ref)	–	–	–
Female	0.373	1.452 (0.829, 2.526)	0.19

Abbreviations: CI, confidence interval; OR, odds ratio; ref, reference; HLA, humane leucozyte antibody; 1y-tl, death-censored transplant loss within one year.

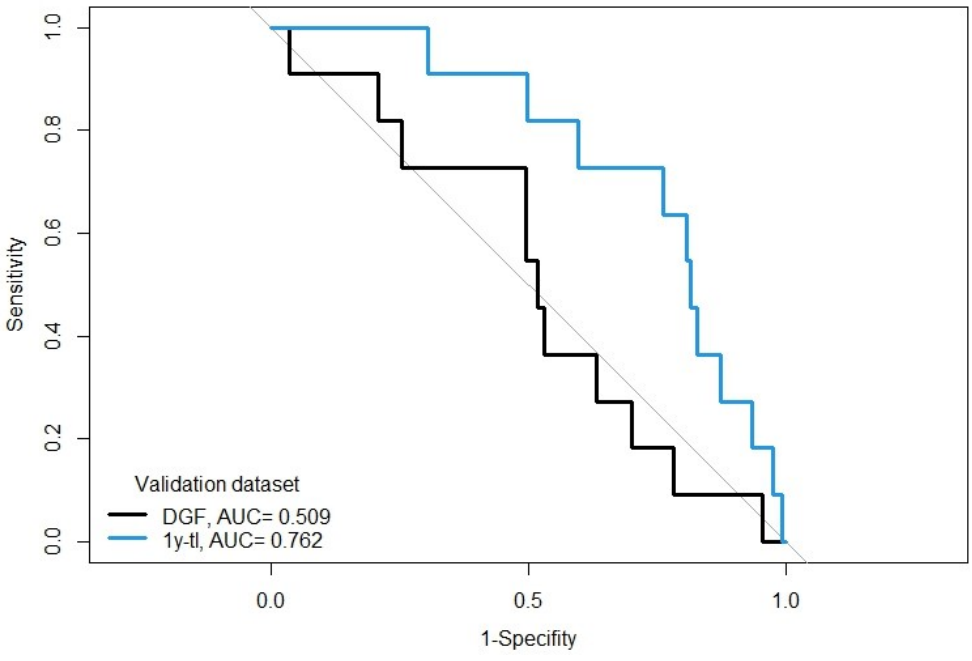


Figure 5.5: ROC of models trained on complete case datasets for prediction of DGF and 1y-tl on the corresponding validation datasets

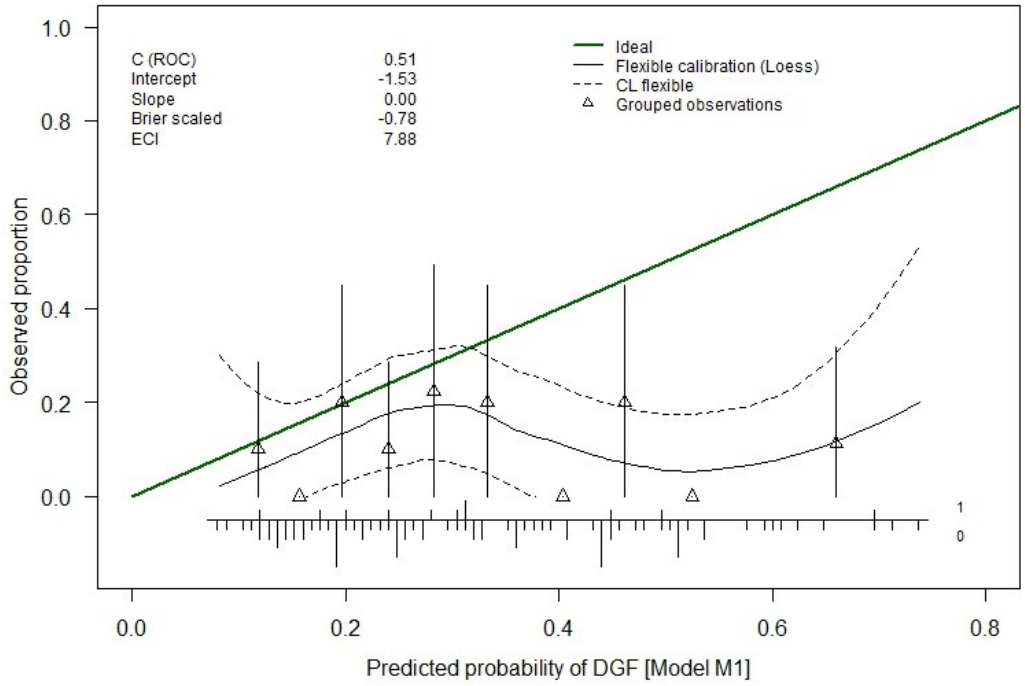


Figure 5.6: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the stepwise logistic regression model from the complete case dataset for prediction of DGF on the validation dataset

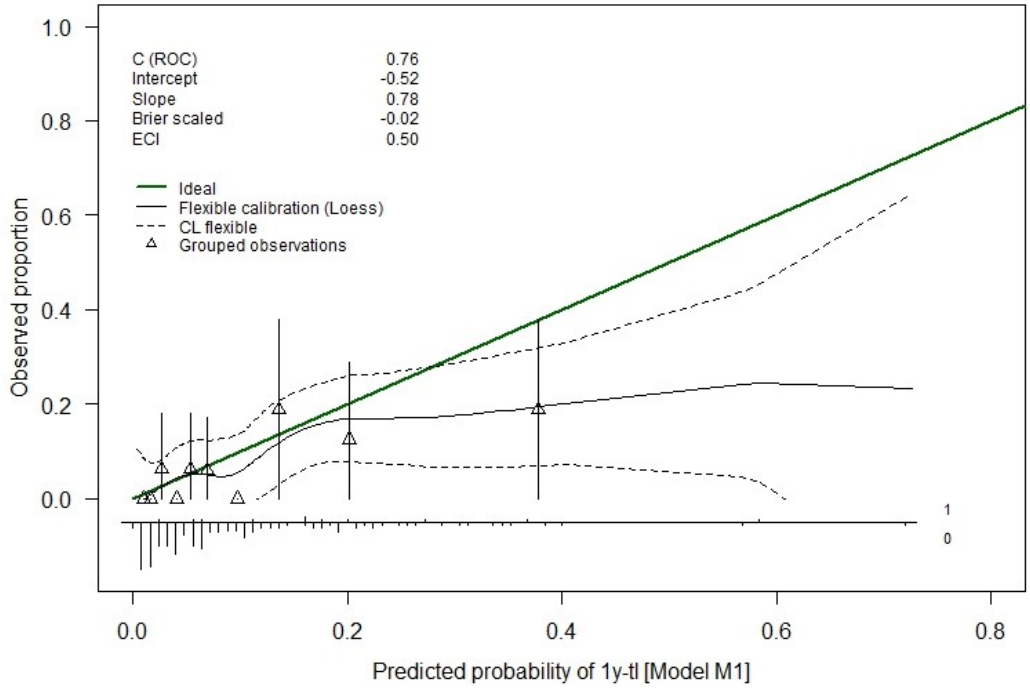


Figure 5.7: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the stepwise logistic regression model from the complete case dataset for prediction of 1y-tl on the validation dataset

5.3.4 Univariable p-value < 0.25

Results of the univariable analyses for the outcomes DGF and 1y-tl are listed for donor variables in Table 5.16, for recipient and transplant variables in Table 5.17 and for histological variables in Tables 5.32 and 5.33.

Regarding DGF, the donors age (y), BMI and history of hypertension, along with the recipients age (y), BMI, dialysis vintage (y), number of previous pregnancies, CMV IgG status, PRA at transplantation and the end-stage renal diseases "hypertensive nephropathy", "Pyelonephritis/Interstitial nephritis/obstructive uropathy/reflux uropathy", "FSGS" and "diabetic nephropathy" were associated. Also, the number of HLA-DR mismatches, sum of HLA A-, B- and DR- mismatches and the cold ischaemia time were associated with an increased odds of DGF.

Regarding 1y-tl, the donors age (y), history of diabetes mellitus or hypertension and last Cockcroft-Gault eGFR, along with the recipients age (y), sex and previous transplantations were associated. Also, besides the HLA-DR mismatches, all transplantation related variables including cold ischaemia time were associated with death-censored transplant loss within the first year after transplantation.

Table 5.15: Multivariable logistic regression model for the outcome DGF based on variables with univariable p-value $< .25$ [Model 2 for DGF, Figure 3.5]

Predictor for DGF	**beta**	**OR (95% CI)**	**p-value**
Intercept	-0.162	–	0.95
Cold ischaemia time (h)	0.040	1.041 (1.014, 1.070)	0.003
Donor BMI [kg/m²]	0.074	1.077 (1.029, 1.127)	0.001
ESRD: hypertensive nephropathy	-1.094	0.335 (0.111, 0.823)	0.03
HLA-DR mismatches			
None (ref)	–	–	–
One	0.360	1.433 (0.885, 2.364)	0.15
Two	0.836	2.307 (1.264, 4.239)	0.007
Recipient CMV-IgG			
Negative (ref)	–	–	–
Positive	0.767	2.154 (1.233, 3.876)	0.008
Unknown	0.433	1.542 (0.865, 2.816)	0.15
Dialysis vintage (years)	0.066	1.069 (0.985, 1.158)	0.11
Recipient BMI [kg/m²]	-0.369	0.692 (0.478, 1.001)	0.05
(Recipient BMI [kg/m²])²	0.008	1.008 (1.001, 1.014)	0.03
ESRD: FSGS	0.870	2.386 (0.930, 6.126)	0.07

Abbreviations: CI, confidence interval; OR, odds ratio; ref, reference; h, hours; HLA, humane leucozyte antibody; ESRD, end-stage renal disease; FSGS, focal segmental glomerulosclerosis; DGF, delayed graft function.

Based on the univariable selection two multivariable logistic regression models were derived applying stepwise selection. Variables age and BMI were included as linear

and quadratic terms, interaction between recipients age and the sum of HLA-A, -B and -DR mismatches was also considered but not selected. As displayed in Table 5.15, adjusted for dialysis vintage, which was not significantly associated with DGF, an increased cold ischaemia time, donor and recipient BMI, a positive recipients CMV-IgG status or FSGS as end-stage renal disease, as well as two HLA-DR mismatches increased odds for DGF. Recipients with hypertensive nephropathy, at the other hand, showed an decreased odds for DGF. Table 5.18 displays the multivariable model for the outcome 1y-tl. Here, the number of HLA-B mismatches, recipients previous transplantations and a donors history of hypertension were included but at a level of 5% not statistically significant. An increased cold ischaemia time and donor age, however, were associated with higher odds for 1y-tl. Both models were validated on their validation datasets.

Table 5.16: Donor characteristics by outcomes DGF and 1y-tl, univariable analyses

Outcome DGF (left) and 1y-tl (right)	Overall	No DGF	DGF	p	Overall	No TXP-loss	TXP-loss within 1y	p
N	***620***	***454***	***166***		***711***	***650***	***61***	
Donor age [y] (median [IQR])	54.00 [43.00, 64.00]	54.00 [42.00, 63.00]	55.00 [45.25, 66.50]	.045	54.00 [44.00, 64.00]	54.00 [43.00, 63.00]	62.00 [52.00, 71.00]	.001
Donor sex (N(%))								
Male	332 (53.5)	242 (53.3)	90 (54.2)	.912	382 (53.7)	354 (54.5)	28 (45.9)	.251
Female	288 (46.5)	212 (46.7)	76 (45.8)		329 (46.3)	296 (45.5)	33 (54.1)	
Donor BMI [kg/m²] (median [IQR])	25.71 [23.37, 27.78]	25.69 [23.15, 27.78]	26.15 [24.22, 29.24]	.005	25.71 [23.36, 27.99]	25.71 [23.31, 27.98]	26.01 [24.07, 28.70]	.313
Donor diabetes mellitus (N(%))								
No	404 (65.2)	297 (65.4)	107 (64.5)	.975	470 (66.1)	424 (65.2)	46 (75.4)	.194
Yes	37 (6.0)	27 (5.9)	10 (6.0)		42 (5.9)	38 (5.8)	4 (6.6)	
Unknown	179 (28.9)	130 (28.6)	49 (29.5)		199 (28.0)	188 (28.9)	11 (18.0)	
Donor hypertension (N(%))								
No	251 (40.5)	193 (42.5)	58 (34.9)	.119	298 (41.9)	279 (42.9)	19 (31.1)	<.001
Yes	233 (37.6)	160 (35.2)	73 (44.0)		262 (36.8)	225 (34.6)	37 (60.7)	
Unknown	136 (21.9)	101 (22.2)	35 (21.1)		151 (21.2)	146 (22.5)	5 (8.2)	
Donor creatinine, first [mg/dl] (median [IQR])	0.84 [0.69, 1.07]	0.84 [0.69, 1.09]	0.83 [0.69, 1.05]	.946	0.83 [0.69, 1.07]	0.83 [0.69, 1.06]	0.84 [0.67, 1.09]	.945
Donor creatinine, last [mg/dl] (median [IQR])	0.81 [0.66, 1.06]	0.81 [0.66, 1.05]	0.82 [0.65, 1.10]	.312	0.81 [0.65, 1.07]	0.81 [0.64, 1.06]	0.80 [0.67, 1.09]	.833
Donor creatinine, highest [mg/dl] (median [IQR])	0.96 [0.77, 1.24]	0.93 [0.77, 1.23]	0.99 [0.76, 1.31]	.607	0.94 [0.76, 1.24]	0.94 [0.77, 1.25]	0.89 [0.69, 1.16]	.154
Donor Cockcroft-Gault eGFR, first (median [IQR])	99.84 [78.11, 128.71]	100.43 [78.45, 128.25]	96.78 [77.64, 128.96]	.943	100.35 [78.07, 127.33]	101.00 [78.87, 128.61]	92.36 [75.56, 120.21]	.124
Donor Cockcroft-Gault eGFR, last (median [IQR])	101.85 [75.59, 136.29]	103.13 [76.73, 136.44]	98.33 [70.52, 134.80]	.264	102.69 [76.39, 135.70]	103.27 [76.73, 136.18]	89.58 [69.33, 126.81]	.157
Donor Cockcroft-Gault eGFR, highest (median [IQR])	87.82 [67.34, 112.54]	89.00 [68.20, 112.52]	83.90 [61.75, 112.37]	.348	88.02 [67.84, 112.85]	88.06 [68.00, 112.96]	85.11 [67.16, 111.64]	.745
Donor CMV IgG (N(%))								
Negative	227 (36.6)	172 (37.9)	55 (33.1)	.521	256 (36.0)	235 (36.2)	21 (34.4)	.935
Positive	384 (61.9)	276 (60.8)	108 (65.1)		446 (62.7)	407 (62.6)	39 (63.9)	
Missing	9 (1.5)	6 (1.3)	3 (1.8)		9 (1.3)	8 (1.2)	1 (1.6)	
Donor CMV IgM (N(%))								
Negative	450 (72.6)	323 (71.1)	127 (76.5)	.392	522 (73.4)	472 (72.6)	50 (82.0)	.242
Positive	7 (1.1)	5 (1.1)	2 (1.2)		9 (1.3)	8 (1.2)	1 (1.6)	
Missing	163 (26.3)	126 (27.8)	37 (22.3)		180 (25.3)	170 (26.2)	10 (16.4)	
Donor HCVAB (N(%))								
Negative	616 (99.4)	452 (99.6)	164 (98.8)	.575	707 (99.4)	646 (99.4)	61 (100.0)	.828
Positive	2 (0.3)	1 (0.2)	1 (0.6)		2 (0.3)	2 (0.3)	0 (0.0)	
Missing	2 (0.3)	1 (0.2)	1 (0.6)		2 (0.3)	2 (0.3)	0 (0.0)	
Type of brain death (N(%))								
Primary	411 (66.3)	301 (66.3)	110 (66.3)	.8	479 (67.4)	436 (67.1)	43 (70.5)	.799
Secondary	71 (11.5)	54 (11.9)	17 (10.2)		80 (11.3)	73 (11.2)	7 (11.5)	
Unknown	138 (22.3)	99 (21.8)	39 (23.5)		152 (21.4)	141 (21.7)	11 (18.0)	
Cause of death (N(%))								
apnea	16 (2.6)	14 (3.1)	2 (1.2)	.308	17 (2.4)	16 (2.5)	1 (1.6)	> .900
bacterial meningitis	3 (0.5)	3 (0.7)	0 (0.0)	.692	4 (0.6)	4 (0.6)	0 (0.0)	> .900
cardiac arrest	27 (4.4)	18 (4.0)	9 (5.4)	.572	31 (4.4)	28 (4.3)	3 (4.9)	> .900
cerebrovascular accident	329 (53.1)	239 (52.6)	90 (54.2)	.835	379 (53.3)	343 (52.8)	36 (59.0)	.437
drug suicide	2 (0.3)	2 (0.4)	0 (0.0)	.955	2 (0.3)	2 (0.3)	0 (0.0)	> .900
primary brain tumor	1 (0.2)	1 (0.2)	0 (0.0)	> .900	1 (0.1)	1 (0.2)	0 (0.0)	> .900
trauma	21 (3.4)	17 (3.7)	4 (2.4)	.574	23 (3.2)	21 (3.2)	2 (3.3)	> .900
traumatic brain injury	76 (12.3)	56 (12.3)	20 (12.0)	> .900	93 (13.1)	86 (13.2)	7 (11.5)	.849
Unknown	145 (23.4)	104 (22.9)	41 (24.7)	.676	161 (22.6)	149 (22.9)	12 (19.7)	.694

Abbreviations: eGFR, estimated glomerular filtration rate; IQR, interquartile range; CMV IgG and IgM, cytomegalovirus antibody IgG and IgM; TXP, transplant; HCVAB, hepatitis C virus antibody-positive; DGF, delayed graft function; 1y-tl, death-censored transplant loss within one year.

Receiver operating curves of the predicted outcomes are displayed in Figure 5.8. Like on the multiple imputed data, the AUC for 1y-tl was with a value of .822 higher than for DGF, which reached an AUC of .661.

Table 5.17: Recipient and transplant characteristics by outcomes DGF and 1y-tl, univariable analyses

Outcome DGF (left) and 1y-tl (right)	*Overall*	No DGF	DGF	p	Overall	No TXP-loss	TXP-loss within 1y	p
N	***620***	***454***	***166***		***711***	***650***	***61***	
Recipient age [y] (median [IQR])	55.0 [45.0, 64.0]	54.0 [45.0, 63.0]	55.50 [45.25, 65.0]	.19	55.0 [45.0, 64.0]	54.0 [45.0, 63.75]	60.00 [48.0, 67.0]	.014
Recipient sex (N(%))								
Male	393 (63.4)	284 (62.6)	109 (65.7)	.537	454 (63.9)	420 (64.6)	34 (55.7)	.215
Female	227 (36.6)	170 (37.4)	57 (34.3)		257 (36.1)	230 (35.4)	27 (44.3)	
Recipient BMI [kg/m²] (median [IQR])	25.41 [22.88, 29.02]	25.51 [22.83, 28.71]	25.25 [23.23, 30.46]	.208	25.52 [22.96, 29.04]	25.47 [22.87, 28.74]	26.00 [23.14, 30.34]	.358
Dialysis vintage [y] (median [IQR])	3.11 [1.79, 4.90]	3.01 [1.65, 4.49]	3.43 [2.13, 5.38]	.006	3.13 [1.86, 4.85]	3.12 [1.86, 4.83]	3.58 [1.86, 4.92]	.731
Recipient no. of previous pregnancies (median [IQR])	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	.019	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	.783
Previous transplantations (%)								
No	564 (91.0)	415 (91.4)	149 (89.8)	.634	643 (90.4)	591 (90.9)	52 (85.2)	.225
Yes	56 (9.0)	39 (8.6)	17 (10.2)		68 (9.6)	59 (9.1)	9 (14.8)	
Recipient CMV IgG (N(%))								
Negative	124 (20.0)	99 (21.8)	25 (15.1)	.111	142 (2.0)	133 (20.5)	9 (14.8)	.524
Positive	254 (41.0)	177 (39.0)	77 (46.4)		277 (39.0)	253 (38.9)	24 (39.3)	
Missing	242 (39.0)	178 (39.2)	64 (38.6)		292 (41.1)	264 (40.6)	28 (45.9)	
Recipient CMV IgM (N(%))								
Negative	281 (45.3)	201 (44.3)	80 (48.2)	.532	320 (45.0)	294 (45.2)	26 (42.6)	.275
Positive	15 (2.4)	10 (2.2)	5 (3.0)		15 (2.1)	12 (1.8)	3 (4.9)	
Missing	324 (52.3)	243 (53.5)	81 (48.8)		376 (52.9)	344 (52.9)	32 (52.5)	
Recipient PRA at transplantation (median [IQR])	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.134	0.0 [.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	.758
Recipient PRA peak historic (median [IQR])	0.0 [0.0, 5.0]	0.0 [0.0, 5.0]	0.0 [0.0, 5.75]	.599	0.0 [0.0, 5.0]	0.0 [0.0, 5.0]	0.0 [0.0, 8.0]	0.654
ESRD (N(%))								
Congenital dysplasia/hypoplasia/ malformation with/without urinary tract malformation	7 (1.1)	5 (1.1)	2 (1.2)	.675	7 (1.0)	7 (1.1)	0 (.0)	.621
Diabetic nephropathy	81 (13.1)	54 (11.9)	27 (16.3)	.195	98 (13.8)	89 (13.7)	9 (14.8)	.971
Drug-induced/toxic	7 (1.1)	7 (1.5)	0 (0.0)	.238	7 (1.0)	7 (1.1)	0 (0.0)	.892
FSGS	21 (3.4)	11 (2.4)	10 (6.0)	.052	25 (3.5)	22 (3.4)	3 (4.9)	.796
Genetic nephropathy/ Glomerulopathy non-FSGS	96 (15.5)	75 (16.5)	21 (12.7)	.292	111 (15.6)	105 (16.2)	6 (9.8)	.265
GN/Acquired Glomerulopathy	146 (23.5)	103 (22.7)	43 (25.9)	.466	175 (24.6)	157 (24.2)	18 (29.5)	.44
Hypertensive nephropathy	46 (7.4)	41 (9.0)	5 (3.0)	.018	50 (7.0)	47 (7.2)	3 (4.9)	.679
Nephrocalcinosis	1 (0.2)	1 (0.2)	0 (0.0)	> .900	1 (0.1)	1 (0.2)	0 (0.0)	>.900
Pyelonephritis/Interstitial nephritis/ obstructive uropathy/reflux uropathy	35 (5.6)	30 (6.6)	5 (3.0)	.128	41 (5.8)	37 (5.7)	4 (6.6)	> .900
Renal vascular disease	21 (3.4)	14 (3.1)	7 (4.2)	.66	24 (3.4)	21 (3.2)	3 (4.9)	.744
TMA/Scleroderma	9 (1.5)	6 (1.3)	3 (1.8)	.945	9 (1.3)	9 (1.4)	0 (0.0)	.744
Tumour/trauma/surgery	5 (0.8)	4 (0.9)	1 (0.6)	> .900	5 (0.7)	5 (0.8)	0 (0.0)	> .900
Unknown cause	145 (23.4)	103 (22.7)	42 (25.3)	.566	158 (22.2)	143 (22.0)	15 (24.6)	.761
HLA mismatches A (N(%))								
None	194 (31.3)	144 (31.7)	50 (30.1)	.911	229 (32.2)	215 (33.1)	14 (23.0)	.208
One	332 (53.5)	241 (53.1)	91 (54.8)		373 (52.5)	340 (52.3)	33 (54.1)	
Two	93 (15.0)	68 (15.0)	25 (15.1)		108 (15.2)	94 (14.5)	14 (23.0)	
Missing	1 (0.2)	1 (0.2)	0 (0.0)		1 (0.1)	1 (0.2)	0 (0.0)	
HLA mismatches B (N(%))								
None	119 (19.2)	89 (19.6)	30 (18.1)	.895	137 (19.3)	128 (19.7)	9 (14.8)	.04
One	305 (49.2)	221 (48.7)	84 (50.6)		351 (49.4)	328 (50.5)	23 (37.7)	
Two	195 (31.5)	143 (31.5)	52 (31.3)		222 (31.2)	193 (29.7)	29 (47.5)	
Missing	1 (0.2)	1 (0.2)	0 (0.0)		1 (0.1)	1 (0.2)	0 (0.0)	
HLA mismatches DR (N(%))								
None	177 (28.5)	139 (30.6)	38 (22.9)	.029	207 (29.1)	193 (29.7)	14 (23.0)	.702
One	338 (54.5)	249 (54.8)	89 (53.6)		393 (55.3)	357 (54.9)	36 (59.0)	
Two	104 (16.8)	65 (14.3)	39 (23.5)		110 (15.5)	99 (15.2)	11 (18.0)	
Missing	1 (0.2)	1 (0.2)	0 (0.0)		1 (0.1)	1 (0.2)	0 (0.0)	
HLA mismatches [sum] (N(%))								
0	61 (9.8)	47 (10.4)	14 (8.4)	.143	73 (10.3)	68 (10.5)	5 (8.2)	.152
1	52 (8.4)	42 (9.3)	10 (6.0)		60 (8.4)	58 (8.9)	2 (3.3)	
2	105 (16.9)	75 (16.5)	30 (18.1)		125 (17.6)	117 (18.0)	8 (13.1)	
3	194 (31.3)	141 (31.1)	53 (31.9)		216 (30.4)	201 (30.9)	15 (24.6)	
4	137 (22.1)	105 (23.1)	32 (19.3)		159 (22.4)	139 (21.4)	20 (32.8)	
5	53 (8.5)	30 (6.6)	23 (13.9)		59 (8.3)	50 (7.7)	9 (14.8)	
6	17 (2.7)	13 (2.9)	4 (2.4)		18 (2.5)	16 (2.5)	2 (3.3)	
Missing	1 (0.2)	1 (0.2)	0 (0.0)		1 (0.1)	1 (0.2)	0 (0.0)	
Cold ischaemia time [h] (median [IQR])	12.28 [6.00, 16.99]	12.00 [5.74, 16.00]	13.69 [7.97, 19.00]	.015	11.27 [0.00, 16.03]	11.00 [0.00, 16.00]	15.00 [9.00, 19.00]	.001
Warm ischaemia time [min] (median [IQR])/(mean (SD))	35.00 [0.00, 46.25]	34.00 [0.00, 45.00]	39.00 [0.00, 52.75]	.021	28.46 (25.19)	27.57 (24.75)	37.92 (28.02)	.002

Abbreviations: IQR, interquartile range; TXP, transplant; HLA, humane leucozyte antibody; CMV IgG and IgM, cytomegalovirus antibody IgG and IgM; h, hours; min, minutes; SD, standard deviation; FSGS, focal segmental glomerulosclerosis; TMA, thrombotic microangiopathy; ESRD, end-stage renal disease; PRA, panel reactive antibody; DGF, delayed graft function; 1y-tl, death-censored transplant loss within one year.

Table 5.18: Multivariable logistic regression model for the outcome 1y-tl based on variables with univariable p-value $< .25$ [Model 2 for 1y-tl, Figure 3.5]

Predictor for 1y-tl	beta	OR (95% CI)	p-value
Intercept	-1.803	–	0.21
Cold ischaemia time (h)	0.047	1.048 (1.013, 1.087)	0.008
Donor history of hypertension			
No (ref)	–	–	–
Yes	0.622	1.862 (0.986, 3.623)	0.06
Unknown	-0.738	0.478 (0.154, 1.233)	0.15
HLA-B mismatches			
None (ref)	–	–	–
One	0.035	1.036 (0.470, 2.469)	0.93
Two	0.637	1.891 (0.867, 4.483)	0.12
Donor age (y)	-0.091	0.913 (0.825, 1.020)	0.09
(Donor age)² (y)²	0.001	1.001 (1.000, 1.002)	0.04
Previous transplantations = yes	0.628	1.873 (0.796, 4.027)	0.13

Abbreviations: OR, odds ratio; CI, confidence interval; HLA, humane leucozyte antibody; ref, reference; y, years.

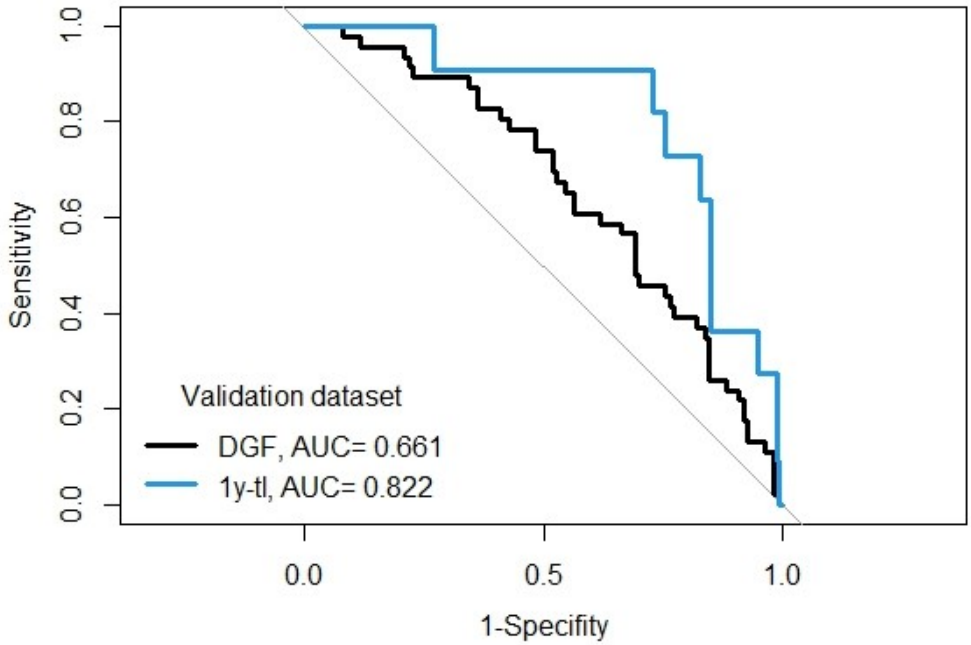


Figure 5.8: ROC of models based on univariable p $< .25$ for prediction of DGF and 1y-tl on the corresponding validation datasets

The calibration plot with the corresponding intercept and slope of the predicted DGF-values indicate a general underestimation of the risk (intercept = -.53, slope =.61) on the validation dataset 5.9. Consequently, the scaled Brier score was with -0.01 very poor. The estimated calibration index (ECI) for DGF was 1.65, a value to be compared to other ECIs of models predicting DGF on the validation dataset. Looking at the calibration plot for 1y-tl, Figure 5.10, one will notice a similar calibration compared to previous models predicting this outcome (intercept = -.55, slope = 1.39). However, the predicted probabilities only assume values between 0% and 45%, which again indicates an underestimation of the risk. Taking the scaled Brier score

of .09 into account, the overall calibration was again very poor while an ECI of .23 is rather good.

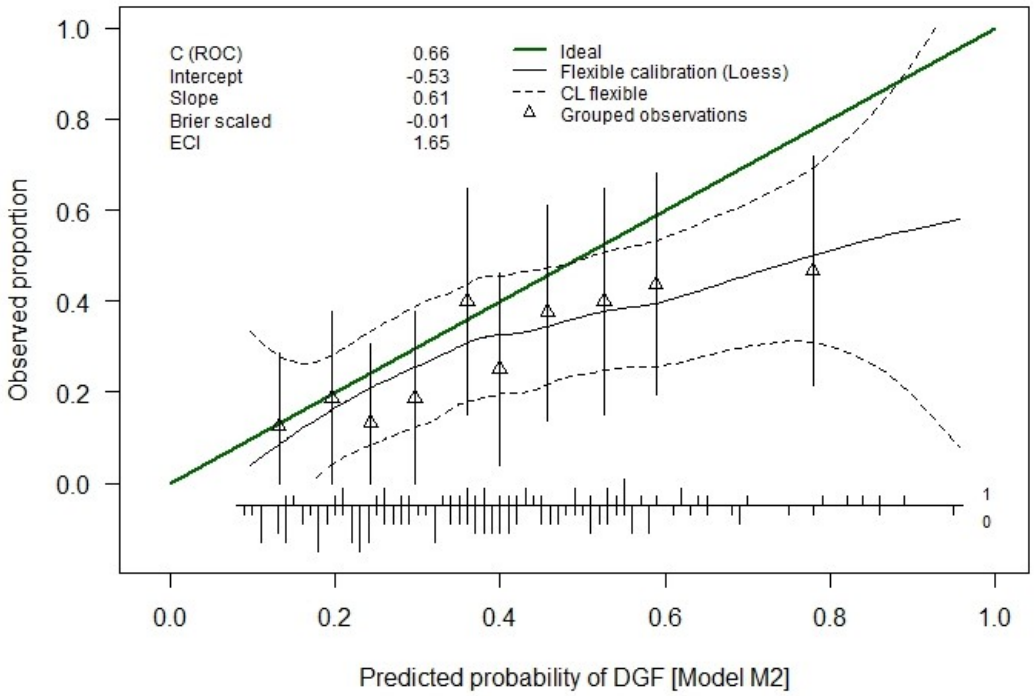


Figure 5.9: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on univariable p-values $< .25$ for prediction of DGF on the validation dataset

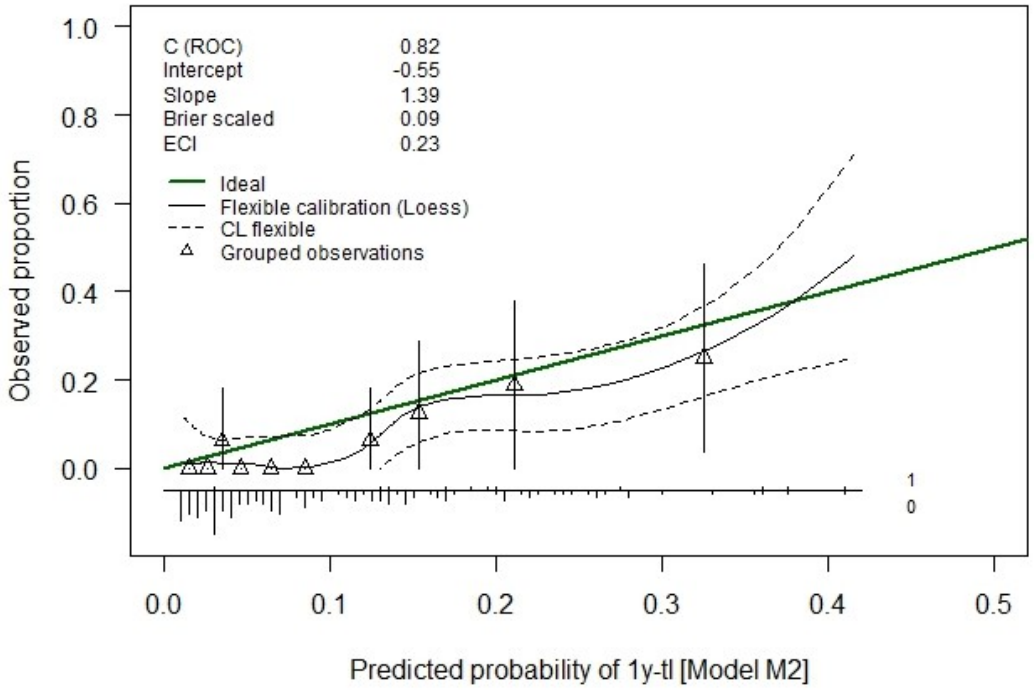


Figure 5.10: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on univariable p-values $< .25$ for prediction of 1y-tl on the validation dataset

5.3.5 LASSO

Since LASSO does not allow factor variables with more than two levels, those with more than two levels had to be dichotomised using dummy coding. Missing values were imputed. Subsequently, leave-one-centre-out cross-validation was used to determine the penalisation parameter lambda using the AUC as measure of goodnes of fit. Results for different values of lambda are displayed in Figure 5.11 and Figure 5.12. Selection of the smallest lambda with yet high AUC resulted in $\lambda_{DGF} =$ 0.02363172 for DGF and $\lambda_{1y-tl} = 0.03442374$ for 1y-tl. The corresponding numbers of the selected independent variables (N_{LASSO}(DGF) = 13 and N_{LASSO}(1y-tl) = 2) can also be seen in the figures and are listed with the penalized regression coefficients in Table 5.19. One variable that was previously excluded due to missing values, but was now considered relevant for both DGF and 1y-tl after imputation, was a previous diagnosis of hypertension in the donor. The same applied to the CMV-IgG positivity of the recipient, which was also included in the LASSO model with DGF. The models derived by LASSO were validated on the independent validation set. Prediction of DGF had an accuracy of 0.734 (95% CI: .658 - .801) with sensitivity and specificity of .696 and .750, given a threshold of 0.373, respectively. Prediction of 1y-tl had an accuracy of 0.879 (95% CI: .801 - .934) with sensitivity and specificity of .444 and .918, given a threshold of 0.098, respectively.

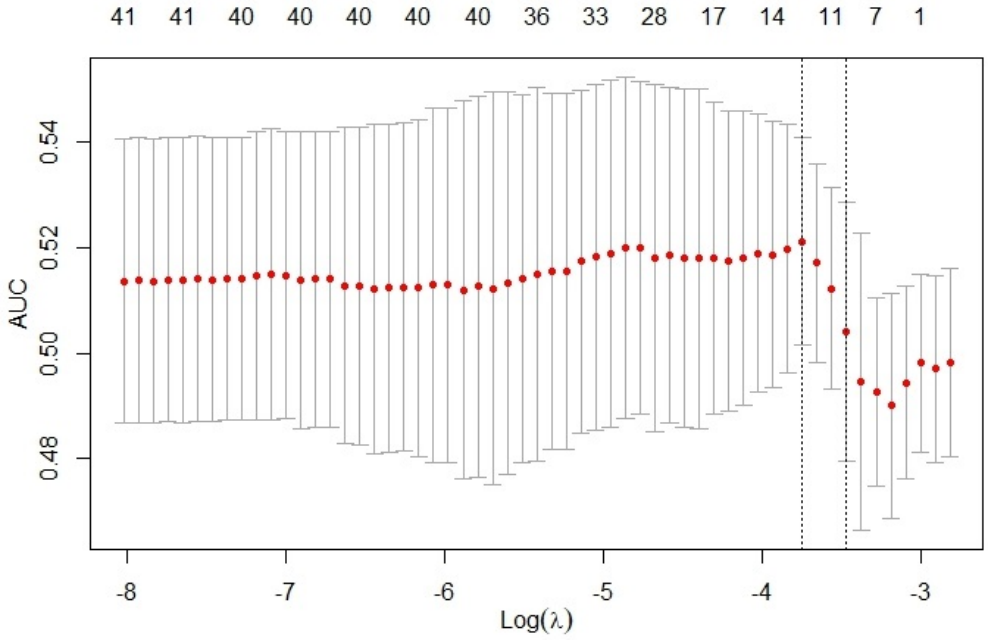


Figure 5.11: LASSO: cross-validation to estimate lambda λ_{DGF} for the outcome DGF

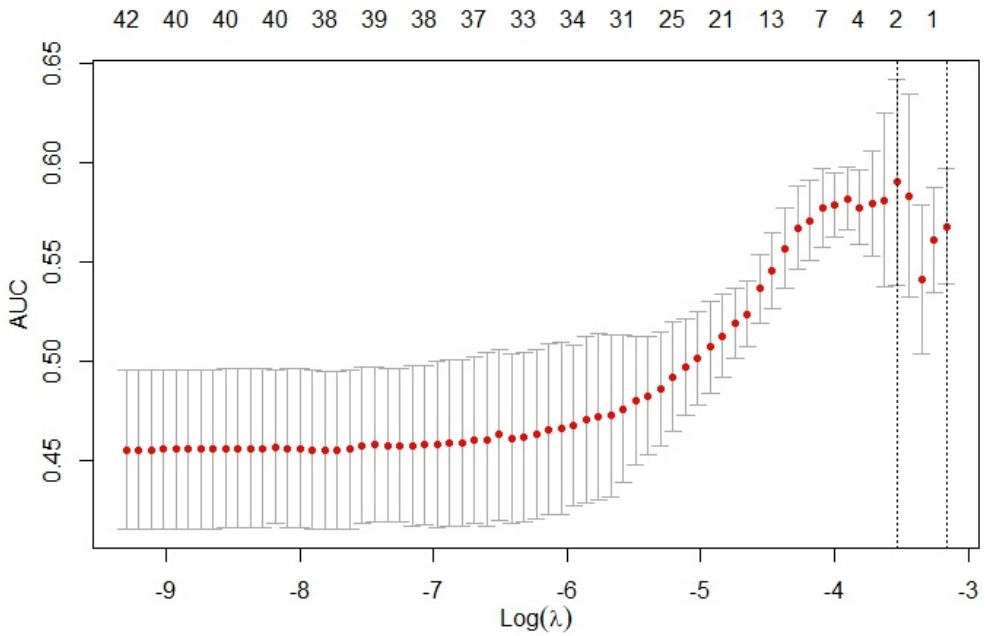


Figure 5.12: LASSO: cross-validation to estimate lambda λ_{1y-tl} for the outcome 1y-tl

Table 5.19: Multivariable LASSO penalized models for the outcome DGF and 1y-tl [M3 for DGF and 1y-tl, Figure 3.5]

Variables and penalized coefficients for outcome DGF, $\lambda_{DGF} = 0.02363$		Variables and penalized coefficients for outcome 1y-tl, $\lambda_{1y-tl} = 0.02918$	
(Intercept)	-3.1277	(Intercept)	-2.5352
Donor age (years)	0.0026	Donor history of hypertension = No	-0.1064
Donor BMI (kg/m^2)	0.0355	Cold ischaemia time (h)	0.0056
Donor history of hypertension	0.0169	HLA-B mismatches = 2	0.0078
Donor serum creatinine, last mg/dl	0.0877	Donor age (years)	0.0030
Recipient BMI (kg/m^2)	0.0174		
Dialysis vintage (years)	0.0422		
Recipient CMV-IgG	0.2550		
ESRD: Hypertensive nephropathy	-0.5307		
ESRD: FSGS	0.3015		
ESRD: Pyelonephritis/Interstitial nephritis/ obstructive uropathy/reflux uropathy	-0.1113		
ESRD = Genetic nephropathy/ Glomerulopathy non FSGS	-0.0042		
Nr. of HLA-DR mismatches = 0	-0.0632		
Nr. of HLA-DR mismatches = 2	0.3086		
Cold ischaemia time (hours)	0.0139		

Abbreviations: HLA, humane leucozyte antibody; CMV IgG, cytomegalovirus antibody IgG; FSGS, focal segmental glomerulosclerosis; ESRD, end-stage renal disease; λ, penalty parameter; h, hours.
DGF model: accuracy=.743 (95% CI: .658 - .801); sensitivity=.696; specificity=.750
1y-tl model: accuracy=.879 (95% CI: .801 - .934); sensitivity=.444; specificity=.918

Table 5.20: Multivariable logistic regression model based on variables derived by LASSO penalization for the outcome DGF [Model 3 for DGF, Figure 3.5]

Predictor for DGF	beta (SE)	OR (95% CI)	p-value
Intercept	-0.044 (3.231)	–	0.99
(Recipient BMI)² (kg/m²)²	0.009 (0.005)	1.009 (1.000, 1.018)	0.05
(Donor age)² (years²)	0.00014 (0.00009)	1.00014 (0.99997, 1.00032)	0.11
Donor BMI (kg/m²)	0.055 (0.031)	1.057 (0.995, 1.124)	0.07
Recipient BMI (kg/m²)	-0.407 (0.244)	0.666 (0.412, 1.081)	0.1
Donor serum creatinine, last mg/dl	0.734 (0.286)	2.084 (1.207, 3.711)	0.01
Dialysis vintage (years)	0.106 (0.048)	1.112 (1.011, 1.223)	0.03
ESRD: Hypertensive nephropathy	-2.273 (1.044)	0.103 (0.006, 0.524)	0.03
Recipient CMV-IgG = positive	0.659 (0.302)	1.932 (1.084, 3.553)	0.03
ESRD = Genetic nephropathy/ Glomerulopathy non FSGS	-1.031 (0.431)	0.357 (0.143, 0.791)	0.02
Cold ischaemia time (hours)	0.023 (0.017)	1.023 (0.988, 1.059)	0.2

Abbreviations: OR, odds ratio; CI, confidence interval; ESRD, end-stage renal disease; FSGS, focal segmental glomerulosclerosis; SE, standard error.

Table 5.21: Multivariable logistic regression model based on variables derived by LASSO penalization for the outcome 1y-tl [Model 3 for 1y-tl, Figure 3.5]

Predictor for 1y-tl	beta (SE)	OR (95% CI)	p-value
Intercept	-5.037 (0.658)	–	<0.001
Cold ischaemia time (hours)	0.053 (0.018)	1.054 (1.019, 1.092)	0.003
Donor history of hypertension = yes	0.760 (0.302)	2.137 (1.195, 3.930)	0.01
HLA-B mismatches			
None (ref)			
One	0.333 (0.456)	1.395 (0.597, 3.665)	0.465
Two	0.932 (0.451)	2.540 (1.102, 6.619)	0.038

Abbreviations: OR, odds ratio; CI, confidence interval; HLA, humane leucozyte antibody; SE, standard error; ref, reference.

The calibration plot with the corresponding intercept and slope of the predicted DGF-values indicate a general underestimation of the risk (intercept = -1.3, slope =.58) on the validation dataset 5.13. Consequently, the scaled Brier score was with -0.08 very poor, like the ECI which equalled 5.42. Looking at the calibration plot for 1y-tl, Figure 5.14, one will notice a similar calibration compared to previous models predicting this outcome (intercept = -.37, slope = 1.79). However, the predicted probabilities only assume values between 0% and 30%, which again indicates an underestimation of the risk. Taking the scaled Brier score of .05 into account, the overall calibration was again very poor while an ECI of .17 is rather good.

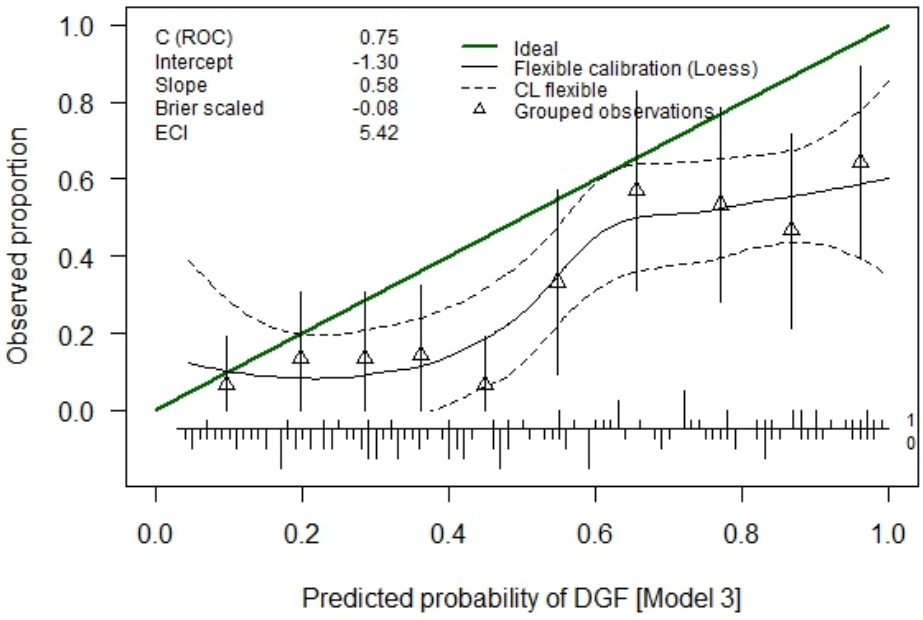


Figure 5.13: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on LASSO for prediction of DGF on the validation dataset

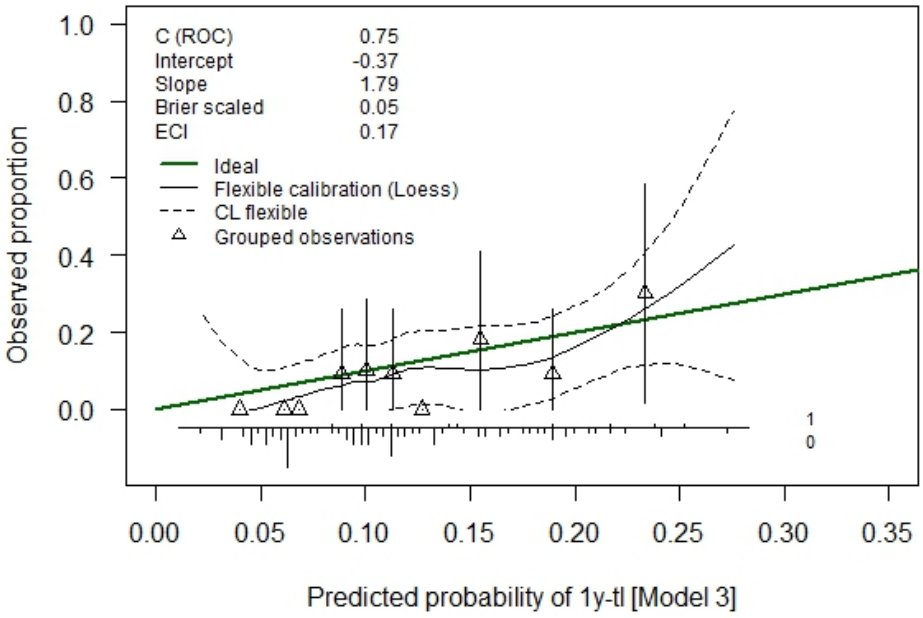


Figure 5.14: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on LASSO for prediction of 1y-tl on the validation dataset

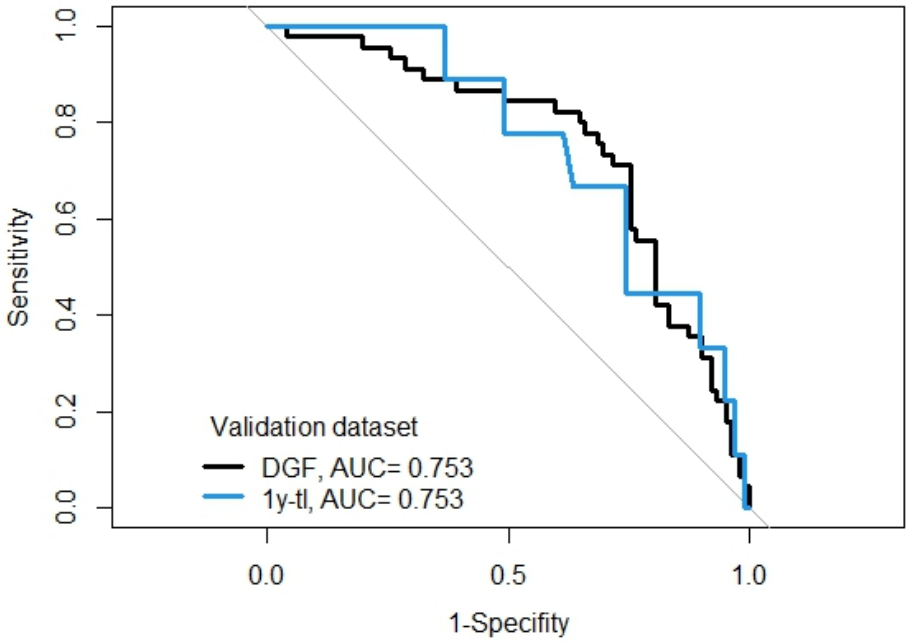


Figure 5.15: ROC of models based on LASSO variable selection for prediction of DGF and 1y-tl on the corresponding validation datasets

5.3.6 CART

As shown in figure 5.16, after pruning (complexity parameter = 0.0164), the recipients and donors BMI were amongst the most important predictors for DGF, along with cold ischaemia time. The odds for DGF in the group of recipients with a BMI $\geq$ 31.9 kg\m^2 and cold ischaemia time $\geq$ 14.28 hours were amongst the ones with highest odds (leave H, Odds=17:9) for DGF. In the subset of recipients with BMI $<$ 31.9 kg\m^2, an increased donor BMI $\geq >$ 30.8 kg\m^2 was also associated with DGF odds. When the donors in this subgroup also showed last serum creatinine values before transplantation $\geq$ 1.52 mg/dl, odds for DGF was 10:1 (leave F). Odds for DGF where reduced to 2:44 in the subset of recipients and donors with low BMI when additionally the recipients end-stage renal disease was one of the following: drug-induced or toxic, nephrocalcinosis, hypertensive nephropathy or tumor/trauma/surgery (leave A, odds = 2:44). Even when the donors BMI and recipients age were increased ($>$ 30.8 kg\m^2 and $>$ 40.5 years), odds for DGF was reduced to 4:33 when the donors last serum creatinine value and age were in the lower subset ($<$ 1.52 mg/dl and $<$ 69.5 years).

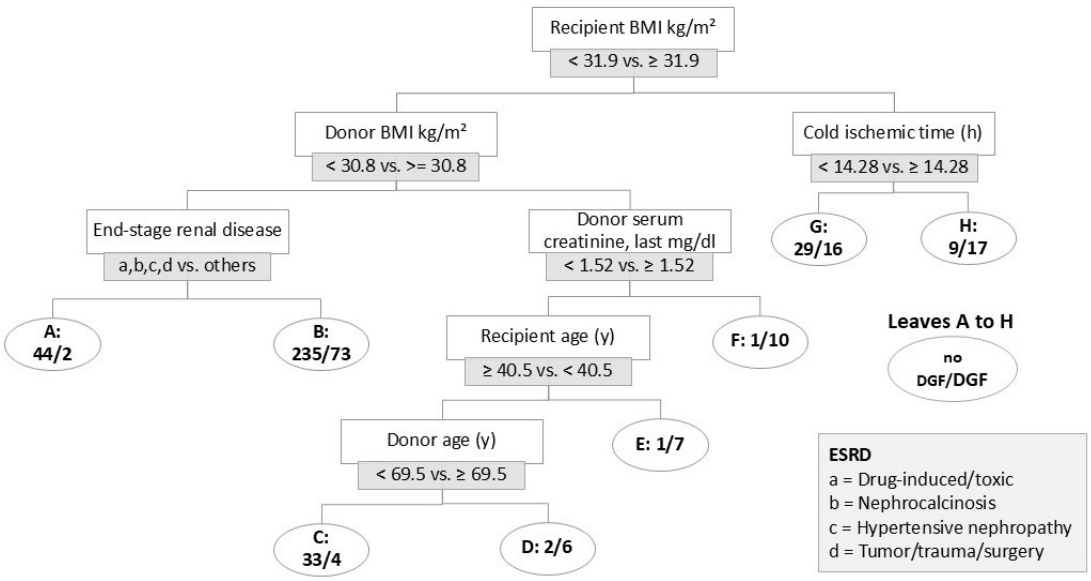


Figure 5.16: CART for outcome DGF [Model M4, Figure 3.5]]

According to the results of the CART algorithm, the three recipient-related variables BMI, age and ESRD, the donor-related variables BMI, age and serum creatinine, and cold were most strongly associated with DGF. When CART was used to analyse the outcome 1y-tl, as displayed in Figure 5.17, the variables that seemed to be important were, besides cold ischaemic time, primarily donor-related. Odds for 1y-tl were with 6:3, leave I, highest in the subset of donors in the age $\geq$ 66.5 with a BMI $\geq$ 23.4 kg\m^2 and last serum creatinine $\geq$ 1.24 mg/dl when the cold ischaemia time was $\geq$ 4.83 and the recipients end-stage renal disease was none of the following: tumor/ trauma/ surgery, diabetic nephropathy, pyelonephritis/ interstitial nephritis/ obstructive uropathy/ reflux uropathy or genetic nephropathy/ glomerulopathy non-FSGS. No 1y-tl was observed in older donors, however, when their BMI was $<$ 23.4 kg\m^2 (Leave E, odds = 0:23) or further down the branches accounting for ESRD in leaves F, odds=0:12 and, additionally cold ischaemia time in leave G, odds = 0:10. In the subset of donors in the age $<$ 45.5 years with a transplants cold ischaemia time $\geq$ 19.8 hours and a recipients ESRD which is neither drug-induced/toxic, nor Congenital dysplasia/hypoplasia/malformation with/without urinary tract malformation, pyelonephritis/interstitial nephritis/obstructive uropathy/reflux uropathy or tumor/trauma/surgery, the odds for 1y-tl was also increased to 5:3.

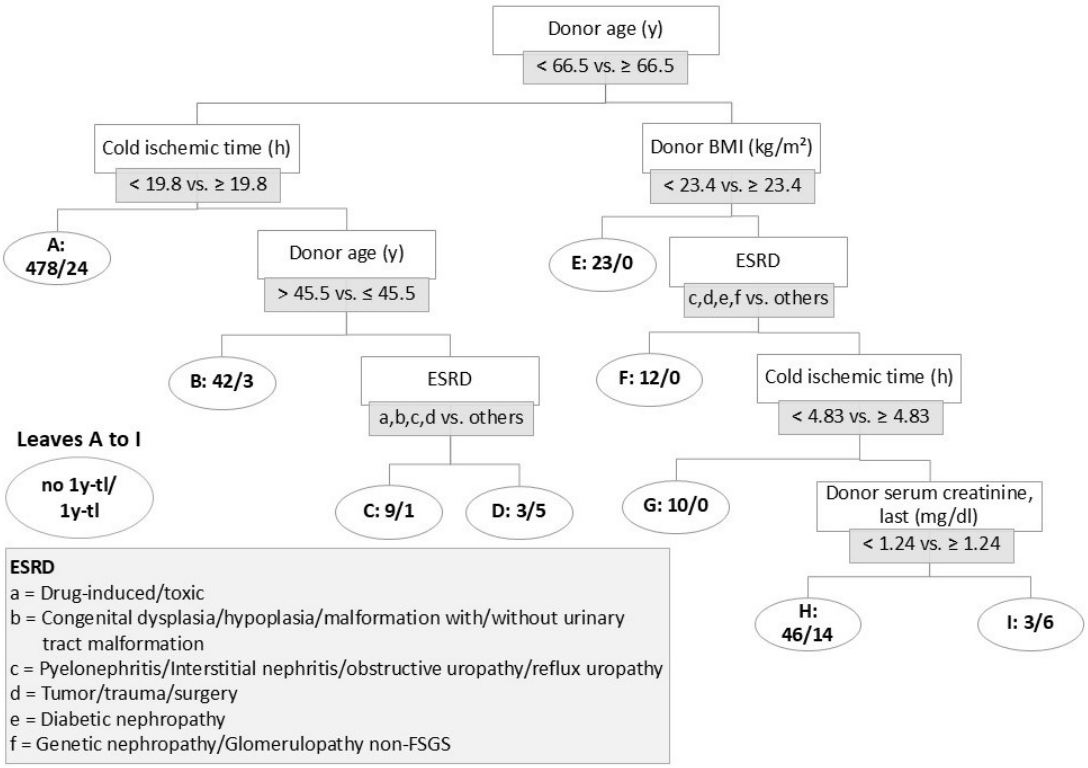


Figure 5.17: CART for outcome death-censored transplant loss within one year [Model M4, Figure 3.5]]

Predictive accuracy, sensitivity and specificity of the trees generated by CARD were measured based on confusion matrixes for both DGF and 1y-tl on the training and validation dataset and are sum up in Table 5.22. In any case, the specificity was high ($> .80$), while sensitivity was low ($< .35$) indicating that the rates of false negative predictions are higher than the false negative ones. Overall accuracy, however, was very good for predictions of 1y-tl (training set: .953, validation set: .905) but only okay for DGF (training set: .779, validation set: .627).

Table 5.22: Sensitivity, Specificity and predictive accuracy of trees generated by the CART algorithm for both DGF and 1y-tl

	Observed outcome				**Observed outcome**		
Training set	**DGF**		**Specificity/Sensitivity**	**Training set**	**1y-tl**		**Specificity/Sensitivity**
Predicted	***No DGF***	***DGF***	.954/.302	***Predicted***	***No loss***	***Loss***	.994/.114
No DGF	433	116	**Accuracy (95%-CI)**	***No loss***	644	50	**Accuracy (95%-CI)**
DGF	21	50	.779 (.744, .811)	***Loss***	6	11	0.921 (.899, .94)
Validation set	**DGF**		**Specificity/Sensitivity**	**Validation set**	**1y-tl**		**Specificity/Sensitivity**
Predicted	***No DGF***	***DGF***	.813/.174	***Predicted***	***No loss***	***Loss***	.953/.200
No DGF	91	38	**Accuracy (95%-CI)**	***No loss***	141	8	**Accuracy (95%-CI)**
DGF	21	8	.627 (.546, .702)	***Loss***	7	2	.905 (.848, .946)

Abbreviations: DGF, delayed graft function; 1y-tl, death-censored transplant loss within one year; CI, confidence interval

Looking at the tree in Figure 5.16 for DGF, two leaves stand out in which a better sensitivity for DGF could be achieved by further splitting. These are located

in particular in branches with opposing splits: low recipient and donor BMI but unfavourable ESRD and high BMI of the recipient with a short cold ischaemia time. The splitting could possibly be improved by less pruning or by taking other factors, especially histological variables, into account in these subgroups. However, a technical implementation of an automated update with additional variables to improve the prediction accuracy based on the leaves of a pre-specified regression tree is not known. Accordingly, further research into the implementation of the 2-step procedure with regression trees would be of interest. It is similar with the tree in Figure 5.17 for the prediction 1y-tl, where the subsets in the leaves A and H could also be further split or updated with further, yet not included variables.

Table 5.23: Multivariable logistic regression model for the outcome DGF based on variables identified by CART [Model 4 for DGF, Figure 5.24]

Predictor for DGF	**beta (SE)**	**OR (95% CI)**	**p-value**
Intercept	-3.821 (0.776)	–	<0.001
Cold ischaemia time (h)	0.025 (0.015)	1.025 (0.996, 1.055)	0.09
Recipient ESRD			
Others (ref)	–	–	–
Drug-induced/toxic or hypertensive nephropathy or nephrocalcinosis or tumour/trauma/surgery	-1.374 (0.455)	0.253 (0.094, 0.575)	0.003
Donr BMI (kg\m²)	0.057 (0.025)	1.059 (1.008, 1.113)	0.02
(Recipient BMI)² (kg\m²)²	0.001 (0.000)	1.001 (1.000, 1.002)	0.07
Donor age (years)	0.012 (0.008)	1.012 (0.997, 1.027)	0.13

Abbreviations: SE, standard error; CI, confidence interval; OR, odds ratio; h, hours; ESRD, end-stage renal disease; ref, reference.

In order to take into account the results of the CART analysis for DGF, the variable ESRD with its 13 groups was first dichotomised according to the tree. The variables of the tree and additionally the dichotomised ESRD variable, possible interactions that can be inferred from the tree structure (donor and recipient BMI) and quadratic terms of the continuous age and BMI variables were then considered in the regression model. The final regression model is shown in Table 5.23. Although prolonged cold ischaemia time was still associated with an increased incidence of DGF, this was no longer significant when taking into account the donor BMI and age, the recipient BMI and dichotomised ESRD. It was shown again that the model generated on the training data tended to underestimate the DGF risk on the validation dataset (calibration intercept = -.7; slope = .53). The AUC was better than for model 1, but worse than in model 2, with a value of .595. The average squared difference of the predicted probabilities with the estimated observed probabilities (ECI = 2.18) was so far the second highest after the one of model 1 and the scaled Brier score even negative -.12.

As for the reference model, donor age (squared) was the only independent predictor for 1y-tl besides cold ischaemia time. Calibration measures were therefore similar: AUC = .736, intercept = -.50, slope = 1.47. Overall measures were also comparable with ECI = .34 and a scaled Brier score of .05. Despite the high slope, the model tends to underestimate the risk of 1y-tl up to a predicted probability of 22%. Again, possible interactions of now donor age and donor BMI were included as well as two slightly different dichotomised versions of the recipients ESRD, none of which were significantly associated with 1y-tl in the binary logistic regression model.

Table 5.24: Multivariable logistic regression model for the outcome 1y-tl based on variables identified by CART [Model 4 for 1y-tl, Figure 5.24]

Predictor for 1y-tl	**beta (SE)**	**OR (95% CI)**	**p-value**
Intercept	-4.661 (0.708)	–	<0.001
Cold ischaemia time (h)	0.050 (0.020)	1.052 (1.012, 1.095)	0.01
(Donor age)² (years²)	0.0003 (0.0001)	1.0003 (1.0001, 1.0005)	0.002

Abbreviations: SE, standard error; OR, odds ratio; CI, confidence interval

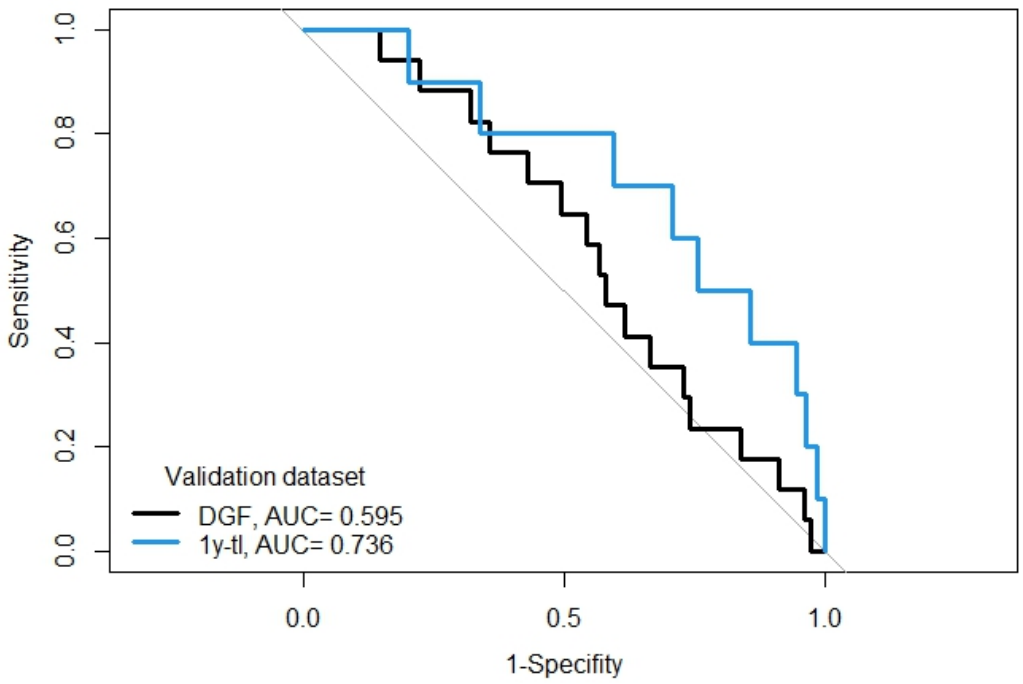


Figure 5.18: ROC of models based on variables identified by CART for prediction of DGF and 1y-tl on the corresponding validation datasets

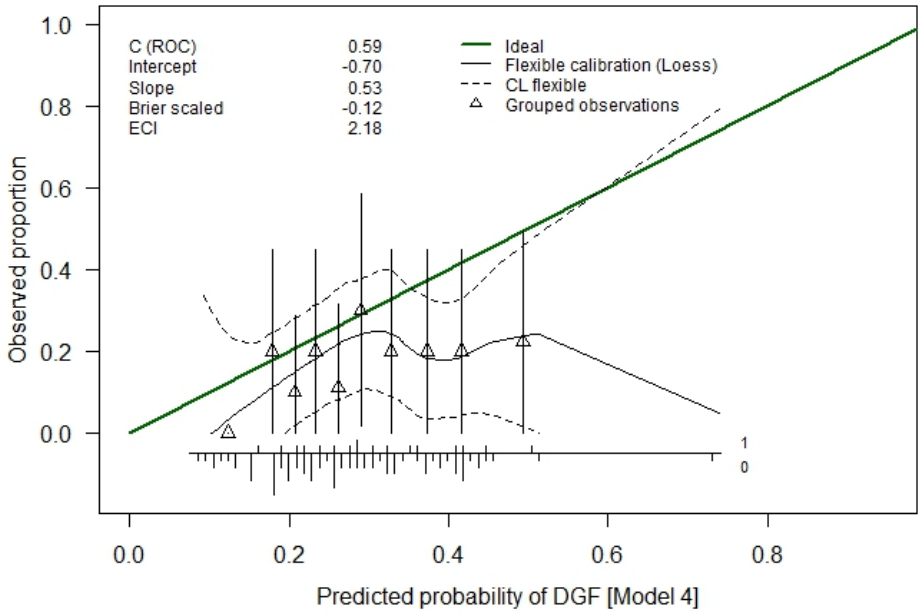


Figure 5.19: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on variables identified by CART for prediction of DGF on the validation dataset

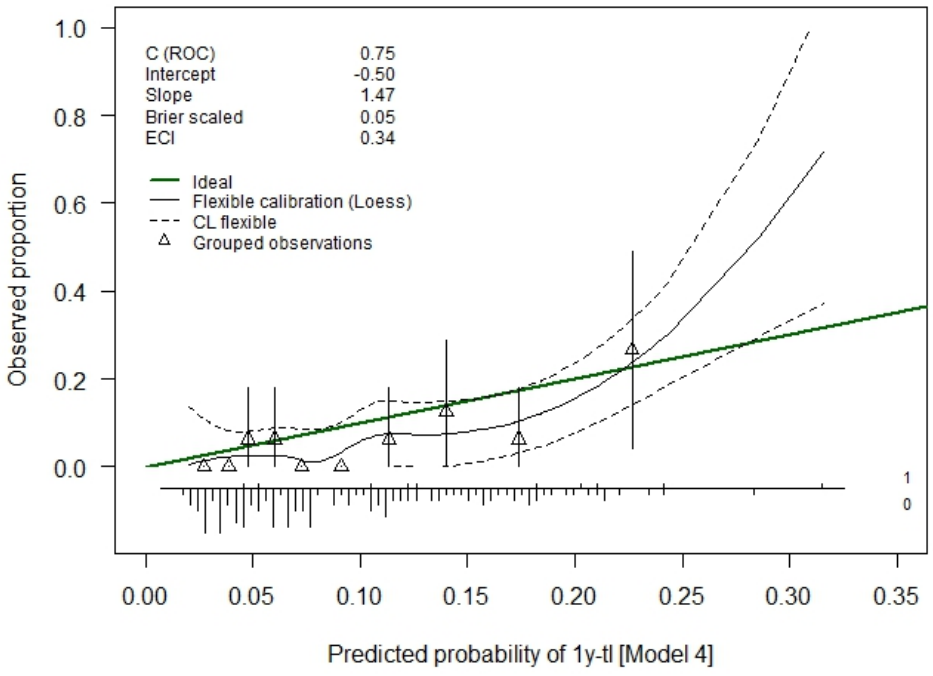


Figure 5.20: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on variables identified by CART for prediction of 1y-tl on the validation dataset

5.3.7 VSURF

In the first step "thresholding", VSURF selected 22 of the dummy-encoded variables, followed by 5 variables at the "interpretation" and "prediction" step regarding the outcome DGF, each. The final variables were recipient and donor BMI (kg/m^2), dialysis vintage, cold ischaemia time and donors last measure of the eGFR (mg/dl). The resulting misclassification rate (out-of-bag error) of the random forest was 19.8% with average variable importance (VI) $<$ .006. The donors last measure of eGFR (mg/dl) was then removed in the logistic regression modelling by the stepwise selection (Table 5.25).

With regard to 1y-tl, 27 dummy-encoded variables were selected in the "thresholding", followed by 5 variables in the "interpretation" and 4 variables in the "prediction" step. These variables were the donors type of brain death, donors last serum creatinine values (mg/dl), the donors last measure of eGFR (mg/dl) and the sum of HLA-A, -B and -DR mismatches. The resulting misclassification rate (out-of-bag error) of the random forest was 8.4% with average variable importance (VI) $<$ .006. As cold ischaemia time was not selected by VSURF, it was forced into the model when developing the final logistic regression-based model, adjusted for the four selected variables. Finally, besides cold ischaemia time only the sum of HLA-A, -B and -DR mismatches was selected (Table 5.26).

Table 5.25: Multivariable logistic regression model for the outcome DGF based on variables identified by VSURV [Model 5 for DGF, Figure 5.24]

Predictor for DGF	**beta (SE)**	**OR (95% CI)**	**p-value**
Intercept	-0.221 (2.478)		0.929
Cold ischaemia time (h)	0.030 (0.013)	1.030 (1.005, 1.057)	0.020
Donor BMI (kg/m^2)	0.076 (0.022)	1.079 (1.033, 1.128)	$<$0.001
(Recipient BMI) ^{2} (kg/m^2)^2	0.006 (0.003)	1.006 (1.000, 1.013)	0.06
Dialysis vintage (years)	0.081 (0.039)	1.084 (1.004, 1.170)	0.037
Recipient BMI (kg/m^2)	-0.300 (0.183)	0.740 (0.516, 1.045)	0.101

Abbreviations: h, hours; SE, standard error; CI, confidence interval; OR, odds ratio.

The calibration plot with the corresponding intercept and slope of the predicted DGF-values indicate a general underestimation of the risk (intercept = -.53, slope =.62) on the validation dataset (Figure 5.22). Consequently, the scaled Brier score was with -0.01 very poor. The estimated calibration index (ECI) for DGF was 1.42.

Table 5.26: Multivariable logistic regression model for the outcome 1y-tl based on variables identified by VSURV [Model 5 for 1y-tl, Figure 5.24]

Predictor for 1y-tl	beta (SE)	OR (95% CI)	p-value
Intercept	-3.781 (0.415)		<0.001
Cold ischaemia time (h)	0.056 (0.018)	1.058 (1.022, 1.096)	0.001
Sum of HLA-A, -B and -DR mismatches	0.252 (0.095)	1.286 (1.072, 1.555)	0.008

Abbreviations: HLA, humane leucozyte antibody; SE, standard error; CI, confidence interval; OR, odds ratio.

Looking at the calibration plot for 1y-tl, Figure 5.23, one will notice a similar calibration compared to previous models predicting this outcome (intercept = -.36, slope = 2.7). The model tends to underestimate risks below the cohorts prevalence (9%) and overestimates risks above this threshold. However, the predicted probabilities only assume values between 0% and 20%, which again indicates an underestimation of the risk. Taking the scaled Brier score of .06 into account, the overall calibration was again very poor while an ECI of .26 is rather good.

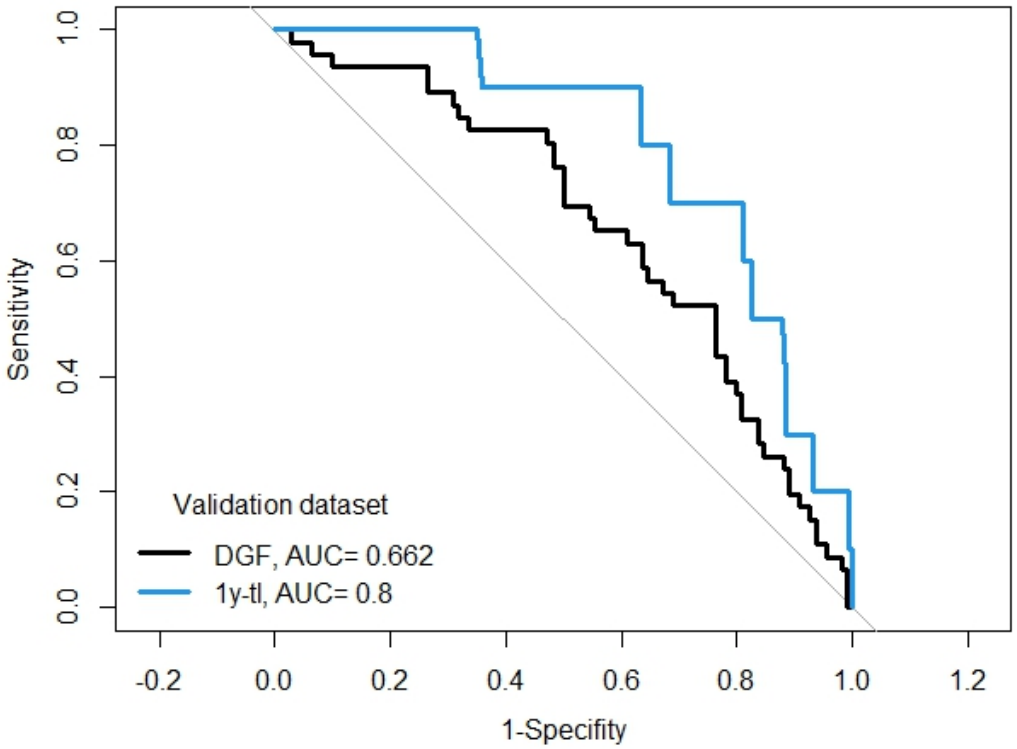


Figure 5.21: ROC of models based on variables identified by VSURF for prediction of DGF and 1y-tl on the corresponding validation datasets

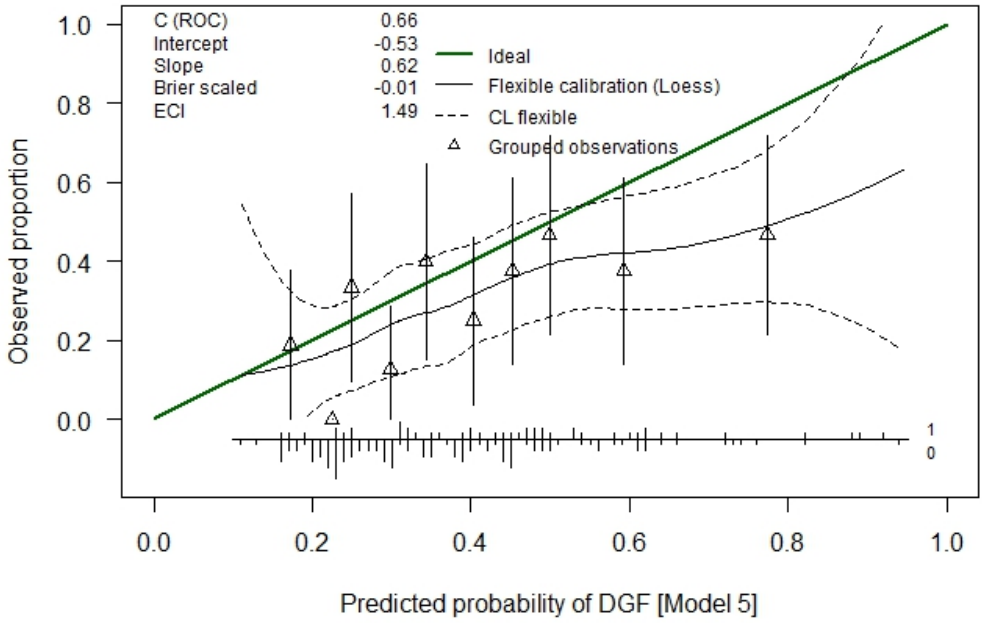


Figure 5.22: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on variables identified by VSURF for prediction of DGF on the validation dataset

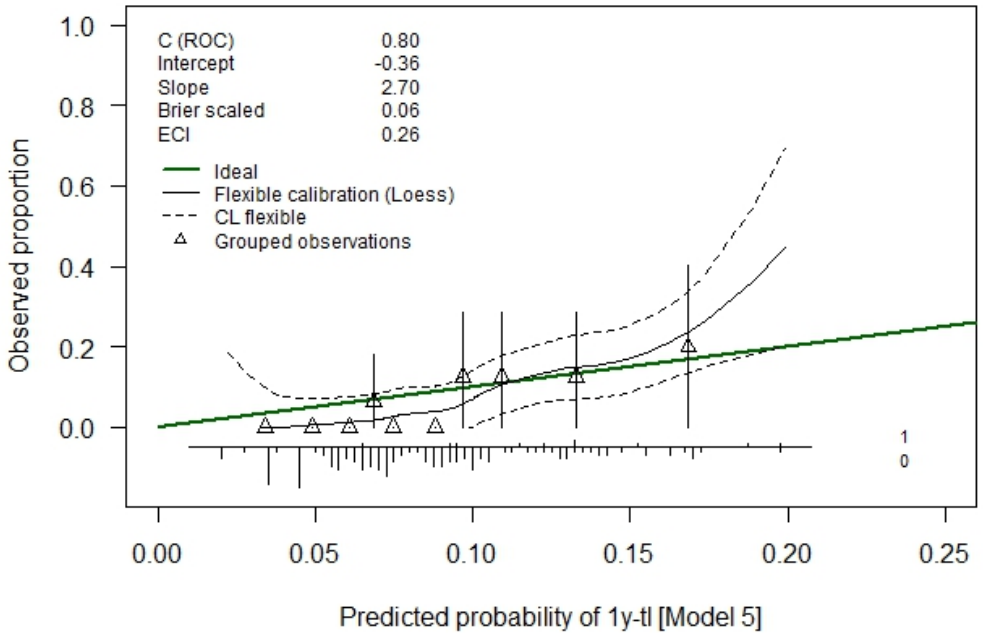


Figure 5.23: Calibration plot and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of the logistic regression model based on variables identified by VSURF for prediction of 1y-tl on the validation dataset

5.3.8 Summary: pre-selection of potential clinical predictors

Before the study began, the minimum required sample size for the analysis of the outcome DGF, as well as the distribution of observations by study centre, was determined in the study plan. In addition, the sources for the training and external validation datasets were defined. A large proportion of the validation cohort came from a sample of the German Organ Procurement Organization that consisted of marginal quality organs. A comparison of the distributions of the donor-, recipient- and transplant-specific variables had shown that the organs in the validation cohort tended to be of poorer quality based on these factors. Donors in the validation cohort were significantly older, had a higher prevalence of diabetes mellitus and hypertension, higher BMI and serum creatinine values and a lower eGFR rate. The recipients in the validation cohort were also significantly older, had been on dialysis for longer and were more often positive for CMV IgG. Regarding transplant-specific variables, the number of HLA-DR mismatches was also higher in the validation cohort. Although warm ischaemia time was not considered in the score development, it was longer by a median of 4 minutes. This imbalance was subsequently reflected in the validation of the models generated in Sections 5.3.2 to 5.3.7. The systematic underestimation of the predicted risks on the validation dataset shows the need for a recalibration of the models with regard to this bias.

Due to missing values in the fewer than planned datasets collected, and a subsequent change in the participation of the study centres, cuts had to be made in the planned modelling. In Sections 5.3.2 to 5.3.7, several methods were used to reduce potential predictors and, based on this, binary logistic regression models were created using stepwise variable selection. As a reference model to evaluate the effect of missing values, a multiple imputed was created in Section 5.3.2 and a binary logistic regression model developed for each outcome.

A summary of the model performance is given in the Table 5.27. In the first model, Section 5.3.3, the preselection was omitted. As a result, a relatively high number of variables (11 and 7) were included for both DGF and 1y-tl, which led to overfitting and ultimately poor performance (high ECI, low AUC) on the validation dataset.

When considering only the outcome DGF, the models performed best after preselection using the univariate P-value (Section 5.3.4: AUC = .661, ECI = 1.65) and LASSO (Section 5.3.5: AUC = .753, ECI = 5.42), with 9 and 10 predictors included, respectively. The univariable P-value and LASSO were also the best methods for reducing variables when predicting 1y-tl (AUC = .822; ECI = .23 and AUC = .753; ECI = .17, respectively). The prevalence of transplant loss in the first year after transplantation is generally lower than that of DGF, which is 27% here, at around

9% in the dataset. Accordingly, a larger number of cases would have been needed for modelling than for DGF. According to the 10 events per predictor rule of thumb, no more than 6 independent clinical and histopathological variables should be considered for the 1y-tl outcome. The final Models 2 and 3 for predicting 1y-tl include 6 and 3 clinical variables, respectively, which is still within the scope of the rule of thumb.

The advantage of CART became particularly apparent when factor variables with more than four levels were included. This applied to the 17 levels of the ESRD and the 10 levels of the donor's cause of death. Some of these were only weakly expressed and were combined into variables with fewer subgroups in CART. This made it possible to determine which illnesses or causes of death could be statistically combined to make a more accurate prediction. However, there was a high proportion of unknown diagnoses for both variables, which is why the variables combined on the basis of the data are considered susceptible to bias. The subgroups were already categorised on the basis of medical expertise, which is more relevant than statistical power for future prediction. In subsequent modelling steps, the combined variables were therefore not considered further.

When considering the results of all logistic regression models, it must be taken into account that the cold ischaemia time had to be included as a predictor. This was done for the reason that the effect of a potential reduction of this in hours should be taken into account when calculating the point scores. As the methods of preselection (CART, P-value $< .25$, and LASSO) have shown, cold ischaemia time is associated with DGF and 1y-tl, independently of confounding variables. Nevertheless, it may be possible to generate models with better performance if multivariate adjustment makes the cold ischaemia time less significant.

Ultimately, a few variables emerged as the most relevant in all the models, and these were used for the final modelling of Step 1.

Table 5.27: Summary of model performances with model names according to Figure 5.24

Outcome	Method	Model	No. of predictors (excl. intercept)	AUC / Accuracy*	Calibration intercept / slope		Scaled Brier	ECI
DGF	Logistic regression on multiple	***RefMod***	6	0.688	-0.64	0.77	0.01	1.79
1y-tl	imputed dataset	***RefMod***	2	0.724	-0.37	1.54	0.05	0.41
DGF	Logistic regression with stepwise variable selection	***M1 = Model 1***	11	0.509	-1.53	0	-0.78	7.88
1y-tl		***M1 = Model 1***	7	0.762	-0.52	0.78	-0.02	0.5
DGF	Logistic regression for variables	***Model 2***	9	0.661	-0.53	0.61	-0.01	1.65
1y-tl	with univariable p-value <.25	***Model 2***	6	0.822	-0.55	1.39	0.09	0.23
DGF	LASSO	***M3***	13	0.743*				
1y-tl		***M3***	4	0.879*				
DGF	Logistic regression	***Model 3***	10	0.753	-1.3	0.58	-0.08	5.42
1y-tl	after LASSO	***Model 3***	3	0.753	-0.37	1.79	0.05	0.17
DGF	CART	***M4***	7	0.627*				
1y-tl		***M4***	6	0.905*				
DGF	Logistic regression	***Model 4***	5	0.595	-0.7	0.53	-0.12	2.18
1y-tl	after CART	***Model 4***	2	0.736	-0.5	1.47	0.05	0.34
DGF	Logistic regression	***Model 5***	5	0.662	-0.53	0.62	-0.01	1.49
1y-tl	after VSURF	***Model 5***	2	0.8	-0.36	2.7	0.06	0.26

Colors: red/orange = worst measures by outcome; green/turkish green = best measures by outcome
Abbreviations: ECI, estimated calibration index; AUC, area under the receiver operating curve; LASSO, Least Absolute Shrinkage and Selection Operator; CART, Classification And Regression Tree; VSURF, Variable Selection Random Forest.

5.4 Score development, step 1: clinical data score

Based on existing knowledge about possible influencing factors from the literature (Sections 2.1 and 2.2) and our own previous variable selections (Sections 5.3.2 to 5.3.7), the number of variables to be included in the modeling of the clinical score could be reduced. For DGF, the relevant variables identified were the donor and recipient BMI, the recipient CMV-IgG status, the number of HLA-DR mismatches, dialysis vintage and cold ischaemia time. With regard to 1y-tl, besides cold ischaemic time, the age of the donor and the sum of the HLA-A, -B and -DR mismatches were considered most relevant. The resulting logistic regression models are sum up in Table 5.28 and Table 5.29 and can be expressed as:

$Logit_{S1}[DGF] = log\left(\frac{P[DGF]}{(1-P[DGF])}\right)$ = -5.66 + .035 (Cold ischaemia time [h]) + .809 (HLA-DR mismatches = two) + .332 (HLA-DR mismatches = one) + .071 (Donor BMI [kg/m^2]) + .045 (Recipient BMI [kg/m^2]) + .099 (Dialysis vintage [y]) + .756 (Recipient CMV-IgG = positive).

Here, the coefficient for unknown recipient CMV-IgG status was left out since it was not significant (p $>$ 0.05). Also, in practice, an "unknown" status is not considered to be informative. Respectively, for the outcome 1y-tl the probability can be estimated by:

$Logit_{S1}[1y-tl] = log\left(\frac{(P[1y-tl])}{(1-P[1y-tl])}\right)$ = -4.97 + 0.053 (Cold ischaemia time [h]) + 0.026 (Donor age [y]) + 0.186 (Sum of HLA-A, -B, and -DR mismatches).

To calculate the predictive probability P of each outcome, the models can be represented in the following form:
P[DGF] = $\frac{exp(Logit_{S1}[DGF])}{(1+exp(Logit_{S1}[DGF]))}$ and
P[1y-tl] = $\frac{exp(Logit_{S1}[1y-tl])}{(1+exp(Logit_{S1}[1y-tl]))}$, respectively.

Table 5.28: Multivariable logistic regression model for the outcome DGF: final model after step 1 [Figure 5.24]

Predictor for DGF	beta (SE)	OR (95% CI)	p-value
Intercept	-5.556 (0.890)	-	<0.001
Cold ischaemia time (h)	0.035 (0.013)	1.035 (1.009, 1.063)	0.009
HLA-DR mismatches			
None (ref)	-	-	
One	0.332 (0.242)	1.393 (0.872, 2.262)	0.17
Two	0.799 (0.295)	2.224 (1.249, 3.982)	0.007
Recipient BMI (kg/m²)	0.042 (0.020)	1.043 (1.004, 1.085)	0.03
Donor BMI (kg/m²)	0.071 (0.022)	1.073 (1.028, 1.121)	0.001
Dialysis vintage (y)	0.096 (0.039)	1.101 (1.019, 1.189)	0.01
Recipient CMV-IgG positivity			
Negative (ref)	-	-	
Positive	0.754 (0.283)	2.126 (1.236, 3.766)	0.008
Unknown	0.418 (0.292)	1.518 (0.865, 2.728)	0.15

Abbreviations: HLA, humane leucozyte antibody; SE, standard error; CI, confidence interval; OR, odds ratio; y, years.

Table 5.29: Multivariable logistic regression model for the outcome 1y-tl: final model after step 1 [Figure 5.24]

Predictor for 1y-tl	beta (SE)	OR (95% CI)	p-value
Intercept	-4.978 (0.644)		<0.001
Donor age (y)	0.026 (0.010)	1.026 (1.006, 1.047)	0.01
Sum of HLA-A, -B and -DR mismatches	0.186 (0.096)	1.205 (1.002, 1.460)	0.05
Cold ischaemia time (h)	0.053 (0.018)	1.054 (1.019, 1.092)	0.003

Abbreviations: HLA, humane leucozyte antibody; SE, standard error; CI, confidence interval; OR, odds ratio; h, hours; y, years.

As described in the methods Section 3.3.7, given the logistic regression estimates, integer-based risk scores were estimated. Cold ischaemia time (cutoffs = 2, 6, 10, 14 and 18 hours), donor and recipient BMI (cutoffs = 18.5, 25, 30 kg/m²) and the dialysis vintage (cutoffs = 1, 3, 4, 6 years) were transformed into categorical variables for developing a point score for prediction of DGF. To build a point score for prediction of 1y-tl, donor age was grouped by cutoffs 30, 40, 50 and 60 years, sum of HLA-A, -B and -DR mismatches by cutoffs = 0, 1, 2, 3, 4, 6 and cold ischaemia time as in the case of DGF. Templates for calculating point-based risk scores for DGF and 1y-tl, already including results of the second score development step, are given in Figures 5.24 and 5.25[30].

Risk factor	Addition to risk score						Risk score
Donor BMI (kg/m²)	< 18.5	18.5 to < 25		25 to < 30		>= 30	
	-2	0		+3		+6	
Recipient BMI (kg/m²)	< 18.5	18.5 to < 25		25 to < 30		>= 30	
	-2	0		+2		+3	
Recipient CVM IgG	Negative			Positive			
	0			+5			
Number of HLA-DR mismatches	Less than two			Two			
	0			+6			
Dialysis vintage in years	< 1	1 to < 3	3 to < 4		4 to < 6	>= 6	
	0	+1	+2		+3	+5	
Cold ischaemia time in hours	< 2	2 to <6	6 to <10	10 to < 14	14 to < 18	>= 18	
	0	+1	+2	+3	+4	+5	
Total score without histological assessment	- 4 to 11 points		12 to 18 points (18% to 36% risk)		19 to 30 points		
	Low risk		Medium risk: histological assessment advised		Increased risk		
Number of glomeruli	< 20	20 to < 30	30 to < 40	>= 40			
	0	-2	-4	-7			
Banff cv	cv0, cv1 or cv2		cv3				
	0		+5				
				Total score =			

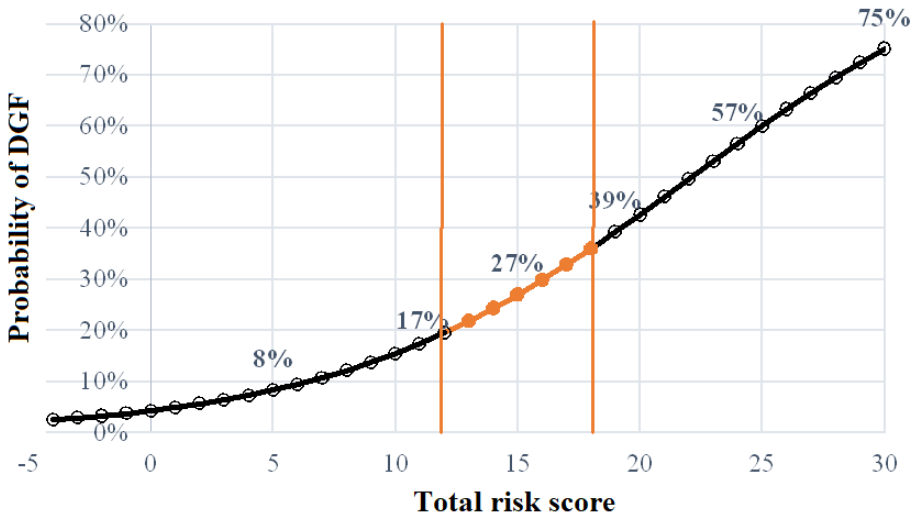


Figure 5.24: Point scoring tool to estimate risk of DGF[30]. For each expected prolongation of the cold ischaemia time by 4 (2 to 6) hours, one additional score point must be added to the **"Total score"** when *number of glomeruli* and *Banff cv* are taken into account.

Risk factor	**Addition to risk score**					**Risk score**
Donor age in years	< 30	30 to < 40	40 to < 50	50 to < 60	>= 60	
	0	+1	+3	+4	+6	
Sum of HLA-A, -B and -DR mismatches	0	1	2	3	4 or 5	6
	0	+1	+2	+3	+4	+5
Cold ischaemia time in hours	< 2	2 to <6	6 to <10	10 to < 14	14 to < 18	>= 18
	0	+1	+2	+3	+4	+5
Total score without histological assessment						
Banff ct	ct0 or ct1			ct2 or ct3		
	0			+6		
					Total score =	

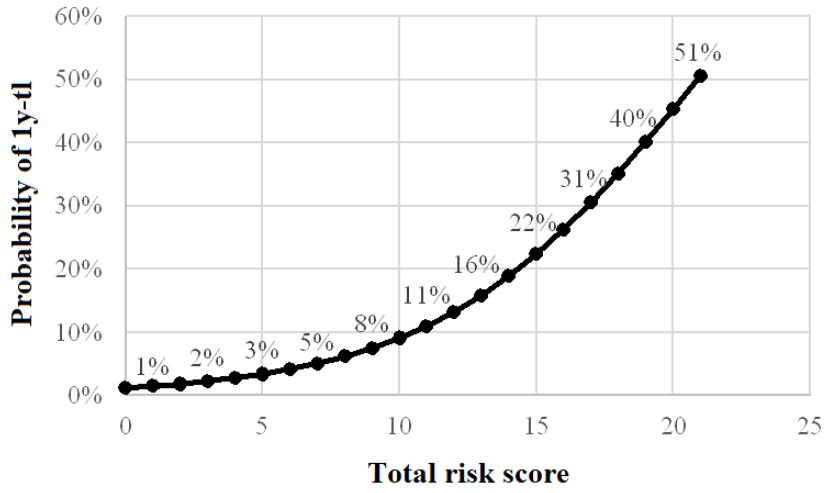


Figure 5.25: Point scoring tool to estimate risk of one-year transplant loss (1y-tl)[30]. For each expected prolongation of the cold ischaemia time by 4 (2 to 6) hours, one additional score point must be added to the **"Total score"** when *Banff ct* is taken into account.

5.5 Score development, step 2: clinical and histological data

As displayed in Tables 5.32[30] and 5.33[30], there was no histological variable significantly univariable associated with DGF and only few with 1y-tl. Amongst the ones associated with 1y-tl were Banff ci (ci2 vs. ci0, OR [95%-CI] = 3.45 [.95-10.07], p=.035) and, with almost identical values, Banff ct, along with the ratio of the number of globally sclerosed glomeruli and all glomeruli (OR [95%-CI] = 10.94 [1.06-89.89], p=.032). As the number of observations with the highest lesion scores were in general low, Banff cv and ci were dichotomised before medelling to cv0 & cv1 & cv2 vs. cv3 and ci0 & ci1 vs. ci2 & ci3. The regression models derived in step 1, Section 5.4, are then updated by including both their logit functions of the predicted probabilities for DGF ($Logit_{S1}[DGF]$) and, respectively, 1y-tl ($Logit_{S1}[1y-tl]$) and the histological variables into a new logistic regression model. As there were no univariable associations between any histological variable described in Tables 5.32[30] and 5.33[30] and DGF, expectations were low to find any multivariable association.

Plotting the histograms of the predicted probabilities of DGF separated by observed DGF, one would expect a bimodal distribution with a valley at the point best separating DGF cases from non-cases. This probability would also be the best threshold according to the AUC. Yet, after the first step, the histogram (Figure 5.26) showed a rather unimodal (normal) distribution of the probabilities with a strong overlap around the observed DGF-prevalence (black vertical line) in the training set. As none of the histological variables was selected when multivariable adjusting for either each of the predictors separately from the final model of step 1, nor the score derived, models were additionally trained on subsets of observations based on the predicted probabilities. In some of these subsets (intervals including the prevalence of DGF), histological variables showed a significant association with DGF. The interval with the widest range including significant histological predictors (predicted probability $>$ 0.18 and $\leq$ 0.36, adding number of glomeruli and dichotomised Banff cv) was then chosen as relevant for updating and recalibration of the model derived in step 1.

Although an update with Banff ct and simultaneous recalibration of the model derived by step 1 for the outcome 1y-tl was possible on the entire training dataset, additional analyses were performed on subsets defined as in the case of DGF. Yet, besides Banff ct no further histological variable was associated with 1y-tl.

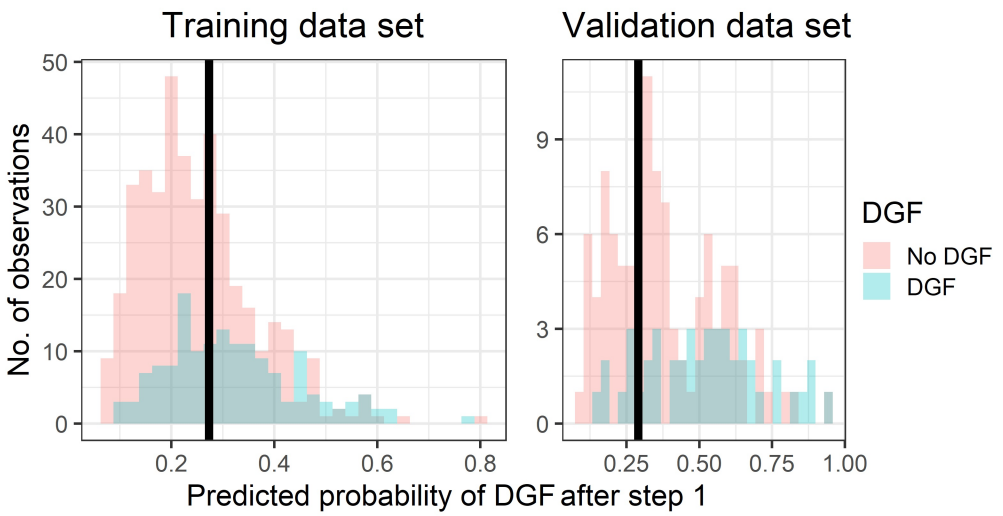


Figure 5.26: Predicted probabilities of DGF after step 1 on training and validation dataset. The vertical, black line represents the DGF prevalence values.

The resulting logistic regression models are sum up in Table 5.30 and Table 5.31 and can be expressed as:

IF

$Logit_{S1}[DGF] <$ 0.18 OR $Logit_{S1}[DGF] >$ 0.36

THEN

$Logit_{S2}[DGF] = Logit_{S1}[DGF]$

ELSE

$Logit_{S2}[DGF] =$ .38 $Logit_{S1}[DGF]$ + .74 (Banff cv = cv3) - .035 (Number of glomeruli) +.035*(expected prolongation of cold ischaemia time [h]).

Table 5.30: Multivariable logistic regression model for the outcome DGF: final model after step 2 for subset of cohort with values of $Logit_{S1}[DGF]$ between .18 and .36 [Figure 5.24]

Predictor for DGF	**beta (SE)**	**OR (95% CI)**	**p-value**
Intercept	-0.31 (0.44)		0.49
$Logit_{S1}[DGF]$	0.38 (0.82)	1.46 (0.29, 7.35)	0.65
Banff cv			
cv0, cv1 or cv2 (ref)			
cv3	0.74 (0.33)	2.10 (1.08, 4.02)	0.03
Number of glomeruli	-0.04 (0.01)	0.96 (0.94, 0.99)	0.01

Abbreviations: SE, standard error; CI, confidence interval; OR, odds ratio.

Respectively, for the outcome 1y-tl the probability can be estimated by:

$Logit_{S2}[1y-tl]$ = -2.6 + 0.98 ($Logit_{S1}[1y-tl]$) + 1.26 (Banff ct = ct2 or ct3) + .053*(expected prolongation of cold ischaemia time [h]).

Table 5.31: Multivariable logistic regression model for the outcome 1y-tl: final model after step 2 [Figure 5.24]

Predictor of 1y-tl	beta (SE)	OR (95% CI)	p-value
Intercept	-2.60 (0.16)		<0.001
$Logit_{S1}[1y-tl]$	0.99 (0.21)	2.68 (1.79, 4.09)	<0.001
Dichotomised Banff ct			
ct0 or ct1 (ref)	-	-	
ct2 or ct3	1.26 (0.56)	3.51 (1.06, 10.04)	0.03

Abbreviations: SE, standard error; CI, confidence interval; OR, odds ratio.

Of note: as displayed in Figure 5.27, there was a strong correlation between the dichotomised variables Banff ci and Banff ct, and the ration of the number of globally sclerosed glomeruli and the total number of glomeruli, which is why these variables can be considered to be interchangeable when included in a prediction mode.

Banff ci vs. ct	*ci0 or ci1*	*ci2 or ci3*
ct0 or ct1	691	0
ct2 or ct3	1	19

Optimal threshold of ratio of globally sclerosed glomeruli/ glomeruli to discriminate amongst kidneys with Banff ci = ci0 or ci1 vs. ci2 or ci3 according to ROC analysis (AUC = .888) equals 19%.

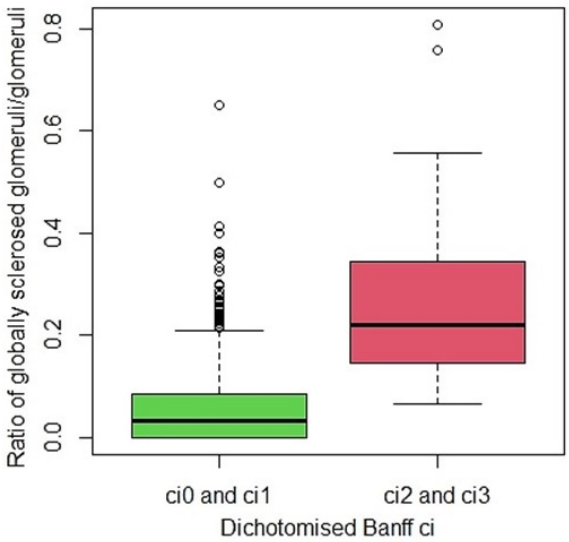


Figure 5.27: Correlation between dichotomised Banff ci and Banff ct, and the ratio of the globally sclerosed glomeruli and all glomeruli

Table 5.32: Variables derived by histopathological evaluation (part 1)[30].

Variable	No DGF	DGF	Univar. OR (95% CI)	No TXP-loss	Loss within 1y	Univar. OR (95% CI)
N	***454***	***166***		***650***	***61***	
Banff i [i0] (N(%))†	454 (100.0)	166 (100.0)	NA	650 (100.0)	61 (100.0)	NA
Banff t [t0] (N(%))†	454 (100.0)	166 (100.0)	NA	650 (100.0)	61 (100.0)	NA
Banff v [v0] (N(%))†	454 (100.0)	166 (100.0)	NA	650 (100.0)	61 (100.0)	NA
Banff g (N(%))†						
g0	450 (99.1)	164 (98.8)	1	639 (98.3)	60 (98.4)	1
g1	1 (0.2)	1 (0.6)	2.74 (0.11-69.65, p=.476)	4 (0.6)	0 (0.0)	NA
g2	2 (0.4)	1 (0.6)	1.37 (0.06-14.41, p=.797)	6 (0.9)	1 (1.6)	1.77 (0.09-10.63, p=.598)
g3	1 (0.2)	0 (0.0)	NA	1 (0.2)	0 (0.0)	NA
Banff ptc [ptc0] (N(%))†	454 (100.0)	166 (100.0)	NA	650 (100.0)	61 (100.0)	NA
Banff ci (%)						
ci0	399 (87.9)	150 (90.4)	1	580 (89.2)	48 (78.7)	1
ci1	44 (9.7)	11 (6.6)	0.66 (0.32-1.28, p=.244)	55 (8.5)	8 (13.1)	1.76 (0.74-3.72, p=.166)
ci2	10 (2.2)	4 (2.4)	1.06 (0.29-3.23, p=.918)	14 (2.2)	4 (6.6)	3.45 (0.95-10.07, p=.035)
ci3	1 (0.2)	1 (0.6)	2.66 (0.10-67.55, p=.490)	1 (0.2)	1 (1.6)	12.08 (0.47-308.77, p=.080)
Banff ct (N(%))†						
ct0	280 (61.7)	101 (60.8)	1	395 (60.8)	35 (57.4)	1
ct1	163 (35.9)	60 (36.1)	1.02 (0.70-1.48, p=.915)	241 (37.1)	21 (34.4)	0.98 (0.55-1.71, p=.954)
ct2	10 (2.2)	4 (2.4)	1.11 (0.30-3.40, p=.864)	13 (2.0)	4 (6.6)	3.47 (0.94-10.43, p=.037)
ct3	1 (0.2)	1 (0.6)	2.77 (0.11-70.54, p=.472)	1 (0.2)	1 (1.6)	11.29 (0.44-289.54, p=.089)
Banff cv (N(%))						
cv0	167 (36.8)	61 (36.7)	1	249 (38.3)	23 (37.7)	1
cv1	86 (18.9)	29 (17.5)	0.92 (0.55-1.53, p=.760)	120 (18.5)	9 (14.8)	0.81 (0.35-1.75, p=.610)
cv2	90 (19.8)	28 (16.9)	0.85 (0.50-1.42, p=.542)	120 (18.5)	15 (24.6)	1.35 (0.67-2.67, p=.387)
cv3	110 (24.2)	48 (28.9)	1.19 (0.76-1.87, p=.437)	160 (24.6)	14 (23.0)	0.95 (0.46-1.87, p=.878)
NA	1 (0.2)	0 (0.0)		1 (0.2)	0 (0.0)	
Banff cg (N(%))†						
cg0	452 (99.6)	166 (100.0)	1	647 (99.5)	61 (100.0)	1
cg1a	1 (0.2)	0 (0.0)	NA	1 (0.2)	0 (0.0)	NA
cg1b	1 (0.2)	0 (0.0)	NA	2 (0.3)	0 (0.0)	NA

Abbreviations: Univar., univariable; OR, odds ratio; CI, confidence interval; NA, not available; y, year.
* Non-parametric Mann-Whitney U-test
† Excluded before variable selection.

Table 5.33: Variables derived by histopathological evaluation (part 2)[30].

Variable	No DGF	DGF	Univar. OR (95% CI)	No TXP-loss	Loss within 1y	Univar. OR (95% CI)
N	***454***	***166***		***650***	***61***	
Acute tubular necrosis (N(%))†						
No	224 (49.3)	88 (53.0)	1	329 (50.6)	36 (59.0)	1
Yes	228 (50.2)	78 (47.0)	0.87 (0.61-1.24, p=0.447)	319 (49.1)	25 (41.0)	0.72 (0.42-1.21, p=0.220)
Unknown	2 (0.4)	0 (0.0)		2 (0.3)	0 (0.0)	
Banff mm (N(%))†						
mm0	451 (99.3)	166 (100.0)	1	646 (99.4)	61 (100.0)	1
mm1	1 (0.2)	0 (0.0)	NA	2 (0.3)	0 (0.0)	NA
mm2	1 (0.2)	0 (0.0)	NA	1 (0.2)	0 (0.0)	NA
mm3	1 (0.2)	0 (0.0)	NA	1 (0.2)	0 (0.0)	NA
Banff ah (N(%))						
ah0	178 (39.2)	64 (38.6)	1	257 (39.5)	19 (31.1)	1
ah1	199 (43.8)	74 (44.6)	1.03 (0.70-1.53, p=0.866)	284 (43.7)	30 (49.2)	1.43 (0.79-2.64, p=0.243)
ah2	57 (12.6)	20 (12.0)	0.98 (0.54-1.73, p=0.935)	82 (12.6)	7 (11.5)	1.15 (0.44-2.73, p=0.755)
ah3	20 (4.4)	8 (4.8)	1.11 (0.44-2.57, p=0.810)	27 (4.2)	5 (8.2)	2.50 (0.78-6.82, p=0.090)
No. of arteries (median [IQR])	3 [2, 4]	3 [2, 4]	1.00 (0.91-1.09, p=0.933)	3 [2, 4]	3 [2, 4]	1.04 (0.92-1.17, p=0.507)
No. of glomeruli (median [IQR])	28 [21, 36]	26 [21, 35]	0.99 (0.98-1.01, p=0.266)	28 [21, 35]	32 [20, 37]	1.01 (0.99-1.03, p=0.195)
No. of GSG (median [IQR])†	1 [0, 2]	1 [0, 2]	0.98 (0.92-1.03, p=0.494)	1 [0, 2]	1 [0, 3]	1.04 (0.97-1.09, p=0.206)
Ratio of the number of GSG to all glomeruli (median [IQR])	0.03 [0, 0.09]	0.04 [0, 0.09]	0.49 (0.06-3.23, p=0.478)	0.03 [0, 0.09]	0.05 [0, 0.12]	10.94 (1.06-89.89, p=0.032)
Microthrombus (N(%))						
No	446 (98.2)	162 (97.6)	1	634 (97.5)	60 (98.4)	1
Yes	7 (1.5)	4 (2.4)	1.57 (0.41-5.28, p=0.474)	15 (2.3)	1 (1.6)	0.70 (0.04-3.57, p=0.737)
Unknown	1 (0.2)	0 (0.0)		1 (0.2)	0 (0.0)	

Abbreviations: IQR, interquartile range; Univar., univariable; No., number; GSG, globally sclerosed glomeruli; CI, confidence interval; OR, odds ratio; y, year; NA, not available; TXP, transplant.
* Non-parametric Mann-Whitney U-test
† Excluded before variable selection

In order to estimate the effect of bias due to missing values when obtaining the estimators, the two steps of score generation were repeated on the multiple imputed dataset. It was shown that the estimators for the outcome 1ytl were practically identical, which is partly due to the low number of missing values of the included variables (Figure 5.28[30]). In the evaluation of the models for DGF, the model for the variables selected in step 1 was again calculated first. Subsequently, on the subset of the cohort whose predicted risk was in the same interval (18% - 36%), the model was updated and recalibrated with the histological variables Banff cv and number of glomeruli. The estimators of the individual variables included in the modelling of step 2 are shown in Figure 5.29[30]. There was a tendency to underestimate the odds for DGF in the number of HLA-DR mismatches (2 vs. 0) and the CMV-IgG status of the recipient, while at the same time overestimating high Banff cv scores in the training compared to the multiple imputed data.

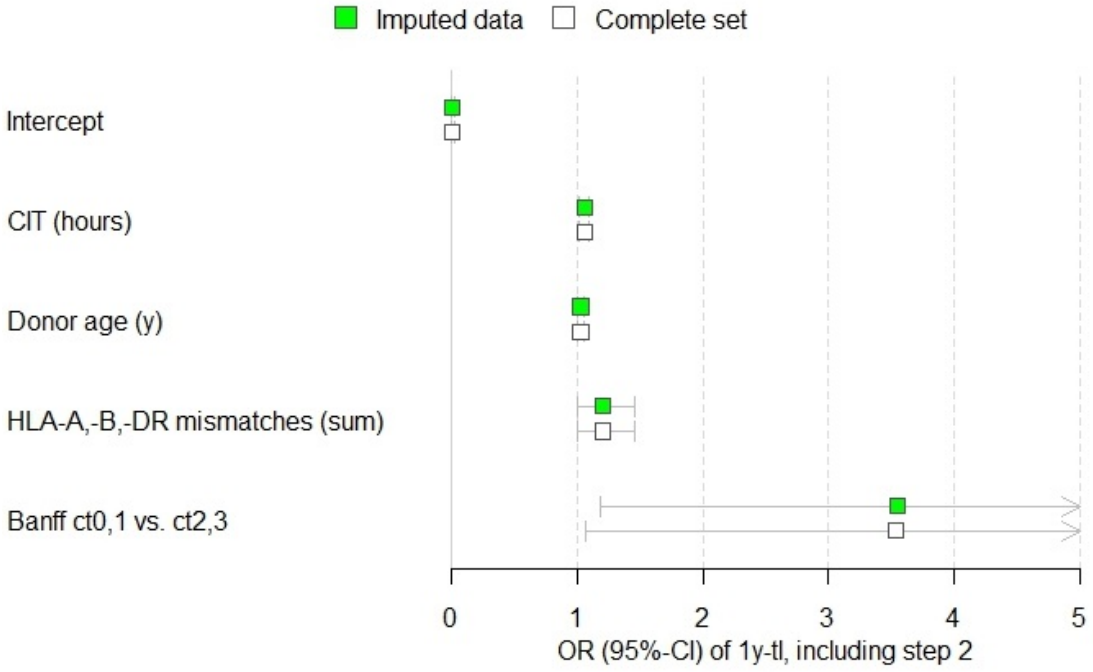


Figure 5.28: Regression estimates of 2-Step score for 1y-tl on imputed vs. non-imputed dataset[30].
Abbreviations: CIT, cold ischaemia time; y, years; HLA-A, B and DR, human leucocyte antigens A, B and DR

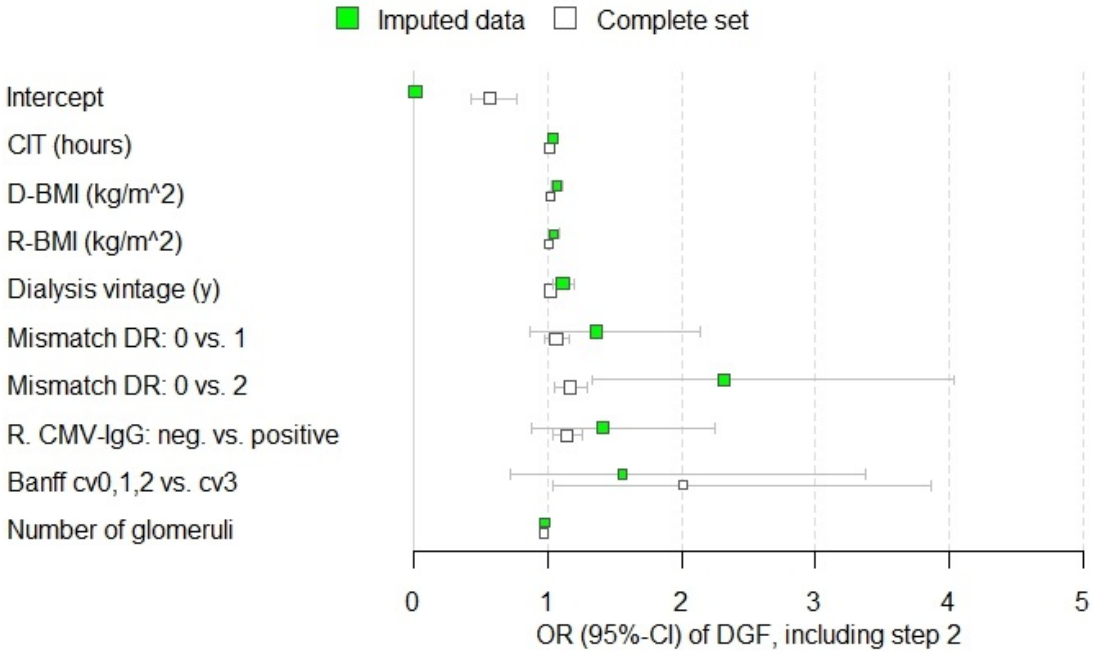


Figure 5.29: Regression estimates of DGF-score including histology on imputed vs. non-imputed dataset. Model trained on subset of observations with DGF-predictions in the range of .18 - .36 which based on the final models after step 1[30].
Abbreviations: CIT, cold ischaemia time; D-BMI, donor BMI; R-BMI, recipient BMI; R. CMV-IgG, recipient cytomegalovirus antibodies IgG

When the histograms of predicted probabilities for DGF are plotted again using the score from step 2, separated according to the actual observed outcome, an improvement in the separation can be seen in the training (Figure 5.30[30]). This was also observed by the net reclassification index, where a significant increase in the NRI for non-events $NRI_{non-event}$ = .504 (95%-CI: .393 to .616) at a simultaneous decrease of the NRI for events NRI_{event}=-.244 (95%-CI: -.454 to -.034) was observed at the subset of recipients with intermediate risk. Overall, a NRI of .260 (95%-CI: .023 - .498) was observed. Measured on the entire dataset, the NRI was less strong pronounced with values of $NRI_{non-event}$ = .319 (95%-CI: .227 - .411), NRI_{event} = -.111 (95%-CI: -.269 to .046) and overall NRI = .208 (95%-CI: .025 - .390). After the second step, the overall AUC and scaled Brier improved slightly from .669 to .695 and .072 to .092 on the training data. As displayed in Figure 5.32[30], calibration intercept and slope were very good (close to 0 and 1), which is expected as the models were trained on the data.

Recalibration and updating did not improve accuracy on the validation dataset, where there were fewer observations in the range between 18% and 36% of the predicted probability. Respectively, the overall NRI on the subset was NRI = .008 (95%-CI: -.659 to .676) and the entire validation NRI = .062 (95%-CI: -.282 to .406). After the second step, the overall AUC and scaled Brier became slightly worse as they decreased from .700 to .692 and .028 to .001 on the validation data. Calibration intercept (-.641 and -.669) and slope (.780 and .620) were rather bad and indicate a severe underestimation of the risk, which can also be observed in Figure 5.32.

For outcome 1y-tl, on the other hand, there was still no clear separation of the predicted probabilities in the histogram (Figure 5.31[30]) on the training after step 2. The AUC after step 2 equalled .707, calibration was, as for DGF, very good (intercept = .005, slope = 1.003, ECI = .047), which can be observed in Figure 5.33[30]. The scaled Brier score, as for other models, was with .041 rather low.

On the validation dataset, the observed transplant losses were significantly more frequently to the right of the vertical line marking the prevalence as compared to the training data. However, there was only a single observation where Banff ct was in ct2 or ct3. Therefore, no significant differences between the prediction after step 1 and step 2 could be expected on the validation which is why the estimation of the NRI is rather useless. The AUC (=.765) was better than on the training data, which was also the case for the scaled Brier score (=.052). Calibration was, as expected, worse (intercept = -.379, slope = 1.452, ECI = .207), with a strong underestimation of the risk below the populations prevalence (9%) and overestimation for higher predicted probabilities.

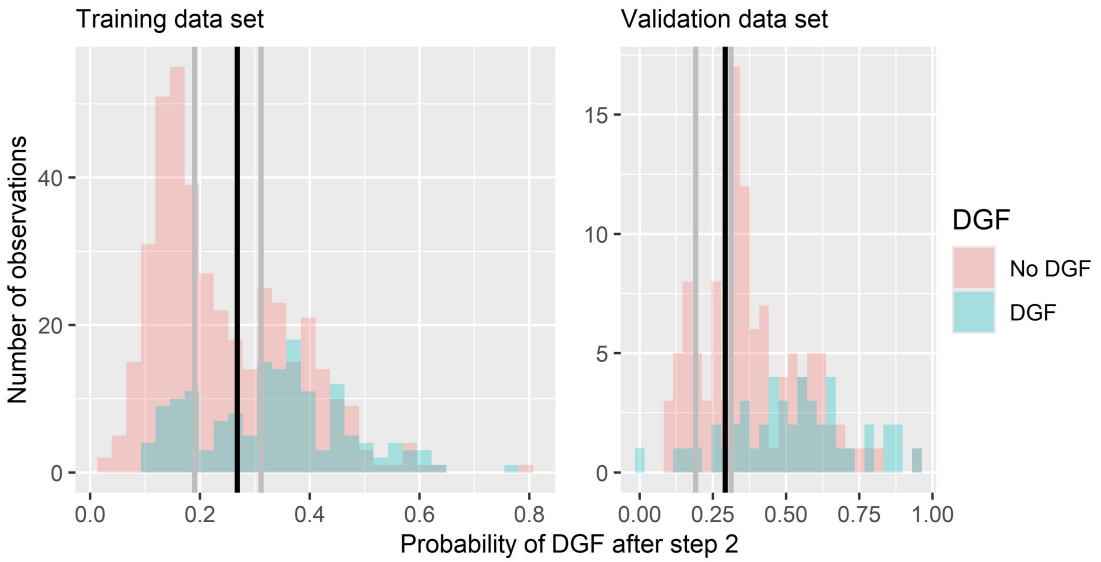


Figure 5.30: Predicted probabilities of DGF after step 2 on training and validation dataset. The vertical, black line represents the DGF prevalence values, grey lines show the interval between .18% and .36%.

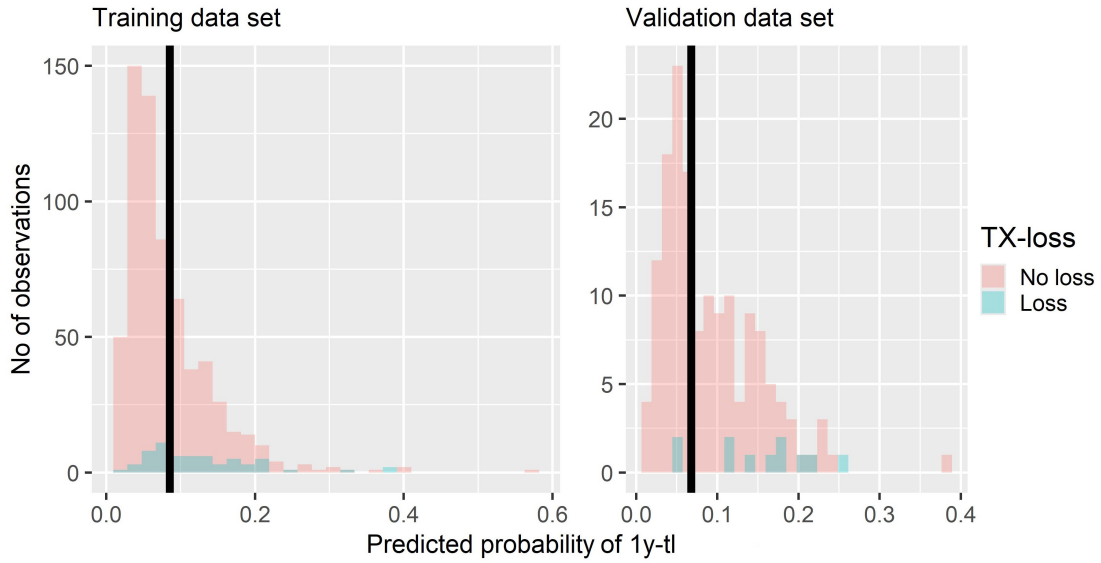


Figure 5.31: Predicted probabilities of 1y-tl after step 2 on training and validation dataset. The vertical, black line represents the prevalence of death-censored transplant losses within the first year after transplantation.

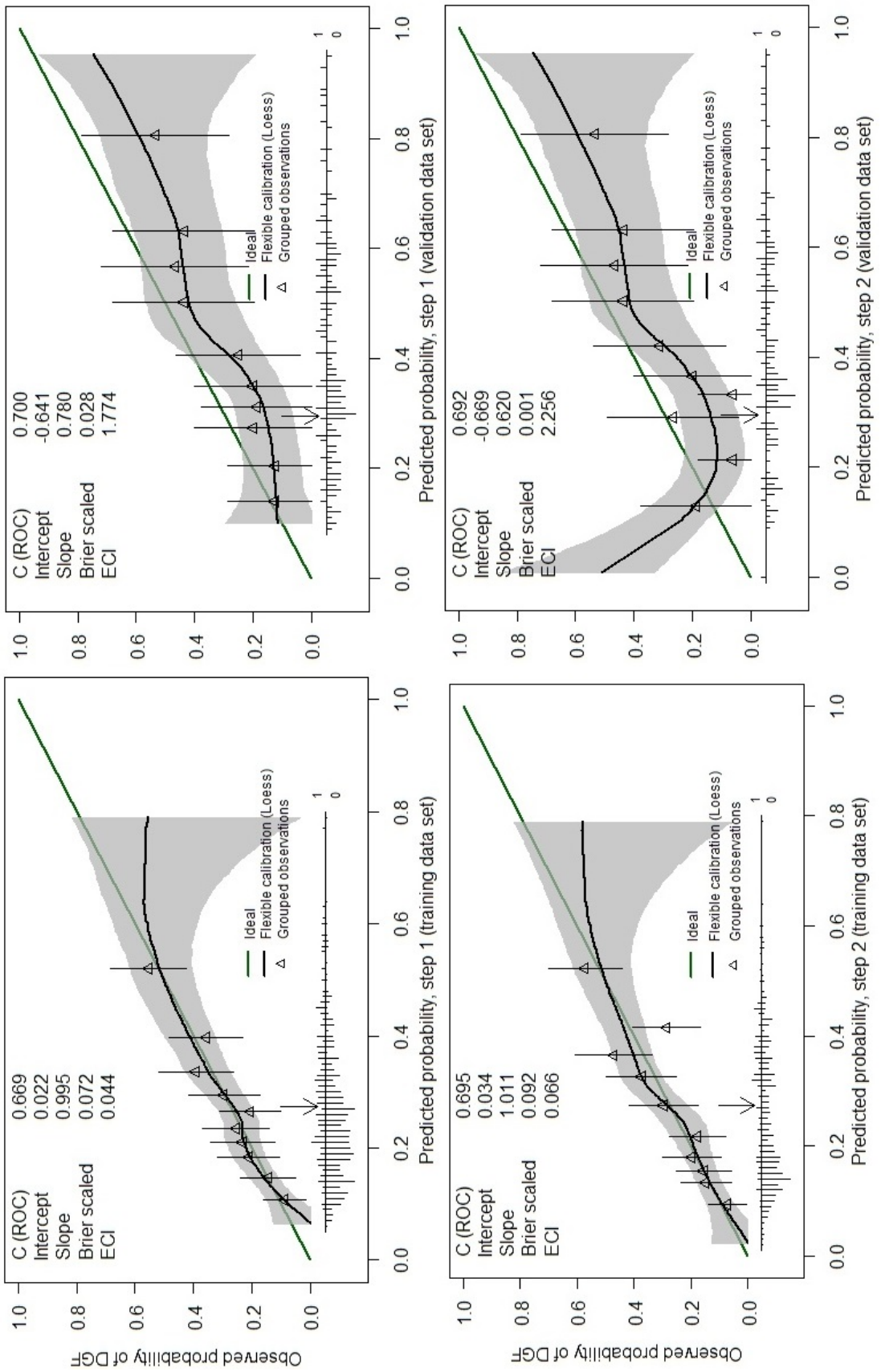


Figure 5.32: Calibration plots and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of DGF after step 1 and 2 on training and validation datasets. [30].

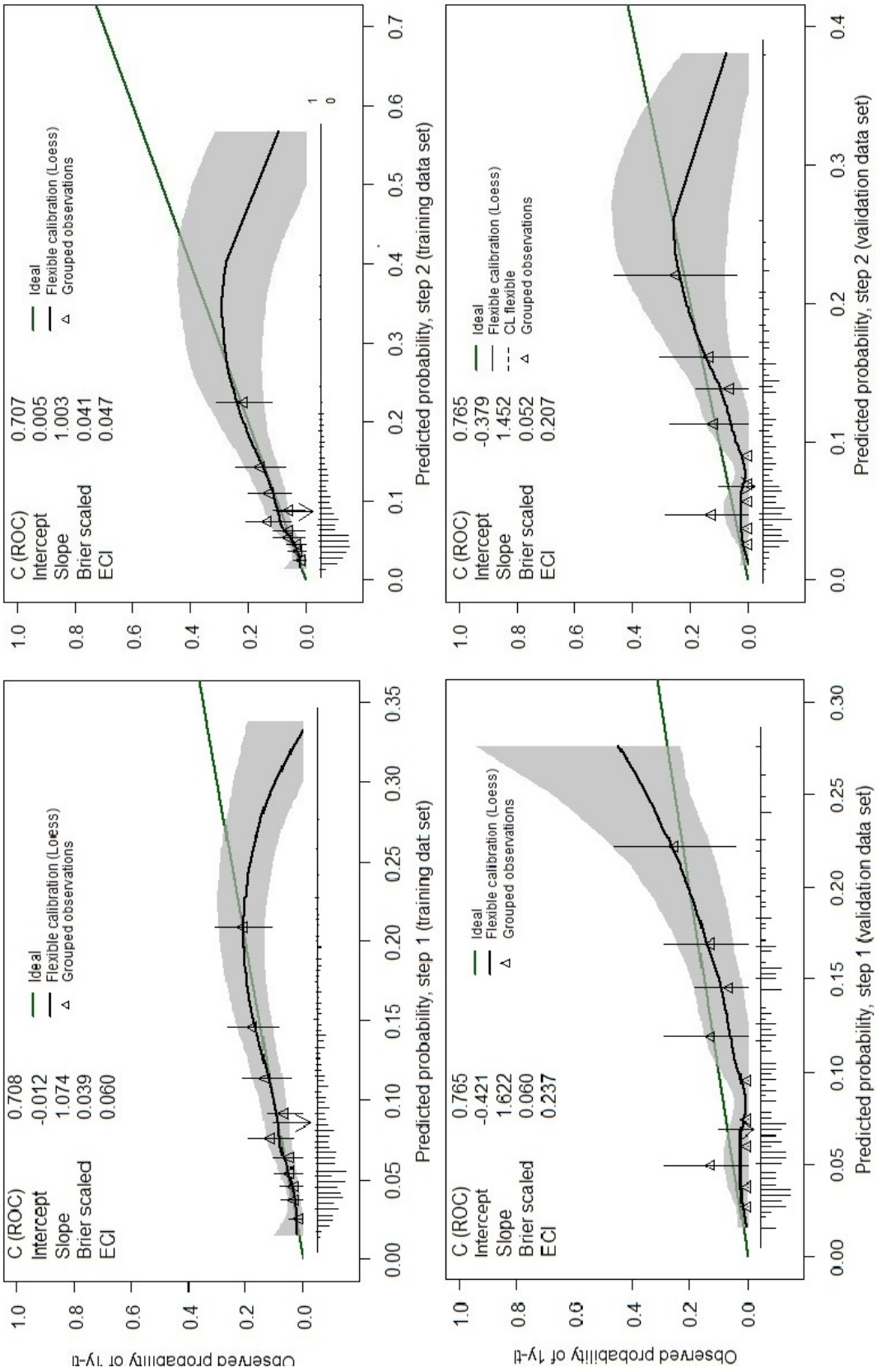


Figure 5.33: Calibration plots and statistics (ECI = estimated calibration index, C(ROC) = AUC, CL = 95% confidence limits) of death-censored transplant loss within one year after step 1 and 2 on training and validation datasets. [30].

5.6 Comparison of performance between established and the new 2-Step scores

As already described in detail in Chapter 2, there are several scores that can be used to predict DGF and 1y-tl. To compare the performance as measured by the AUC of the 2-Step DGF-score, the well-known scores from Irish (KDRI, version 2010 [62]), Balaz [63] and Chapal [64] were used. Balaz presented two scores called "CIV", which bases on Banff ci and Banff cv, and "CIV + donor age + donor cause of death", adding the donors age and cause of death to "CIV". The performance of the developed 2-Step 1y-tl-score was compared with the established ones by Snoeijs [103], Port [79], and De Vusser (Leuven score [102]), as well as the recently on Eurotransplant data generated KTOP-score from Miller [37]. In case of the solely on Banff lesion scores based score by Snoeijs, the "Snoeijs, score", as a sum of cg+mm+ct+cv+ah+(fraction sclerosed glomeruli*3) was chosen rather than the Remuzzi score, with cutoff of ≤ 3 vs. > 3 defining a dichotomised version "Snoeijs, binary". The AUC was chosen, because it is independent of the calibration (with regard to intercept and slope). Therefore, even if the intercept of a regression-based score is not published, the AUC can still be correctly estimated. As the cohorts of the training and validation datasets differed in terms of donor and recipient risk profiles, validation was performed on both datasets. In addition to transplant loss within the first year after transplantation, 3-, 5-, and 10-year transplant survival were also evaluated for the Leuven score and 5-year transplant survival for the scores by Snoeijs and Miller (Figure 5.36[30]). These scores base on Cox-regression models with time-dependent baseline hazard functions instead of a constant intercept term and were created to preliminarily predict long-term outcomes. Three of the scores (Balaz, Snoeijs, De Vusser) also considered zero-hour biopsies, the others were based purely on clinical factors. In all studies, the entire population of recipients of a clinic or registry was always included, with the exclusion of minors. They are therefore not calibrated for any specific risk group.

For the outcome DGF, in addition to the 2-Step score, only Irish's model achieved an AUC value of $> 60\%$ on the training (AUC [95%-CI] = 62.7 [55.2-70.3]%), followed by Chapal (AUC [95%-CI] = 59.6 [54.3 - 64.9]%) and the two scores by Balaz, both of which only achieved AUCs just above 50%. Although none of the models were specifically trained on high-risk cohorts, the AUC's on the validation dataset were higher in all cases, but still below 70% as shown in Figure 5.34[30]. In particular, Balaz's model, which had a AUC of 60.6% when only Banff ci and cv were considered as compared to when donor age and cause of death were also taken into account (AUC = 51.8%), was surprising. On the training dataset, a similar picture emerged for transplant loss within the first year, with KTOP (recipient score AUC [95%-CI] =

65 [57.3-72.7]% vs. recipient and donor score AUC [95%-CI] = 66.2 [59-73.3]%) and Leuven (AUC [95%-CI] = 64.4 [56.3-72.5]%) performing best following the 2-Step score, as shown in Figure 5.35[30]. The AUC of Snoeijs' dichotomized score indicated that the prediction was barely better than a coin toss AUC [95%-CI] = 50.8 [44.4-57.2]%, while the continuous score still had an AUC of 54.6 [46.7-62.4]%. Port's score was comparable to that of Leuven with AUC [95%-CI] of 62.8 [53.5-71.7]%.

Apart from Snoeijs' continuous score, the AUCs improved in all models on the validation dataset, but with a significant widening of the 95% confidence intervals. KTOP again performed best after the 2-Step score with AUC [95%-CI] = 72.6 [60-85.2]% when both donor and recipient variables were considered.

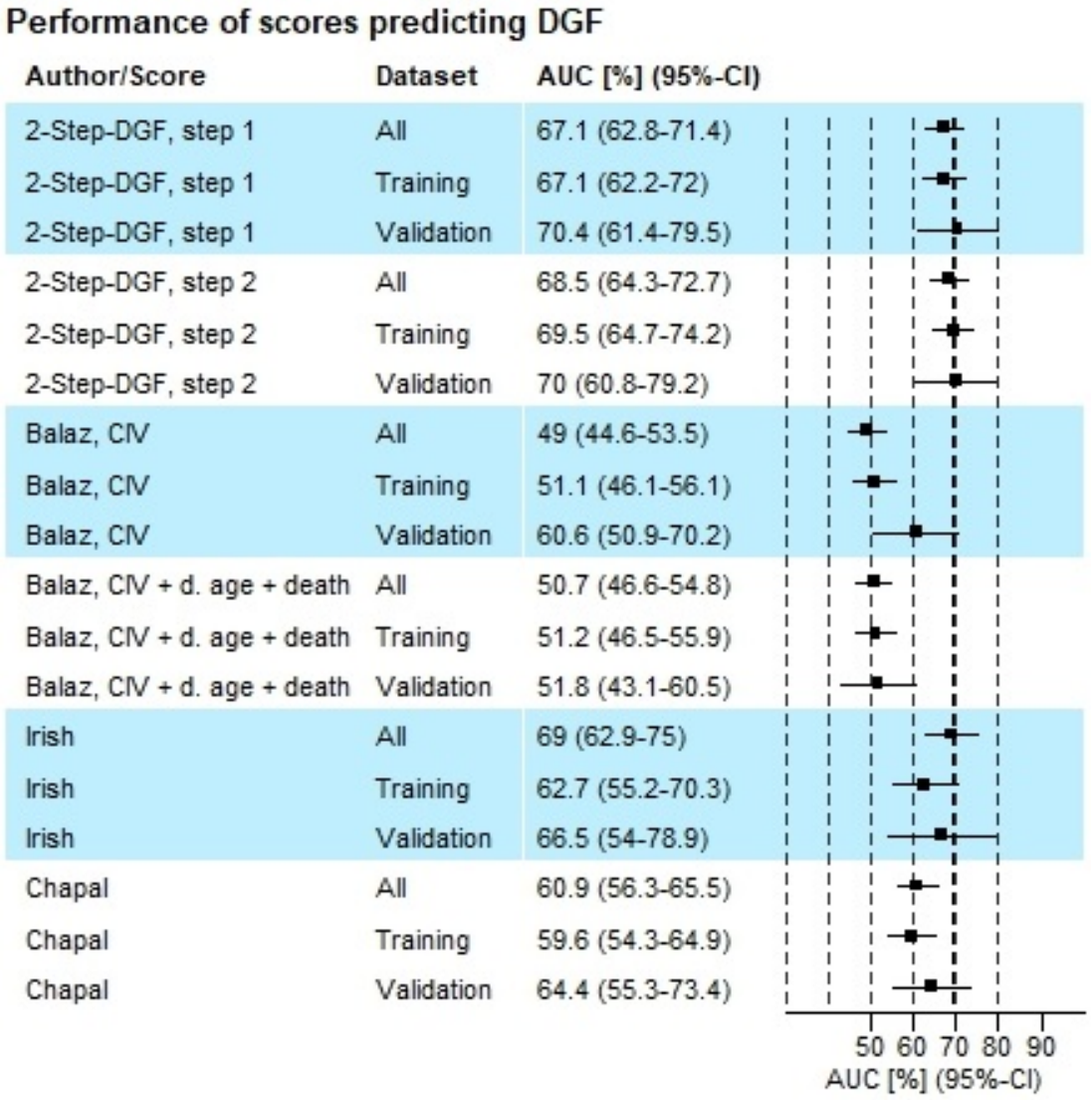


Figure 5.34: AUC values of scores predicting DGF (Balaz, Irish, Chapal, 2-Step)[30]. "Balaz, CIV" bases on Banff ci and Banff cv; "Balaz, CIV + d. age + death" adds the weighted donors age and cause of death to "Balaz, CIV"

The fact that Snoeijs' score, which based purely on Banff lesion scores, performs the worst is hardly surprising, as it has already been shown that the Banff lesion scores in this dataset were univariate only slightly associated with 1y-tl, and continue to do so when sum up. When looking at transplant loss at a later point in time, KTOP continued to perform best for the five years, at least on the training set (AUC = 67.8% and 67.6% without and with recipient information).

On the validation dataset, the prediction by KTOP was comparable to a coin toss in terms of risk discrimination. Snoeijs' score again performed worst (AUC = 51.2%), with a slight improvement on the validation data (AUC = 56.4%). For 5-year transplant loss (5y-tl), De Vusser's Leuven Score performed slightly worse than the KTOP, on both training and validation datasets. Three-year transplant loss (3y-tl) was only determined for the Leuven Score. The AUC for the training data was the same as for 5y-tl (64.5% and 65%, respectively) and for the validation data as for 10y-tl (61.5% and 61.2%, respectively).

In summary, it can be said that scores based purely on clinical parameters showed comparable results with those that also took the Banff lesion scores into account. The Snoeijs score, which was based exclusively on histological parameters, performed worst across the board. The KTOP, which was also trained with Eurotransplant data, showed comparable results to the 2-Step scores, although the training cohort for KTOP was significantly larger.

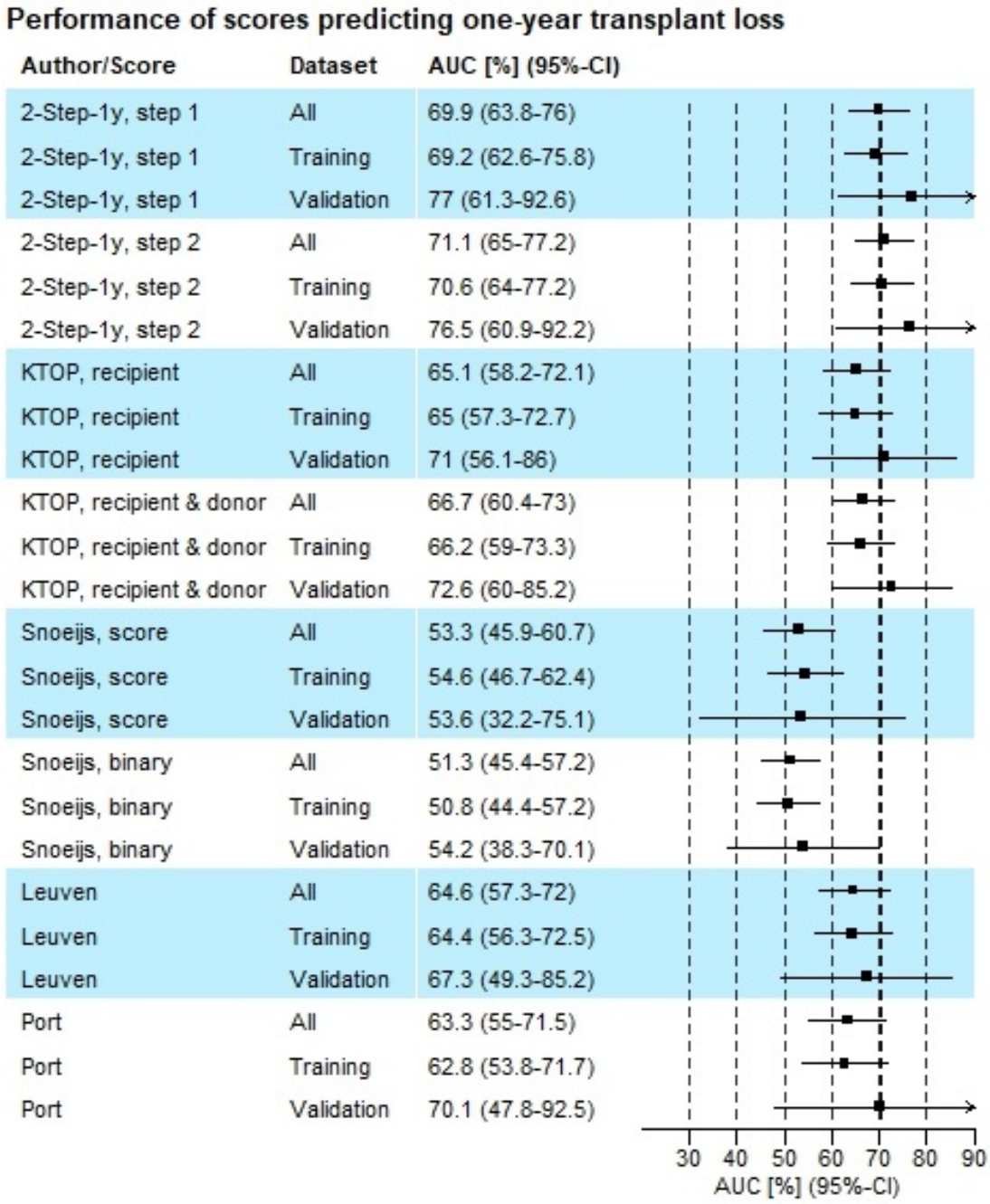


Figure 5.35: AUC values of scores predicting 1y-tl (by Miller "KTOP", Snoeijs, Leuven, Port, 2-Step)[30]. KTOP is based on recipient "KTOP, recipient" or recipient, transplant and donor variables "KTOP, recipient & donor". "Snoeijs, score", is a sum of cg+mm+ct+cv+ah+(fraction sclerosed glomeruli*3), with cutoff ≤ 3 vs. > 3 it defines a dichotomised score "Snoeijs, binary".

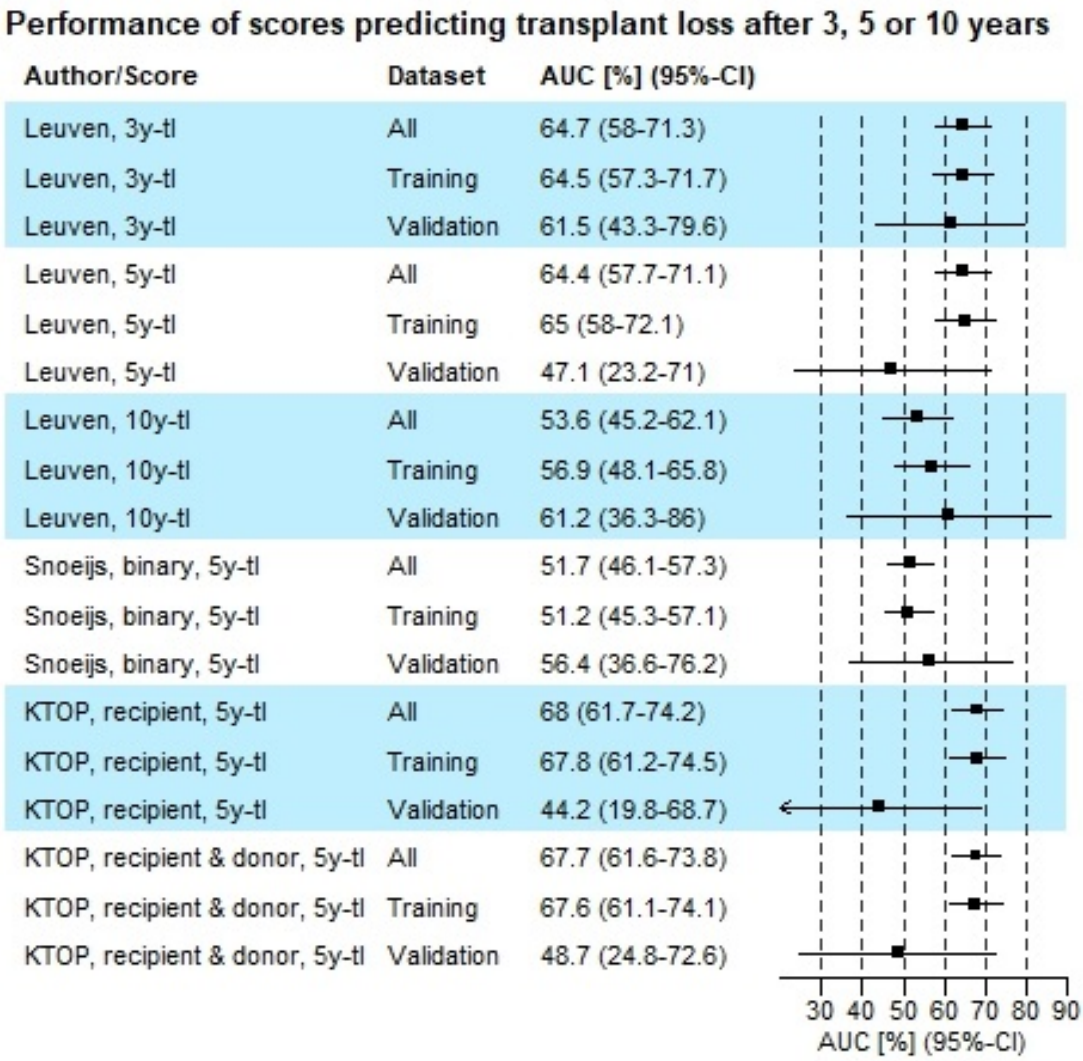


Figure 5.36: AUC values of scores predicting 3-, 5- and 10-year transplant loss (by Leuven, Snoeijs, Miller "KTOP)[30]. KTOP is based on recipient "KTOP, recipient" or recipient, transplant and donor variables "KTOP, recipient & donor". "Snoeijs, binary", is a sum of cg+mm+ct+cv+ah+(fraction sclerosed glomeruli*3), with cutoff ≤ 3 vs. > 3 defining a dichotomised score.

Chapter 6

Discussion and Conclusion

6.1 Summary of findings

Before the actual development of the score began, a systematic literature review of existing scores that predict DGF or one-year (death-censored) graft loss was conducted. As a result of this search, thirty prediction models for the outcome DGF were identified, five of which were based on both clinical and histological factors. None of the 9 scores that predict 1y-tl take into account results from zero-hour or procurement biopsies. However, one-year graft survival could also be predicted with models that take into account a longer observation period after transplantation, provided that appropriate methods, such as Cox regression with the baseline hazard specified at one year, are used in the modelling as with KTOP. Models that consider histological examination as an option have not existed so far. What is new for kidney transplantation is already established and integrated into guidelines in other areas, particularly in oncology and cardiology diagnostics [189, 190, 191].

The available training and validation dataset was then described in terms of the distribution of the possible influencing variables and outcomes. This showed that the donors and recipients in the validation dataset, in terms of their characteristics, were comparatively at higher risk for DGF and the death-censored transplant loss within the first year.

The actual steps in score development were described in Sections 5.3 to 5.5. Finally, a group of recipients at intermediate risk of DGF was identified, for whom an additional histological examination may be useful to predict the outcome. In contrast, an additional histological evaluation to predict later outcomes, such as the 1y-tl, seemed to lead to an improvement in accuracy in every case.

Score building was followed by a validation and a comparison with predictions based on other established scores in Section 5.6. This revealed the negative effects of the different cohorts used for training and validation with regard to the distributions

of the risk factors. In general, the risks for DGF and 1y-tl were underestimated in the validation cohort, which was clearly visible in the calibration curves of the 2-Step scores. AUCs were, however, sometimes higher in the validation cohort than in the training dataset, even for the comparison scores. This clarifies how important it is to consider the distribution of influencing factors in a cohort if you want to apply trained risk scores to new populations. In particular, if you use very heterogeneous cohorts, as is the case within the Eurotransplant region, one should take into account different regional or risk-based clusterings within these cohorts in order to generate results that are more specific and thus more correct.

6.2 Strengths and limitations

One *strength* is certainly that the data used come from several countries of the Eurotransplant region, representing a variety of healthcare systems in which different medical-technical, transplant-political and logistical conditions prevail. In total, transplant centres from six countries were involved. What can be seen as a strength in this project has also shown *limitations*. With a significantly lower number of cases than planned in the application, and an unbalanced distribution among the centres, centre effects in the form of baseline probabilities for the two outcomes could not be taken into account as planned. The number of observations per centre in the training dataset varied between 44 and 442, with DGF rates between 4% and 44% and 1y-tl rates between 3.3% and 9.8%, clearly indicating that the cohorts are not representative and that the baseline estimates are biased. It is even more surprising that inter-centre differences in terms of the relative frequency of outcomes were not taken into account in any risk score to date. When the KTOP score was created, based on more than 32.000 Eurotransplant transplants from eight countries and modelled using Cox regression, at least the predictive accuracy was calculated on a leave-one-country-out basis [37].

As there is a recent but ongoing debate about the ability of nephropathological parameters to predict the transplant outcome, it is a *strength* of this work that these factors were included in the study data in addition to clinical variables. Biopsies were provided by the participating centres and evaluated by an experienced nephropathologist specifically for the purpose of the study. As a result, the histopathological variables were almost complete. The situation was different for variables that did not need to be reported to Eurotransplant for identification and allocation. Variables with $> 40\%$ missing values were directly excluded from the analysis, but even lower proportions, if they occur in many variables, can lead to distorted results and are a source of systematic bias. As a result, missing data can be considered a major *weakness* and *limiting factor* in the modelling, despite the possibility of replacing them using multi-

ple imputation, and the existence of modelling methods, such as random forests, that do not rely on completeness to include them in the analysis. However, analyses on the multiple imputed dataset yielded comparable estimates except for the number of HLA-DR mismatches where the OR for DGF (2 vs. 0 mismatches) was higher in the imputed data (0.23 vs. 0.15). Also, the role of the number of glomeruli, a frequently assessed histological feature, is unclear. The number of glomeruli may instead be a proxy for glomerular density in the cortex and should be evaluated accordingly as it is highly dependent on the type of biopsy used (needle, wedge, punch). The observed negative association between the number of glomeruli and post- transplantation may be biased if transplant centres with a low DGF or 1y-tl rate have preferably used alternatives to needle biopsy. Further research on this issue and on how best to assess glomerular density (in 2-dimensional slides or the 3-dimensional cortex) is needed.

An interesting question which cannot be answered by the study design used, is that of the organs discarded within the Eurotransplant region due to quality concerns, which would still be considered good quality according to the new scores. To collect data on organ quality, Eurotransplant introduced organ quality forms which include information on the quality of the organ, procurement and packaging [92]. According to chapter 9 of the ET Manual, "findings on anatomic abnormalities, possible iatrogenic and packaging and/or transportation related injuries should be indicated" [92]. Eurotransplant has not only committed itself to register the reason for the rejection of a potential transplant, but it also takes the consent of the Eurotransplant duty office to discard an organ from the allocation list, as it could be placed on the rescue allocation list instead.

With access to the information regularly collected by Eurotransplant on all transplanted organs with recipient data, and additional clinical and histopathological information on the quality of the rejected organs, the proportion of rejected organs with the same quality as a comparable, transplanted graft could be quantified. The reliability of the nephropathological assessment according to the Banff criteria was also not recorded, although it was planned to do so. Before a medical device is approved, it is tested for reliability. Similarly, the ability to correctly apply scores to classify pathological abnormalities should be ensured. The Banff scheme of classification was first proposed in 1991 by an international group of pathologists, nephrologists and surgeons and has since then been regularly revised, reviewed and expanded [192]. However, in clinical practice, where decisions about organ quality have to be made on short notice and sometimes by less experienced nephrologists, pathologists and surgeons, the reliability of the Banff classification has been questioned [193, 194, 195]. As a result, the potential impact of pathological abnormalities on transplant outcome could be mitigated by measurement error. In the age of digital pathology and AI-based image analysis, it was also decided to form a separate Banff Digital

Pathology Working Group with the following objectives to "establish a digital pathology repository; develop, validate, and share models for image analysis; and foster collaborations" [196]. With the support of such new tools (a decision based solely on AI is explicitly not advocated here), accuracy could be improved and correlations with transplant outcome could be more precisely determined. Outside this working group, other research groups are working on measuring reliability and how deep learning and semi-supervised learning can support quality measurement and is already underway as a continuation of this project [197].

One medical-technical example in which ET-countries differ are perfusion machines which are not yet widely used in Germany for the storage of transplants prior to transplantation, although the DSO is planning and recommending their increased use. Instead, the kidneys are mainly stored in static cold storage. In the Netherlands, on the other hand, the techniques of normothermic regional perfusion (NRP) and hypothermic machine perfusion (HMP) are widely accepted and have become the standard, even for kidneys at increased risk of failure after transplantation [198]. Several advantages of machine perfusion have been reported in the literature: lower DGF and transplant loss rates [199, 200], additional information on organ quality through real-time monitoring, cost savings [200], immunomodulation and renal repair [201]. However, the logistical requirements of NRP are significantly more demanding than for alternative methods and the protocols used vary widely between transplant centres [202]. Additional costs should be justified by better outcomes after transplantation. However, cost-benefit analyses need to be carried out individually for each region to take account of local health systems and are therefore difficult to estimate. However, a recent Cochrane review showed that compared with static cold storage, hypothermic machine perfusion reduces the rate of DGF and transplant loss, while studies from the USA and Europe have shown that it is cost-saving for kidneys from deceased donors [200]. Despite this existing evidence, information on cold storage versus machine perfusion was not included in the analysis, which is another *limitation*.

The decision on who is accepted as a potential donor is made by the ET countries on an individual basis, taking into account international conventions such as the Declaration of Istanbul on Organ Trafficking and Transplant Tourism [203]. There are also differences between countries in this respect. The cause of death of the donor, i.e. whether it is brain death (DBD donors) or merely circulatory death (DCD donors) according to the Maastricht categories 1 to 4, or even imminent death [204], plays a role. Austria, Belgium and the Netherlands were the first countries to accept kidneys from DCD donors 20 years ago. Germany was one of the 8 European countries, and the only one in Eurotransplant, to accepted mainly organs from DBD donors [205]. A comprehensive discussion of the potential of donation after circular death was already given by Snoeijs et al. in 2007 [206] and was further discussed by Wind et al. in

2013 [207] amongst others, while Potter et al. [204] discuss the potential benefits of imminent death donations. Besides those of young age or with missing biopsies, donors without diagnosis of brain death were excluded in this project. As, according to the Eurotransplant Annual Report of 2022, approximately 50% of deceased donors in the Netherlands and Belgium were DCD, the exclusion of such a number of donors represents a significant *limitation* in the generalisability of the results. Therefore, future studies based on ET data should include both DCD and DBD donors.

Established binary logistic regression was chosen as the modelling method. One of the *strengths* of regression models is that they are easy to interpret and apply to new data, as the result of such a model is an additive scoring system. As shown in chapter 2, most of the prediction models for DGF and 1y-tl are based on regression, which once again emphasises the acceptance of this method in medical statistics when it comes to risk prediction. However, it must also has to be emphasized that in most of the available regression-based scores, the baseline outcome probability also known as "intercept", has not been published, making it impossible to use it them as prediction tool for new data. Regression models are also known to be adaptable in a way that they can be calibrated, updated or fusioned [208], flexible enough to include interactions, nominal and ordinal variables with more than two characteristics, cardinally scaled variables including polynomials and splines. The two-step approach required a statistical or machine learning method that allows recalibration and updating of all or a identified subset of observations for which the accuracy of the clinical- and transplant-based variables was possible. Using logistic regression in the first step and updating of the resulting predictions in a second step by entering histopathological variables resulted in two explainable 2-Step scores, which is a *strength* of the results.

6.3 Informed decision-making

> The whole is greater than the sum of its parts
>
> *Aristotle*
> *384-322 b.C.*

Nephrologists and transplant surgeons who assess the quality of the kidney are certainly at the forefront of the decision to release an organ for donation. Policies, guidelines and centre- or hospital-specific specifications must be taken into account individually to meet the needs of those involved, but experience and confidence in one's own abilities also play an important role.

Ultimately, the recipient must also agree to the transplant after being informed of the potential risks and weighing up the benefits and alternatives. A 2011 survey

of 104 patients and 62 transplant surgeons in the United States compared what clinicians consider relevant information with what patients need to make decisions [209]. The quality of the donor kidney was the most important factor for both groups. Donor age, which is associated with organ quality, and matching difficulties were primarily important to surgeons, while the expected duration of organ preservation after transplantation and similarity to the donor were important to recipients in terms of matching criteria.

Informed decision making describes the process of making decisions based on an understanding of the available information. It involves gathering, analysing and interpreting data to assess potential risks and benefits, alternatives and uncertainties. Figure 6.1 shows some of the pieces of the puzzle that go into the decisions of different stakeholders.

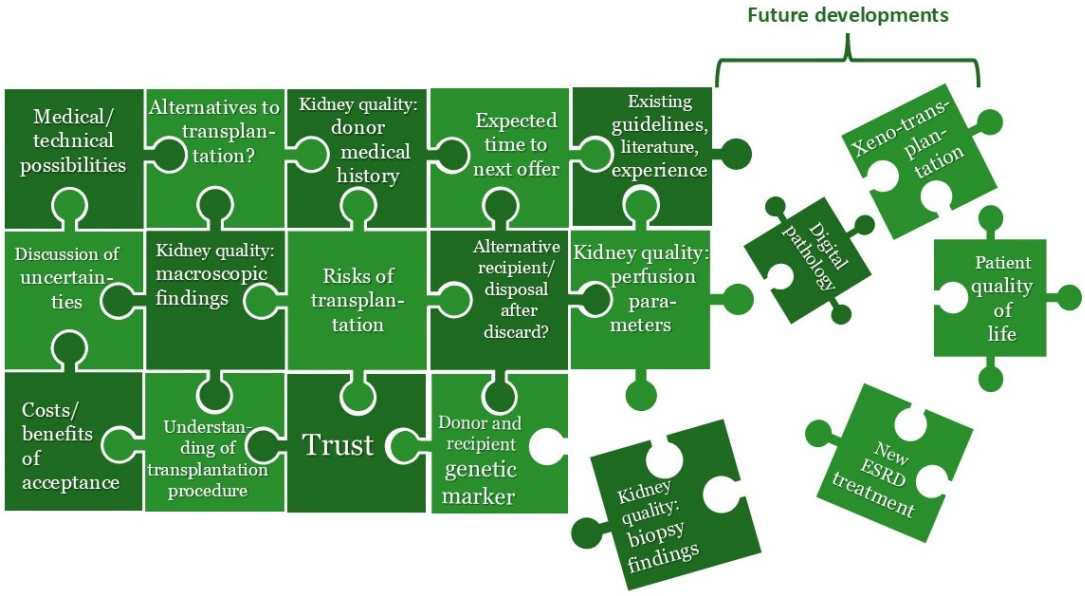


Figure 6.1: Informed decision making: puzzle of stakeholders needs

The system-of-systems approach, which is already widely used in the industry and the social sciences and has been proposed by Dzieran et al. for transplant acceptance, is an approach that is designed to take into account precisely this multidimensionality of decision making [210, 211]. To paraphrase Aristotle's much-quoted sentence "the whole is greater than the sum of its parts". Consequently, the experience of the transplant surgeon, established risk scores (including predicted time to next offer, probability of death before next offer, short and long term transplant outcomes), the literature, clinical and nephropathological factors, medical and technical options for transplant preservation, should all be considered and weighted at the same time and a decision made on this basis.

With regard to the importance of guidelines, a survey of 19 nephrologists in

Australia has shown that confidence in the CARI (Caring for Australasians With Renal Impairment) guideline applied there is quite high [212]. CARI covers the areas of "chronic kidney disease", "dialysis" and "transplantation", with transplantation also including donor suitability [213]. Evidence-based guidelines were generally found to be helpful there, also for decision-making. Another survey amongst clinicians in the USA also indicated the wish for standardization and quality control in how biopsies are performed and findings are interpreted [214].

There is no consensus and little evidence, particularly in the Eurotransplant region, on whether or when to perform renal biopsies and what weight they should have in decision-making. Although there is an ESOT guideline on how to perform a biopsy and what parameters should be taken into account, there are still no guidelines from Eurotransplant or the ESOT on when to do it [39]. However, in chapters 7-10 of the 8th edition of the "Guide To The Quality And Safety Of Organs For Transplantation", the European Directorate for the Quality of Medicines and HealthCare describes kidney-specific evaluation and selection criteria including kidney biopsies, which are not recommended to be performed systematically [215]. In the case of macroscopic findings indicating marginal or bad quality (e.g. tumour, space-occupying lesions), additional imaging and biopsy may be recommended. If a procurement biopsy is performed, it is recommended to follow the Banff classification. With regard to graft acceptance, grafts with low Banff lesion scores (< 2), which are considered to be mild histopathological changes, may be accepted when harvested from older donors or donors with cardiovascular risk factors [215].

The fact that the reliability of the histopathological evaluation of a kidney biopsy is generally considered to be rather low is often criticised in the literature and has been demonstrated several times [216]. In the USA, the proportion of kidneys rejected on the basis of histopathological examination was around 20% in 2019 [217]. Accordingly, there has been and continues to be a critical discussion about the weighting in the decision-making process. Although a survey including 41 kidney transplant surgeons and 27 transplant nephrologists in the USA suggests that good biopsy results increase the acceptance of kidneys with acute kidney injury or of low serum creatinine donor offers, the same study showed that overall more kidneys were discarded than accepted when biopsy findings were available [217]. Another survey of 40% of US kidney transplant centres on deceased donor procurement practices by Lentine et al., 2022, also found that favourable biopsy results could increase acceptance of "marginal" kidneys while increasing the rejection of standard criteria kidneys [218].

6.4 Steps for technical implementation

Ultimately, this project aimed to develop a tool that can be directly integrated into the Eurotransplant database system. As all factors used in the score development are already available in a standardised form in the Eurotransplant database system, and the 2-Step scores represent simple additive models, such an integration would be technically feasible. The implementation of the 2-Step scores should not only provide an evidence-based tool for estimating post-transplant risks, but also reduce the number of rejected organs. As a result, it can be used by all to make an evidence-based decision to accept or reject a transplant. To return to the CRISP-DM data mining process from Section 3.1, the first five steps of creating a prognostic model have already been covered in this project. The model has already been validated internally and presented at conferences (see supplemental material) and in a publication [30]. Validation has shown that the model has weaknesses in calibration when applied to a systematically different cohort. Accordingly, it should be considered whether a new modeling with the first five steps should be carried out for specific subgroups before the current 2-Step scores are included in impact studies.

Implementation would then only be successful if

1 Impact studies were successful with regard to

- the cost-effectiveness and clinical effectiveness
- the rate of rejected or discarded organs that should get reduced and
- the post-transplant risk that should not increase
- the population for which the score is best suited for prediction has been identified as no score fits all (e.g. children, sensitised or older patients)

2 The scores are considered and accepted as a decision-making tool by those who are ultimately involved in the transplantation process and

3 Scores are updateable, e.g. by Bayes theorem, to account for changes in the population and to include new marker.

The KDPI, the KDRI and the Estimated Post Transplant Survival (EPTS) calculator have already been incorporated into UNOS and are regularly updated, to take into account new observations. The first two are even used as part of the OPTN allocation system. A comparison of the by the 2-Step scores predicted risks for DGF and 1y-tl of transplanted and rejected or discarded organs is pending but is already being planned. If the proportion of rejected organs with a low calculated risk is significantly higher than for comparable, successfully transplanted kidneys, this might

be an indication that the number of transplantable organs can be further increased. If modelling, validation and implementation were successful, one could identify and close potential donor detection gap in clinics.

6.5 Ethical considerations

It is widely accepted that there is a chronic shortage of donor organs worldwide, but also in the Eurotransplant region. Accordingly, the fair distribution of the few donated organs is essential, as well as the maximum utilization concerning a possible rejection. Distribution in the Eurotransplant region is regulated by the allocation system there, as described in Section 2.4. This already takes into account the subgroup of older patients who have a lower chance of receiving a donor organ due to increased risks after transplantation. Donor and recipient age are recognized risk factors that often appear in risk scores. If high-risk results in systematically not receiving an organ, despite being at the top of the transplant list, this would be a form of discrimination. In the USA, for example, the Organ Procurement and Transplantation Network (OPTN) is proposing to refit the KDRI, which is an integral part of UNOS, without inclusion to the race factor [219]. "Race" is likely to be a proxy for a genetic variation (APOL1) that is more common in black Americans and is associated with poorer transplant outcomes. A high KDRI is a reason for refusal of transplantation in the USA, which puts this population group at a disadvantage. A comparison between American and French rates of organ acceptance has shown that the proportion of rejected kidneys with a higher risk according to KDRI was significantly higher in the USA than in France, which indicates a low risk appetite, but may also indicate a limited predictive accuracy of the KDRI if the rates of poor transplant outcomes are not lower at the same time [220]. An example about why transparency in how a score was developed is important was given by the UK liver transplantation allocation system in 2022 which uses the Transplant Benefit Score (TBS) (`https://transplantbenefit.org`). Researchers found "implausible predictions that simulated patients with chronic liver disease survive longer if they develop cancer. In so doing, the algorithm actively deprioritized simulated patients with cancer" [221]. What counts here is not only the scientific basis, the quality of the data, and the methods used to develop the score, but also the transparent communication of the weaknesses of the score and the weighting of the results in the decision-making process. The weaknesses and potential strengths of the 2-Step scores were discussed in Section 6.2.

When it comes to the question of who receives an organ, sicker patients are clearly at the top of the waiting list. However, this system could also be exploited and is also subject to errors. A transplant scandal that came to light in Germany in 2012 exploited precisely this type of waiting list prioritisation. It was found that patients

had been classified as sicker at several transplant clinics in order to move them higher up the list. In the US, however, "A racially biased test kept thousands of Black people from getting a kidney transplant" [222] because they appeared to be healthier than they actually were.

Eurotransplant is already working to reduce ethical concerns about fairness and possible discrimination and to maximize the use of the donor organ pool, not only by offering special programs for patients with a low chance of receiving a transplant, but also by implementing a rescue allocation system whereby organs that are considered by a centre to be unsuitable for their recipient or of insufficient quality are given a second or even third chance of being transplanted. The 2-Step score is intended to support these efforts by re-evaluating the quality of organs that would be rejected on the basis of purely clinical, macroscopic observations, or that fall into a medium to high risk group with an unclear outcome, using additional histological markers. Ultimately, it can also be considered unethical to completely ignore the findings of nephropathology if this could save more organs. To quote Marie-Francois Xavier Bichat (1771-1802) at this point:

> 'The more one observes diseases (...), the more one is convinced of the necessity to consider local diseases not from the perspective of complex organs, but from the perspective of individual tissues' [223].

6.6 Conclusion

Finally, despite a limited observational cohort, it was shown that, taking into account clinical parameters, the additional consideration of histological parameters can improve the accuracy of prediction of both DGF and 1y-tl. Furthermore, a knowledge gap in statistics and machine learning methods was revealed, which needs to be addressed by further research. Although the observed effects of additional nephropathology were relatively weak, this project provided an impetus for further follow-up projects with large, possible multinational, cohorts. It was also possible to illustrate how important the impetus supplied by medical issues is for the further development of mathematical methods. Nevertheless, the potential ethical, economic, stakeholder, and transplant surgeon concerns of new methods must not be ignored, and the consequences of implementing such two-stage models in clinical decision-making should be investigated accordingly.

Bibliography

[1] ERA Registry. "ERA Registry Annual Report 2021". Amsterdam UMC, location AMC, Department of Medical Informatics, Amsterdam, the Netherlands. 2023. URL: https://www.era-online.org/wp-content/uploads/2023/08/ERA-Registry-Annual-Report-2021.pdf (visited on 09/06/2023).

[2] Dennis Häckl, Nils Kossack, and Tonio Schoenfelder. "Prävalenz, Kosten der Versorgung und Formen des dialysepflichtigen chronischen Nierenversagens in Deutschland: Vergleich der Dialyseversorgung innerhalb und außerhalb stationärer Pflegeeinrichtungen". In: *Das Gesundheitswesen* 83.10 (Oct. 2021), pp. 818–828. ISSN: 0941-3790, 1439-4421. DOI: 10.1055/a-1330-7152. URL: http://www.thieme-connect.de/DOI/DOI?10.1055/a-1330-7152 (visited on 09/06/2023).

[3] A. Rajkomar, J. Dean, and I. Kohane. "Machine Learning in Medicine". In: *N Engl J Med* 380.14 (2019), pp. 1347–1358. ISSN: 1533-4406 (Electronic) 0028-4793 (Linking). DOI: 10.1056/NEJMra1814259. URL: https://www.ncbi.nlm.nih.gov/pubmed/30943338.

[4] M. Tonelli et al. "Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes". In: *Am J Transplant* 11.10 (2011), pp. 2093–109. ISSN: 1600-6143 (Electronic) 1600-6135 (Linking). DOI: 10.1111/j.1600-6143.2011.03686.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/21883901.

[5] Eurotransplant. Web Page. URL: https://www.eurotransplant.org.

[6] Eurotransplant International Foundation. *Annual Report 2021*. Report. 2022. URL: https://www.eurotransplant.org/wp-content/uploads/2022/06/Annual-Report-2021_LR.pdf.

[7] D. Mudiayi et al. "Global Estimates of Capacity for Kidney Transplantation in World Countries and Regions". In: *Transplantation* 106.6 (2022), pp. 1113–1122. ISSN: 1534-6080 (Electronic) 0041-1337 (Print) 0041-1337 (Linking). DOI: 10.1097/TP.0000000000003943. URL: https://www.ncbi.nlm.nih.gov/pubmed/34495014.

[8] K. L. King, S. A. Husain, and S. Mohan. "Geographic Variation in the Availability of Deceased Donor Kidneys per Wait-Listed Candidate in the United States". In: *Kidney Int Rep* 4.11 (2019), pp. 1630–1633. ISSN: 2468-0249 (Electronic) 2468-0249 (Linking). DOI: 10.1016/j.ekir.2019.08.018. URL: https://www.ncbi.nlm.nih.gov/pubmed/31891004.

[9] D. C. Cron, S. A. Husain, and J. T. Adler. "The New Distance-Based Kidney Allocation System: Implications for Patients, Transplant Centers, and Organ Procurement Organizations". In: *Curr Transplant Rep* 9.4 (2022), pp. 302–307. ISSN: 2196-3029 (Print) 2196-3029 (Electronic). DOI: 10.1007/s40472-022-00384-z. URL: https://www.ncbi.nlm.nih.gov/pubmed/36254174.

[10] M. Y. Yeung, P. T. Coates, and P. K. Li. "Kidney Organ Allocation System: How to Be Fair". In: *Semin Nephrol* 42.4 (2022), p. 151274. ISSN: 1558-4488 (Electronic) 0270-9295 (Linking). DOI: 10.1016/j.semnephrol.2022.09.002. URL: https://www.ncbi.nlm.nih.gov/pubmed/36566139.

[11] C. Ponticelli, F. Reggiani, and G. Moroni. "Delayed Graft Function in Kidney Transplant: Risk Factors, Consequences and Prevention Strategies". In: *J Pers Med* 12.10 (2022). ISSN: 2075-4426 (Print) 2075-4426 (Electronic) 2075-4426 (Linking). DOI: 10.3390/jpm12101557. URL: https://www.ncbi.nlm.nih.gov/pubmed/36294695.

[12] G. J. Alangaden et al. "Infectious complications after kidney transplantation: current epidemiology and associated risk factors". In: *Clin Transplant* 20.4 (2006), pp. 401–9. ISSN: 0902-0063 (Print) 0902-0063 (Linking). DOI: 10.1111/j.1399-0012.2006.00519.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/16842513.

[13] D. W. Kim et al. "Financial impact of delayed graft function in kidney transplantation". In: *Clin Transplant* 34.10 (2020), e14022. ISSN: 1399-0012 (Electronic) 0902-0063 (Print) 0902-0063 (Linking). DOI: 10.1111/ctr.14022. URL: https://www.ncbi.nlm.nih.gov/pubmed/32573812.

[14] M. T. Li et al. "Effects of Delayed Graft Function on Transplant Outcomes: A Meta-analysis". In: *Transplant Direct* 9.2 (2023), e1433. ISSN: 2373-8731 (Print) 2373-8731 (Electronic) 2373-8731 (Linking). DOI: 10.1097/TXD.0000000000001433. URL: https://www.ncbi.nlm.nih.gov/pubmed/36700066.

[15] Ranjit Narayanan et al. "Delayed graft function and the risk of death with graft function in living donor kidney transplant recipients". In: *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 56.5

(Nov. 2010), pp. 961–970. ISSN: 1523-6838. DOI: 10.1053/j.ajkd.2010.06.024.

[16] Shruti N. Tapiawala et al. "Delayed graft function and the risk for death with a functioning graft". In: *Journal of the American Society of Nephrology: JASN* 21.1 (Jan. 2010), pp. 153–161. ISSN: 1533-3450. DOI: 10.1681/ASN.2009040412.

[17] W. D. Irish et al. "A risk prediction model for delayed graft function in the current era of deceased donor renal transplantation". In: *Am J Transplant* 10.10 (2010), pp. 2279–86. ISSN: 1600-6143 (Electronic) 1600-6135 (Linking). DOI: 10.1111/j.1600-6143.2010.03179.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/20883559.

[18] J. D. Schold et al. "The broad spectrum of quality in deceased donor kidneys". In: *Am J Transplant* 5.4 Pt 1 (2005), pp. 757–65. ISSN: 1600-6135 (Print) 1600-6135 (Linking). DOI: 10.1111/j.1600-6143.2005.00770.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/15760399.

[19] C. Jeldres et al. "Prediction of delayed graft function after renal transplantation". In: *Can Urol Assoc J* 3.5 (2009), pp. 377–82. ISSN: 1911-6470 (Print) 1911-6470 (Linking). DOI: 10.5489/cuaj.1147. URL: https://www.ncbi.nlm.nih.gov/pubmed/19829730.

[20] P. Balaz et al. "Identification of expanded-criteria donor kidney grafts at lower risk of delayed graft function". In: *Transplantation* 96.7 (2013), pp. 633–8. ISSN: 1534-6080 (Electronic) 0041-1337 (Linking). DOI: 10.1097/TP.0b013e31829d9225. URL: https://www.ncbi.nlm.nih.gov/pubmed/23912171.

[21] S. Kawakita et al. "Personalized prediction of delayed graft function for recipients of deceased donor kidney transplants with machine learning". In: *Sci Rep* 10.1 (2020), p. 18409. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-020-75473-z. URL: https://www.ncbi.nlm.nih.gov/pubmed/33110142%20https://www.nature.com/articles/s41598-020-75473-z.pdf.

[22] M. Chapal et al. "A useful scoring system for the prediction and management of delayed graft function following kidney transplantation from cadaveric donors". In: *Kidney Int* 86.6 (2014), pp. 1130–9. ISSN: 1523-1755 (Electronic) 0085-2538 (Linking). DOI: 10.1038/ki.2014.188. URL: https://www.ncbi.nlm.nih.gov/pubmed/24897036.

[23] G. Zaza et al. "Predictive model for delayed graft function based on easily available pre-renal transplant variables". In: *Intern Emerg Med* 10.2 (2015), pp. 135–41. ISSN: 1970-9366 (Electronic) 1828-0447 (Linking). DOI: 10.1007/s11739-014-1119-y. URL: https://www.ncbi.nlm.nih.gov/pubmed/25164408.

[24] J. Moore et al. "Predicting early renal allograft function using clinical variables". In: *Nephrology Dialysis Transplantation* 22.9 (June 5, 2007), pp. 2669–2677. ISSN: 0931-0509, 1460-2385. DOI: 10.1093/ndt/gfm249. URL: https://academic.oup.com/ndt/article-lookup/doi/10.1093/ndt/gfm249 (visited on 09/22/2023).

[25] C. G. Ding et al. "Predictive Score Model for Delayed Graft Function Based on Hypothermic Machine Perfusion Variables in Kidney Transplantation". In: *Chin Med J (Engl)* 131.22 (2018), pp. 2651–2657. ISSN: 2542-5641 (Electronic) 0366-6999 (Print) 0366-6999 (Linking). DOI: 10.4103/0366-6999.245278. URL: https://www.ncbi.nlm.nih.gov/pubmed/30425191.

[26] R. Chen et al. "Predictors and one-year outcomes of patients with delayed graft function after deceased donor kidney transplantation". In: *BMC Nephrol* 21.1 (2020), p. 526. ISSN: 1471-2369 (Electronic) 1471-2369 (Linking). DOI: 10.1186/s12882-020-02181-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/33276737.

[27] W. Xue et al. "A prediction model of delayed graft function in deceased donor for renal transplant: a multi-center study from China". In: *Ren Fail* 43.1 (2021), pp. 520–529. ISSN: 1525-6049 (Electronic) 0886-022X (Print) 0886-022X (Linking). DOI: 10.1080/0886022X.2021.1895838. URL: https://www.ncbi.nlm.nih.gov/pubmed/33719820.

[28] Jiali Wang et al. "Combining Clinical Parameters and Acute Tubular Injury Grading Is Superior in Predicting the Prognosis of Deceased-Donor Kidney Transplantation: A 7-Year Observational Study". In: *Frontiers in Immunology* 13 (June 30, 2022), p. 912749. ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.912749. URL: https://www.frontiersin.org/articles/10.3389/fimmu.2022.912749/full (visited on 08/17/2024).

[29] S. Zhao et al. "Prediction model of delayed graft function based on clinical characteristics combined with serum IL-2 levels". In: *BMC Nephrol* 23.1 (2022), p. 284. ISSN: 1471-2369 (Electronic) 1471-2369 (Linking). DOI: 10.1186/s12882-022-02908-2. URL: https://www.ncbi.nlm.nih.gov/pubmed/35971094.

[30] Angela Ernst et al. "2-Step-Scores with optional nephropathology for the prediction of adverse outcomes for brain-dead donor kidneys in Eurotransplant". In: *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association* (Apr. 17, 2024), gfae093. ISSN: 1460-2385. DOI: 10.1093/ndt/gfae093.

[31] Alexander Decruyenaere et al. "Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods". In: *BMC Medical Informatics and Decision Making* 15.1 (Dec. 2015), p. 83. ISSN: 1472-6947. DOI: 10.1186/s12911-015-0206-y. URL: http://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-015-0206-y (visited on 08/17/2024).

[32] Alexander Decruyenaere et al. "Validation in a Single-Center Cohort of Existing Predictive Models for Delayed Graft Function After Kidney Transplantation". In: *Annals of Transplantation* 20 (2015), pp. 544–552. ISSN: 2329-0358. DOI: 10.12659/AOT.894034. URL: http://www.annalsoftransplantation.com/abstract/index/idArt/894034 (visited on 08/17/2024).

[33] Martin Chaumont et al. "Delayed Graft Function in Kidney Transplants: Time Evolution, Role of Acute Rejection, Risk Factors, and Impact on Patient and Graft Outcome". In: *Journal of Transplantation* 2015 (2015), pp. 1–9. ISSN: 2090-0007, 2090-0015. DOI: 10.1155/2015/163757. URL: http://www.hindawi.com/journals/jtrans/2015/163757/ (visited on 08/17/2024).

[34] Olivier H.J. Koning et al. "RISK FACTORS FOR DELAYED GRAFT FUNCTION IN CADAVERIC KIDNEY TRANSPLANTATION: A Prospective Study of Renal Function and Graft Survival after Preservation with University of Wisconsin Solution in Multi-Organ Donors". In: *Transplantation* 63.11 (June 1997), pp. 1620–1628. ISSN: 0041-1337. DOI: 10.1097/00007890-199706150-00015. URL: http://journals.lww.com/00007890-199706150-00015 (visited on 08/18/2024).

[35] F. K. Port et al. "Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors". In: *Transplantation* 74.9 (2002), pp. 1281–6. ISSN: 0041-1337 (Print) 0041-1337 (Linking). DOI: 10.1097/00007890-200211150-00014. URL: https://www.ncbi.nlm.nih.gov/pubmed/12451266.

[36] P. S. Rao et al. "A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index". In: *Transplantation* 88.2 (2009), pp. 231–6. ISSN: 1534-6080 (Electronic) 0041-1337 (Linking). DOI: 10.1097/TP.0b013e3181ac620b. URL: https://www.ncbi.nlm.nih.gov/pubmed/19623019.

[37] Gregor Miller et al. "Kidney Transplantation Outcome Predictions (KTOP): A Risk Prediction Tool for Kidney Transplants from Brain-dead Deceased Donors Based on a Large European Cohort". In: *European Urology* 83.2 (Feb. 2023), pp. 173–179. ISSN: 03022838. DOI: 10.1016/j.eururo.2021.12.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S0302283821022168 (visited on 08/19/2024).

[38] Maarten Naesens. "Zero-Time Renal Transplant Biopsies: A Comprehensive Review". In: *Transplantation* 100.7 (July 2016), pp. 1425–1439. ISSN: 0041-1337. DOI: 10.1097/TP.0000000000001018. URL: https://journals.lww.com/00007890-201607000-00012 (visited on 09/24/2023).

[39] Gianluigi Zaza et al. "European Society for Organ Transplantation (ESOT)-TLJ 3.0 Consensus on Histopathological Analysis of Pre-Implantation Donor Kidney Biopsy: Redefining the Role in the Process of Graft Assessment". In: *Transplant International* 36 (July 4, 2023), p. 11410. ISSN: 1432-2277. DOI: 10.3389/ti.2023.11410. URL: https://www.frontierspartnerships.org/articles/10.3389/ti.2023.11410/full (visited on 08/25/2024).

[40] Dustin Carpenter et al. "Procurement Biopsies in the Evaluation of Deceased Donor Kidneys". In: *Clinical Journal of the American Society of Nephrology* 13.12 (Dec. 2018), pp. 1876–1885. ISSN: 1555-9041, 1555-905X. DOI: 10.2215/CJN.04150418. URL: https://journals.lww.com/01277230-201812000-00017 (visited on 09/24/2023).

[41] S. Nasic et al. "Histological diagnosis from kidney transplant biopsy can contribute to prediction of graft survival". In: *Nephrology (Carlton)* 27.6 (2022), pp. 528–536. ISSN: 1440-1797 (Electronic) 1320-5358 (Print) 1320-5358 (Linking). DOI: 10.1111/nep.14028. URL: https://www.ncbi.nlm.nih.gov/pubmed/35150598.

[42] Krista L. Lentine, Bertram Kasiske, and David A. Axelrod. "Procurement Biopsies in Kidney Transplantation: More Information May Not Lead to Better Decisions". In: *Journal of the American Society of Nephrology* 32.8 (Aug. 2021), pp. 1835–1837. ISSN: 1046-6673, 1533-3450. DOI: 10.1681/ASN.2021030403. URL: https://journals.lww.com/10.1681/ASN.2021030403 (visited on 09/24/2023).

[43] M. Antonieta Azancot et al. "The reproducibility and predictive value on outcome of renal biopsies from expanded criteria donors". In: *Kidney International* 85.5 (May 2014), pp. 1161–1168. ISSN: 1523-1755. DOI: 10.1038/ki.2013.461.

[44] A. Ernst and J. Becker. "Clinicopathological scores with optional nephropathology for the prognostication of outcome for deceased donor kidneys in Eurotransplant". In: *32. Jahrestagung der Deutschen Transplantationsgesellschaft, Jena, Germany*. 26.10. - 28.10.2023.

[45] Chepman P. Clinton J. Kerber R. Khabaza T. Reinartz T. Shearer C. Wirth R. *CRISP-DM 1.0*. https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws201213/kdd/files/CRISPWP-0800.pdf. CRISP-DM consortium: NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringen en Bank Groep B.V (The Netherlands. 1999.

[46] Pengfei Wei, Zhenzhou Lu, and Jingwen Song. "Variable importance analysis: A comprehensive review". In: *Reliability Engineering & System Safety* 142 (Oct. 2015), pp. 399–432. ISSN: 09518320. DOI: 10.1016/j.ress.2015.05.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S0951832015001672 (visited on 09/09/2023).

[47] Kubben. *Fundamentals of clinical data science*. New York, NY: Springer Berlin Heidelberg, 2018. ISBN: 9783319997124.

[48] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau. "Random Forests: some methodological insights". In: (2008). DOI: 10.48550/ARXIV.0811.3619. URL: https://arxiv.org/abs/0811.3619 (visited on 09/09/2023).

[49] Georg Heinze, Christine Wallisch, and Daniela Dunkler. "Variable selection - A review and recommendations for the practicing statistician". In: *Biometrical Journal* 60.3 (May 2018), pp. 431–449. ISSN: 03233847. DOI: 10.1002/bimj.201700067. URL: https://onlinelibrary.wiley.com/doi/10.1002/bimj.201700067 (visited on 09/09/2023).

[50] Sonja Grill et al. "Comparison of approaches for incorporating new information into existing risk prediction models". In: *Statistics in medicine* 36.7 (2017), pp. 1134–1156.

[51] William Nigel Strang et al. "The French organ transplant data system". In: *Studies in Health Technology and Informatics* 116 (2005), pp. 77–82. ISSN: 1879-8365.

[52] Ilse Duus Weinreich et al. "Scandiatransplant Exchange Program (STEP): Development and Results From an International Kidney Exchange Program". In: *Transplantation Direct* 9.11 (Oct. 16, 2023), e1549. ISSN: 2373-8731. DOI: 10.1097/TXD.0000000000001549. URL: https://journals.lww.com/10.1097/TXD.0000000000001549 (visited on 10/31/2024).

[53] Borja M. Fernandez-Felix et al. "CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models". In: *BMC Medical Research Methodology* 23.1 (Feb. 17, 2023), p. 44. ISSN: 1471-2288. DOI: 10.1186/s12874-023-01849-0. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-01849-0 (visited on 08/17/2024).

[54] Andrzej Konieczny et al. "Artificial Intelligence—A Tool for Risk Assessment of Delayed-Graft Function in Kidney Transplant". In: *Journal of Clinical Medicine* 10.22 (Nov. 11, 2021), p. 5244. ISSN: 2077-0383. DOI: 10.3390/jcm10225244. URL: https://www.mdpi.com/2077-0383/10/22/5244 (visited on 08/17/2024).

[55] D.A. Shoskes et al. "Prediction of Early Graft Function in Renal Transplantation Using a Computer Neural Network". In: *Transplantation Proceedings* 30.4 (June 1998), pp. 1316–1317. ISSN: 00411345. DOI: 10.1016/S0041-1345(98)00257-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0041134598002577 (visited on 08/17/2024).

[56] M. E. Brier, P. C. Ray, and J. B. Klein. "Prediction of delayed renal allograft function using an artificial neural network". In: *Nephrology Dialysis Transplantation* 18.12 (Dec. 1, 2003), pp. 2655–2659. ISSN: 0931-0509, 1460-2385. DOI: 10.1093/ndt/gfg439. URL: https://academic.oup.com/ndt/article-lookup/doi/10.1093/ndt/gfg439 (visited on 08/17/2024).

[57] Alfonso H. Santos et al. "Common predictors of adverse outcomes in adult deceased donor kidney transplant recipients with varying sensitization". In: *American Journal of Clinical and Experimental Urology* 11.3 (2023), pp. 235–248. ISSN: 2330-1910. DOI: https://pmc.ncbi.nlm.nih.gov/articles/PMC10333136/.

[58] Vincenzo Losappio et al. "A Single-Center Cohort Study to Define the Role of Pretransplant Biopsy Score in the Long-term Outcome of Kidney Transplantation". In: *Transplantation* 97.9 (May 15, 2014), pp. 934–939. ISSN: 0041-1337. DOI: 10.1097/01.TP.0000438208.50089.29. URL: https://journals.lww.com/00007890-201405150-00010 (visited on 08/17/2024).

[59] Sunjae Bae et al. "Machine learning to predict transplant outcomes: helpful or hype? A national cohort study". In: *Transplant International* 33.11 (Nov. 2020), pp. 1472–1480. ISSN: 0934-0874, 1432-2277. DOI: 10.1111/tri.13695. URL: https://onlinelibrary.wiley.com/doi/10.1111/tri.13695 (visited on 08/17/2024).

[60] William D. Irish et al. "Nomogram for Predicting the Likelihood of Delayed Graft Function in Adult Cadaveric Renal Transplant Recipients". In: *Journal of the American Society of Nephrology* 14.11 (Nov. 2003), pp. 2967–2974. ISSN: 1046-6673. DOI: 10.1097/01.ASN.0000093254.31868.85. URL: https://journals.lww.com/00001751-200311000-00032 (visited on 01/18/2024).

[61] Claudio Jeldres et al. "Prediction of delayed graft function after renal transplantation". In: *Canadian Urological Association Journal* 3.5 (May 1, 2013), p. 377. ISSN: 1920-1214. DOI: 10.5489/cuaj.1147. URL: http://cuaj.ca/index.php/journal/article/view/1147 (visited on 08/17/2024).

[62] W.D Irish et al. "A Risk Prediction Model for Delayed Graft Function in the Current Era of Deceased Donor Renal Transplantation". In: *American Journal of Transplantation* 10.10 (Oct. 2010), pp. 2279–2286. ISSN: 16006135. DOI: 10.1111/j.1600-6143.2010.03179.x. URL: https://linkinghub.elsevier.com/retrieve/pii/S1600613522144382 (visited on 08/17/2024).

[63] Peter Balaz et al. "Identification of Expanded-Criteria Donor Kidney Grafts at Lower Risk of Delayed Graft Function". In: *Transplantation* 96.7 (Oct. 15, 2013), pp. 633–638. ISSN: 0041-1337. DOI: 10.1097/TP.0b013e31829d9225. URL: https://journals.lww.com/00007890-201310150-00008 (visited on 08/17/2024).

[64] Marion Chapal et al. "A useful scoring system for the prediction and management of delayed graft function following kidney transplantation from cadaveric donors". In: *Kidney International* 86.6 (Dec. 2014), pp. 1130–1139. ISSN: 00852538. DOI: 10.1038/ki.2014.188. URL: https://linkinghub.elsevier.com/retrieve/pii/S0085253815304464 (visited on 08/17/2024).

[65] Chen-Guang Ding et al. "Predictive Score Model for Delayed Graft Function Based on Easily Available Variables before Kidney Donation after Cardiac Death". In: *Chinese Medical Journal* 130.20 (Oct. 20, 2017), pp. 2429–2434. ISSN: 0366-6999. DOI: 10.4103/0366-6999.216409. URL: https://journals.lww.com/00029330-201710200-00006 (visited on 08/17/2024).

[66] Qipeng Sun et al. "New Factors Predicting Delayed Graft Function: a Multi-Center Cohort Study of Kidney Donation After Brain Death Followed by Circulatory Death". In: *Kidney and Blood Pressure Research* 43.3 (2018), pp. 893–903. ISSN: 1420-4096, 1423-0143. DOI: 10.1159/000490337. URL: https://karger.com/KBR/article/doi/10.1159/000490337 (visited on 08/17/2024).

[67] Rao Chen et al. "Predictors and one-year outcomes of patients with delayed graft function after deceased donor kidney transplantation". In: *BMC Nephrology* 21.1 (Dec. 2020), p. 526. ISSN: 1471-2369. DOI: 10.1186/s12882-020-02181-1. URL: https://bmcnephrol.biomedcentral.com/articles/10.1186/s12882-020-02181-1 (visited on 08/17/2024).

[68] Jin Zheng et al. "Comprehensive assessment of deceased donor kidneys with clinical characteristics, pre-implant biopsy histopathology and hypothermic mechanical perfusion parameters is highly predictive of delayed graft function". In: *Renal Failure* 42.1 (Jan. 2020), pp. 369–376. ISSN: 0886-022X, 1525-6049. DOI: 10.1080/0886022X.2020.1752716. URL: https://www.tandfonline.com/doi/full/10.1080/0886022X.2020.1752716 (visited on 08/17/2024).

[69] Satoru Kawakita et al. "Personalized prediction of delayed graft function for recipients of deceased donor kidney transplants with machine learning". In: *Scientific Reports* 10.1 (Oct. 27, 2020), p. 18409. ISSN: 2045-2322. DOI: 10.1038/s41598-020-75473-z. URL: https://www.nature.com/articles/s41598-020-75473-z (visited on 08/17/2024).

[70] You Luo et al. "Deep Learning Algorithms for the Prediction of Posttransplant Renal Function in Deceased-Donor Kidney Recipients: A Preliminary Study Based on Pretransplant Biopsy". In: *Frontiers in Medicine* 8 (Jan. 18, 2022), p. 676461. ISSN: 2296-858X. DOI: 10.3389/fmed.2021.676461. URL: https://www.frontiersin.org/articles/10.3389/fmed.2021.676461/full (visited on 08/17/2024).

[71] Wujun Xue et al. "A prediction model of delayed graft function in deceased donor for renal transplant: a multi-center study from China". In: *Renal Failure* 43.1 (Jan. 1, 2021), pp. 520–529. ISSN: 0886-022X, 1525-6049. DOI: 10.1080/0886022X.2021.1895838. URL: https://www.tandfonline.com/doi/full/10.1080/0886022X.2021.1895838 (visited on 08/17/2024).

[72] Shitao Zhao et al. "Prediction model of delayed graft function based on clinical characteristics combined with serum IL-2 levels". In: *BMC Nephrology* 23.1 (Aug. 15, 2022), p. 284. ISSN: 1471-2369. DOI: 10.1186/s12882-022-02908-2. URL: https://bmcnephrol.biomedcentral.com/articles/10.1186/s12882-022-02908-2 (visited on 08/17/2024).

[73] Jiashan Pan and Guiyi Liao. "Development and Validation of Nomogram for Predicting Delayed Graft Function After Kidney Transplantation of Deceased Donor". In: *International Journal of General Medicine* Volume 14 (Nov. 2021), pp. 9103–9115. ISSN: 1178-7074. DOI: 10.2147/IJGM.S331854. URL: https://www.dovepress.com/development-and-validation-

of-nomogram-for-predicting-delayed-graft-fu-peer-reviewed-fulltext-article-IJGM (visited on 08/17/2024).

[74] Silvana Daher Costa et al. "The impact of deceased donor maintenance on delayed kidney allograft function: A machine learning analysis". In: *PLOS ONE* 15.2 (Feb. 6, 2020). Ed. by Frank Jmf Dor, e0228597. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0228597. URL: https://dx.plos.org/10.1371/journal.pone.0228597 (visited on 08/17/2024).

[75] Yuxi Qiao et al. "Predictive value of hypothermic machine perfusion parameters combined perfusate biomarkers in deceased donor kidney transplantation". In: *Chinese Medical Journal* 135.2 (Jan. 20, 2022), pp. 181–186. ISSN: 0366-6999, 2542-5641. DOI: 10.1097/CM9.0000000000001867. URL: https://journals.lww.com/10.1097/CM9.0000000000001867 (visited on 08/17/2024).

[76] Kuang-Yu Jen et al. "Automated En Masse Machine Learning Model Generation Shows Comparable Performance as Classic Regression Models for Predicting Delayed Graft Function in Renal Allografts". In: *Transplantation* 105.12 (Dec. 2021), pp. 2646–2654. ISSN: 0041-1337. DOI: 10.1097/TP.0000000000003640. URL: https://journals.lww.com/10.1097/TP.0000000000003640 (visited on 11/10/2024).

[77] Chen-Guang Ding et al. "Predictive Score Model for Delayed Graft Function Based on Hypothermic Machine Perfusion Variables in Kidney Transplantation". In: *Chinese Medical Journal* 131.22 (Nov. 20, 2018), pp. 2651–2657. ISSN: 0366-6999. DOI: 10.4103/0366-6999.245278. URL: https://journals.lww.com/00029330-201811200-00002 (visited on 08/17/2024).

[78] E. Vittinghoff and C. E. McCulloch. "Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression". In: *American Journal of Epidemiology* 165.6 (Jan. 12, 2007), pp. 710–718. ISSN: 0002-9262, 1476-6256. DOI: 10.1093/aje/kwk052. URL: https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwk052 (visited on 08/19/2024).

[79] Friedrich K. Port et al. "Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors1:" in: *Transplantation* 74.9 (Nov. 2002), pp. 1281–1286. ISSN: 0041-1337. DOI: 10.1097/00007890-200211150-00014. URL: http://journals.lww.com/00007890-200211150-00014 (visited on 08/17/2024).

[80] Ray S. Lin et al. "Single and multiple time-point prediction models in kidney transplant outcomes". In: *Journal of Biomedical Informatics* 41.6 (Dec. 2008), pp. 944–952. ISSN: 15320464. DOI: 10.1016/j.jbi.2008.03.

005. URL: https://linkinghub.elsevier.com/retrieve/pii/S1532046408000439 (visited on 08/17/2024).

[81] François-Xavier Paquette et al. "Machine Learning Support for Decision-Making in Kidney Transplantation: Step-by-step Development of a Technological Solution". In: *JMIR Medical Informatics* 10.6 (June 14, 2022), e34554. ISSN: 2291-9694. DOI: 10.2196/34554. URL: https://medinform.jmir.org/2022/6/e34554 (visited on 08/17/2024).

[82] Jesse D. Schold et al. "The Broad Spectrum of Quality in Deceased Donor Kidneys". In: *American Journal of Transplantation* 5.4 (Apr. 2005), pp. 757–765. ISSN: 16006135. DOI: 10.1111/j.1600-6143.2005.00770.x. URL: https://linkinghub.elsevier.com/retrieve/pii/S1600613522036358 (visited on 08/17/2024).

[83] Sergey Krikov et al. "Predicting Kidney Transplant Survival Using Tree-Based Modeling". In: *ASAIO Journal* 53.5 (Sept. 2007), pp. 592–600. ISSN: 1058-2916. DOI: 10.1097/MAT.0b013e318145b9f7. URL: https://pubmed.ncbi.nlm.nih.gov/17885333/ (visited on 08/17/2024).

[84] Trevor S. Brown et al. "Bayesian Modeling of Pretransplant Variables Accurately Predicts Kidney Graft Survival". In: *American Journal of Nephrology* 36.6 (2012), pp. 561–569. ISSN: 0250-8095, 1421-9670. DOI: 10.1159/000345552. URL: https://karger.com/AJN/article/doi/10.1159/000345552 (visited on 08/17/2024).

[85] Syed Asil Ali Naqvi et al. "Predicting Kidney Graft Survival Using Machine Learning Methods: Prediction Model Development and Feature Significance Analysis Study". In: *Journal of Medical Internet Research* 23.8 (Aug. 27, 2021), e26843. ISSN: 1438-8871. DOI: 10.2196/26843. URL: https://www.jmir.org/2021/8/e26843 (visited on 08/17/2024).

[86] Leila Shahmoradi et al. "Comparing Three Data Mining Methods to Predict Kidney Transplant Survival". In: *Acta Informatica Medica* 24.5 (2016), p. 322. ISSN: 0353-8109. DOI: 10.5455/aim.2016.24.322-327. URL: http://www.scopemed.org/fulltextpdf.php?mno=247776 (visited on 08/17/2024).

[87] Jiakai Li et al. "Bayes Net Classifiers For Prediction Of Renal Graft Status And Survival Period". In: (Mar. 25, 2010). DOI: 10.5281/ZENODO.1334844. URL: https://zenodo.org/record/1334844 (visited on 08/17/2024).

[88] C.J. Wang et al. "The Donor Kidney Biopsy and Its Implications in Predicting Graft Outcomes: A Systematic Review". In: *American Journal of Transplantation* 15.7 (July 2015), pp. 1903–1914. ISSN: 16006135. DOI: 10.1111/ajt.13213. URL: https://linkinghub.elsevier.com/retrieve/pii/S1600613522003379 (visited on 08/25/2024).

[89] Beat Moeckli et al. "Evaluation of donor kidneys prior to transplantation: an update of current and emerging methods". In: *Transplant International* 32.5 (May 2019), pp. 459–469. ISSN: 0934-0874, 1432-2277. DOI: 10.1111/tri.13430. URL: https://onlinelibrary.wiley.com/doi/10.1111/tri.13430 (visited on 08/25/2024).

[90] S. Ali Husain et al. "Reproducibility of Deceased Donor Kidney Procurement Biopsies". In: *Clinical Journal of the American Society of Nephrology* 15.2 (Feb. 2020), pp. 257–264. ISSN: 1555-9041, 1555-905X. DOI: 10.2215/CJN.09170819. URL: https://journals.lww.com/10.2215/CJN.09170819 (visited on 08/25/2024).

[91] M. Antonieta Azancot et al. "The reproducibility and predictive value on outcome of renal biopsies from expanded criteria donors". In: *Kidney International* 85.5 (May 2014), pp. 1161–1168. ISSN: 00852538. DOI: 10.1038/ki.2013.461. URL: https://linkinghub.elsevier.com/retrieve/pii/S0085253815563229 (visited on 08/25/2024).

[92] Eurotransplant. *Eurotransplant Manual*. https://www.eurotransplant.org/allocation/eurotransplant-manual/. Accessed: 01.05.2024. January 2024.

[93] ANZDATA. *ANZDATA Data Set Specification*. Accessed: 01.04.2024. 2023.

[94] UK Renal Registry. *UKRR: CKD/AKI clinical dataset*. Accessed: 01.04.2024. 2024.

[95] John Oo Ayorinde et al. "PreImplantation Trial of Histopathology In renal Allografts (PITHIA): a stepped-wedge cluster randomised controlled trial protocol". In: *BMJ Open* 9.1 (Jan. 2019), e026166. ISSN: 2044-6055, 2044-6055. DOI: 10.1136/bmjopen-2018-026166. URL: https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2018-026166 (visited on 12/07/2024).

[96] Erik Von Elm et al. "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies". In: *Annals of Internal Medicine* 147.8 (Oct. 16, 2007), p. 573. ISSN: 0003-4819. DOI: 10.7326/0003-4819-147-8-200710160-00010. URL: http://annals.org/article.aspx?doi=10.7326/0003-4819-147-8-200710160-00010 (visited on 08/26/2024).

[97] R.B. Munivenkatappa et al. "The Maryland Aggregate Pathology Index: A Deceased Donor Kidney Biopsy Scoring System for Predicting Graft Failure". In: *American Journal of Transplantation* 8.11 (Nov. 2008), pp. 2316–2324. ISSN: 16006135. DOI: 10.1111/j.1600-6143.2008.02370.x. URL: https://linkinghub.elsevier.com/retrieve/pii/S1600613522029951 (visited on 08/25/2024).

[98] Satu R. K. Lehtonen, Eero I. Taskinen, and Helena M. Isoniemi. "Histopathological findings in renal allografts at time of transplantation and correlation with onset of graft function*". In: *APMIS* 107.7 (Mar. 1999), pp. 945–950. ISSN: 0903-4641, 1600-0463. DOI: 10.1111/j.1699-0463.1999.tb01495.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1699-0463.1999.tb01495.x (visited on 08/25/2024).

[99] Giuseppe Remuzzi et al. "Long-Term Outcome of Renal Transplantation from Older Donors". In: *New England Journal of Medicine* 354.4 (Jan. 26, 2006), pp. 343–352. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa052891. URL: http://www.nejm.org/doi/abs/10.1056/NEJMoa052891 (visited on 08/25/2024).

[100] Alexandre Loupy et al. "The Banff 2019 Kidney Meeting Report (I): Updates on and clarification of criteria for T cell- and antibody-mediated rejection". In: *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 20.9 (Sept. 2020), pp. 2318–2331. ISSN: 1600-6143. DOI: 10.1111/ajt.15898.

[101] D. Anglicheau et al. "A Simple Clinico-Histopathological Composite Scoring System Is Highly Predictive of Graft Outcomes in Marginal Donors". In: *American Journal of Transplantation* 8.11 (Nov. 2008), pp. 2325–2334. ISSN: 16006135. DOI: 10.1111/j.1600-6143.2008.02394.x. URL: https://linkinghub.elsevier.com/retrieve/pii/S1600613522029963 (visited on 08/25/2024).

[102] Katrien De Vusser et al. "The Predictive Value of Kidney Allograft Baseline Biopsies for Long-Term Graft Survival". In: *Journal of the American Society of Nephrology* 24.11 (Nov. 2013), pp. 1913–1923. ISSN: 1046-6673. DOI: 10.1681/ASN.2012111081. URL: https://journals.lww.com/00001751-201311000-00023 (visited on 08/25/2024).

[103] M.G.J. Snoeijs et al. "Histological Assessment of Preimplantation Biopsies May Improve Selection of Kidneys from Old Donors After Cardiac Death". In: *American Journal of Transplantation* 8.9 (Sept. 2008), pp. 1844–1851. ISSN: 16006135. DOI: 10.1111/j.1600-6143.2008.02318.x. URL: https://

linkinghub.elsevier.com/retrieve/pii/S1600613522056933 (visited on 11/23/2023).

[104] María Dolores Navarro et al. "Significance of Preimplantation Analysis of Kidney Biopsies From Expanded Criteria Donors in Long-Term Outcome". In: *Transplantation* 91.4 (Feb. 27, 2011), pp. 432–439. ISSN: 0041-1337. DOI: 10.1097/TP.0b013e318204bdd7. URL: https://journals.lww.com/00007890-201102270-00009 (visited on 08/25/2024).

[105] André Costa Teixeira et al. "Procurement Biopsies Can Predict Unfavorable Outcomes in Kidneys With Low MAPI Score Values". In: *Transplantation Proceedings* 53.2 (Mar. 2021), pp. 602–606. ISSN: 00411345. DOI: 10.1016/j.transproceed.2020.09.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0041134520304401 (visited on 08/25/2024).

[106] Stefano Marletta et al. "Pre-Implantation Kidney Biopsies in Extended Criteria Donors: From On Call to Expert Pathologist, from Conventional Microscope to Digital Pathology". In: *Critical Reviews™ in Oncogenesis* 28.3 (2023), pp. 7–20. ISSN: 0893-9675. DOI: 10.1615/CritRevOncog.2023049007. URL: https://www.dl.begellhouse.com/journals/439f422d0783386a,52b4a9bc5c9ba950,529334e5040cac4c.html (visited on 08/25/2024).

[107] Jesper Kers et al. "Deep learning-based classification of kidney transplant pathology: a retrospective, multicentre, proof-of-concept study". In: *The Lancet Digital Health* 4.1 (Jan. 2022), e18–e26. ISSN: 25897500. DOI: 10.1016/S2589-7500(21)00211-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S2589750021002119 (visited on 08/25/2024).

[108] Zhengzi Yi et al. "Deep learning identified pathological abnormalities predictive of graft loss in kidney transplant biopsies". In: *Kidney International* 101.2 (Feb. 2022), pp. 288–298. ISSN: 00852538. DOI: 10.1016/j.kint.2021.09.028. URL: https://linkinghub.elsevier.com/retrieve/pii/S0085253821010255 (visited on 08/25/2024).

[109] European Parliament. *DIRECTIVE 2010/45/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 7 July 2010 on standards of quality and safety of human organs intended for transplantation*. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:207:0014:0029:EN:PDF. Accessed: 01.05.2024. July 2010.

[110] Houneida Sakly et al. "Medical decision-Making based on Combined CRISP-DM Approach and CNN Classification for Cardiac MRI". In: *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*. 2022 IEEE 9th International Conference on

Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). Hammamet, Tunisia: IEEE, May 28, 2022, pp. 25–30. ISBN: 978-1-72818-442-5. DOI: 10.1109/SETIT54465.2022.9875820. URL: https://ieeexplore.ieee.org/document/9875820/ (visited on 09/11/2023).

[111] Stefan Studer et al. "Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology". In: *Machine Learning and Knowledge Extraction* 3.2 (Apr. 22, 2021), pp. 392–413. ISSN: 2504-4990. DOI: 10.3390/make3020020. URL: https://www.mdpi.com/2504-4990/3/2/20 (visited on 09/11/2023).

[112] E. W. Steyerberg and Y. Vergouwe. "Towards better clinical prediction models: seven steps for development and an ABCD for validation". In: *European Heart Journal* 35.29 (Aug. 1, 2014), pp. 1925–1931. ISSN: 0195-668X, 1522-9645. DOI: 10.1093/eurheartj/ehu207. URL: https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehu207 (visited on 09/11/2023).

[113] Patrick Royston and Willi Sauerbrei. *Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. Wiley series in probability and statistics. OCLC: ocn191697435. Chichester, England ; Hoboken, NJ: John Wiley, 2008. 303 pp. ISBN: 978-0-470-02842-1.

[114] David J. C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge, UK ; New York: Cambridge University Press, 2003. 628 pp. ISBN: 978-0-521-64298-9.

[115] Kuhn and Max. "Building Predictive Models in R Using the caret Package". In: *Journal of Statistical Software* 28.5 (2008), pp. 1–26. DOI: 10.18637/jss.v028.i05. URL: https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

[116] Ewout W. Steyerberg. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Statistics for Biology and Health. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-16398-3 978-3-030-16399-0. DOI: 10.1007/978-3-030-16399-0. URL: http://link.springer.com/10.1007/978-3-030-16399-0 (visited on 09/30/2023).

[117] Georg Heinze and Daniela Dunkler. "Five myths about variable selection". In: *Transplant International* 30.1 (Jan. 2017), pp. 6–10. ISSN: 09340874. DOI: 10.1111/tri.12895. URL: https://onlinelibrary.wiley.com/doi/10.1111/tri.12895 (visited on 09/23/2023).

[118] Frank E. Harrell. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer Series in Statistics. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-19424-0 978-3-319-19425-7. DOI: 10.1007/978-3-319-19425-7. URL: https://link.springer.com/10.1007/978-3-319-19425-7 (visited on 09/25/2023).

[119] P. Royston et al. "Prognosis and prognostic research: Developing a prognostic model". In: *BMJ* 338 (mar31 1 Mar. 31, 2009), b604–b604. ISSN: 0959-8138, 1468-5833. DOI: 10.1136/bmj.b604. URL: https://www.bmj.com/lookup/doi/10.1136/bmj.b604 (visited on 09/25/2023).

[120] Katerina M. Marcoulides and Tenko Raykov. "Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods". In: *Educational and Psychological Measurement* 79.5 (Oct. 2019), pp. 874–882. ISSN: 0013-1644, 1552-3888. DOI: 10.1177/0013164418817803. URL: https://journals.sagepub.com/doi/10.1177/0013164418817803 (visited on 10/11/2024).

[121] Babak Naimi et al. "Where is positional uncertainty a problem for species distribution modelling". In: *Ecography* 37 (2014), pp. 191–203. DOI: 10.1111/j.1600-0587.2013.00205.x.

[122] John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage, 2019. URL: https://www.john-fox.ca/Companion/.

[123] Menelaos Pavlou, Gareth Ambler, and Rumana Z. Omar. "Risk prediction in multicentre studies when there is confounding by cluster or informative cluster size". In: *BMC Medical Research Methodology* 21.1 (Dec. 2021), p. 135. ISSN: 1471-2288. DOI: 10.1186/s12874-021-01321-x. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-021-01321-x (visited on 09/25/2023).

[124] Rose Sisk et al. "Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: A simulation study". In: *Statistical Methods in Medical Research* 32.8 (Aug. 2023), pp. 1461–1477. ISSN: 0962-2802, 1477-0334. DOI: 10.1177/09622802231165001. URL: http://journals.sagepub.com/doi/10.1177/09622802231165001 (visited on 09/24/2023).

[125] Sebastian Jäger, Arndt Allhorn, and Felix Bießmann. "A Benchmark for Data Imputation Methods". In: *Frontiers in Big Data* 4 (July 8, 2021), p. 693674. ISSN: 2624-909X. DOI: 10.3389/fdata.2021.693674. URL: https://

www.frontiersin.org/articles/10.3389/fdata.2021.693674/full (visited on 09/25/2023).

[126] Benjamin Yaël Gravesteijn et al. "Missing Data in Prediction Research: A Five-Step Approach for Multiple Imputation, Illustrated in the CENTER-TBI Study". In: *Journal of Neurotrauma* 38.13 (July 1, 2021), pp. 1842–1857. ISSN: 0897-7151, 1557-9042. DOI: 10.1089/neu.2020.7218. URL: https://www.liebertpub.com/doi/10.1089/neu.2020.7218 (visited on 09/25/2023).

[127] Zhongmin Tian and Mingyu Liang. "Renal metabolism and hypertension". In: *Nature Communications* 12.1 (Feb. 11, 2021), p. 963. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21301-5. URL: https://www.nature.com/articles/s41467-021-21301-5 (visited on 09/30/2023).

[128] Megumi Oshima et al. "Trajectories of kidney function in diabetes: a clinicopathological update". In: *Nature Reviews Nephrology* 17.11 (Nov. 2021), pp. 740–750. ISSN: 1759-5061, 1759-507X. DOI: 10.1038/s41581-021-00462-y. URL: https://www.nature.com/articles/s41581-021-00462-y (visited on 09/30/2023).

[129] David W. Hosmer, Stanley Lemeshow, and Rodney X. Sturdivant. *Applied Logistic Regression*. 1st ed. Wiley Series in Probability and Statistics. Wiley, Mar. 22, 2013. ISBN: 978-0-470-58247-3 978-1-118-54838-7. DOI: 10.1002/9781118548387. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118548387 (visited on 09/22/2023).

[130] Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. 1st ed. Wiley Series in Probability and Statistics. Wiley, June 9, 1987. ISBN: 978-0-471-08705-2 978-0-470-31669-6. DOI: 10.1002/9780470316696. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316696 (visited on 09/30/2023).

[131] Peter C. Austin et al. "Effect of Variable Selection Strategy on the Performance of Prognostic Models When Using Multiple Imputation". In: *Circulation: Cardiovascular Quality and Outcomes* 12.11 (Nov. 2019), e005927. ISSN: 1941-7713, 1941-7705. DOI: 10.1161/CIRCOUTCOMES.119.005927. URL: https://www.ahajournals.org/doi/10.1161/CIRCOUTCOMES.119.005927 (visited on 09/30/2023).

[132] Andrea Marshall et al. "Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines". In: *BMC medical research methodology* 9 (July 28, 2009), p. 57. ISSN: 1471-2288. DOI: 10.1186/1471-2288-9-57.

[133] Iris Eekhout, Mark A. Van De Wiel, and Martijn W. Heymans. "Methods for significance testing of categorical covariates in logistic regression models after multiple imputation: power and applicability analysis". In: *BMC Medical Research Methodology* 17.1 (Dec. 2017), p. 129. ISSN: 1471-2288. DOI: 10.1186/s12874-017-0404-7. URL: http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-017-0404-7 (visited on 10/22/2023).

[134] M. A. Van De Wiel, J. Berkhof, and W. N. Van Wieringen. "Testing the prediction error difference between 2 predictors". In: *Biostatistics* 10.3 (July 1, 2009), pp. 550–560. ISSN: 1465-4644, 1468-4357. DOI: 10.1093/biostatistics/kxp011. URL: https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/kxp011 (visited on 10/22/2023).

[135] Stef van Buuren and Karin Groothuis-Oudshoorn. "mice: Multivariate Imputation by Chained Equations in R". In: *Journal of Statistical Software* 45.3 (2011), pp. 1–67. DOI: 10.18637/jss.v045.i03.

[136] Martijn Heymans. *miceafter: Data and Statistical Analyses after Multiple Imputation*. R package version 0.5.0. 2022. URL: https://CRAN.R-project.org/package=miceafter.

[137] H. Akaike. "On Newer Statistical Approaches to Parameter Estimation and Structure Determination". In: *IFAC Proceedings Volumes* (1978). DOI: https://doi.org/10.1016/S1474-6670(17)66162-7.

[138] Jerome Friedman, Robert Tibshirani, and Trevor Hastie. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (2010), pp. 1–22. DOI: 10.18637/jss.v033.i01.

[139] Robert B. Bendel and A. A. Afifi. "Comparison of Stopping Rules in Forward "Stepwise" Regression". In: *Journal of the American Statistical Association* 72.357 (Mar. 1977), p. 46. ISSN: 01621459. DOI: 10.2307/2286904. URL: https://www.jstor.org/stable/2286904?origin=crossref (visited on 09/23/2023).

[140] Ruth M. Mickey and Sander Greenland. "THE IMPACT OF CONFOUNDER SELECTION CRITERIA ON EFFECT ESTIMATION". In: *American Journal of Epidemiology* 129.1 (Jan. 1989), pp. 125–137. ISSN: 1476-6256, 0002-9262. DOI: 10.1093/oxfordjournals.aje.a115101. URL: https://academic.oup.com/aje/article/58697/THE (visited on 09/23/2023).

[141] Max Halperin, William C. Blackwelder, and Joel I. Verter. "Estimation of the multivariate logistic risk function: A comparison of the discriminant function and maximum likelihood approaches". In: *Journal of Chronic Diseases* 24.2 (July 1971), pp. 125–158. ISSN: 00219681. DOI: 10.1016/0021-9681(71)90106-8. URL: https://linkinghub.elsevier.com/retrieve/pii/0021968171901068 (visited on 09/23/2023).

[142] Sam Doerken et al. "Penalized logistic regression with low prevalence exposures beyond high dimensional settings". In: *PloS One* 14.5 (2019), e0217057. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0217057.

[143] Emine Ozgur Bayman and Franklin Dexter. "Multicollinearity in Logistic Regression Models". In: *Anesthesia & Analgesia* 133.2 (Aug. 2021), pp. 362–365. ISSN: 0003-2999. DOI: 10.1213/ANE.0000000000005593. URL: https://journals.lww.com/10.1213/ANE.0000000000005593 (visited on 10/23/2023).

[144] Carlisle Rainey and Kelly McCaskey. "Estimating logit models with small samples". In: *Political Science Research and Methods* 9.3 (July 2021), pp. 549–564. ISSN: 2049-8470, 2049-8489. DOI: 10.1017/psrm.2021.9. URL: https://www.cambridge.org/core/product/identifier/S2049847021000091/type/journal_article (visited on 10/29/2023).

[145] Tibshirani, R. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society, Serias B (Methological)* 58.1 (1996), pp. 267–88. URL: http://www.jstor.org/stable/2346178.

[146] Pau Rodríguez et al. "Beyond one-hot encoding: Lower dimensional target embedding". In: *Image and Vision Computing* 75 (July 2018), pp. 21–31. ISSN: 02628856. DOI: 10.1016/j.imavis.2018.04.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0262885618300623 (visited on 11/19/2023).

[147] Ewout W. Steyerberg. *Clinical prediction models: a practical approach to development, validation, and updating*. Statistics for biology and health. OCLC: ocn295033353. New York, NY: Springer, 2009. 497 pp. ISBN: 9780387772431 9780387772448.

[148] Achim Ahrens, Christian B. Hansen, and Mark E. Schaffer. "lassopack: Model selection and prediction with regularized regression in Stata". In: *The Stata Journal: Promoting communications on statistics and Stata* 20.1 (Mar. 2020), pp. 176–235. ISSN: 1536-867X, 1536-8734. DOI: 10.1177/1536867X20909697. URL: http://journals.sagepub.com/doi/10.1177/1536867X20909697 (visited on 01/08/2024).

[149] Alexandre Belloni and Victor Chernozhukov. "Least squares after model selection in high-dimensional sparse models". In: *Bernoulli* 19.2 (May 1, 2013). ISSN: 1350-7265. DOI: 10.3150/11-BEJ410. URL: https://projecteuclid.org/journals/bernoulli/volume-19/issue-2/Least-squares-after-model-selection-in-high-dimensional-sparse-models/10.3150/11-BEJ410.full (visited on 01/08/2024).

[150] Kuhn and Max. "Building Predictive Models in R Using the caret Package". In: *Journal of Statistical Software* 28.5 (2008), pp. 1–26. DOI: 10.18637/jss.v028.i05. URL: https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

[151] Robert Tibshirani et al. "Strong Rules for Discarding Predictors in Lasso-Type Problems". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 74.2 (Mar. 1, 2012), pp. 245–266. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/j.1467-9868.2011.01004.x. URL: https://academic.oup.com/jrsssb/article/74/2/245/7075310 (visited on 11/19/2023).

[152] James N. Morgan and John A. Sonquist. "Problems in the Analysis of Survey Data, and a Proposal". In: *Journal of the American Statistical Association* 58.302 (June 1963), pp. 415–434. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1963.10500855. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500855 (visited on 01/15/2024).

[153] Leo Breiman, ed. *Classification and regression trees*. 1. CRC Press repr. Boca Raton, Fla.: Chapman & Hall/CRC, 1998. 358 pp. ISBN: 9780412048418.

[154] Wei-Yin Loh. "Fifty Years of Classification and Regression Trees". In: *International Statistical Review* 82.3 (Dec. 2014), pp. 329–348. ISSN: 0306-7734, 1751-5823. DOI: 10.1111/insr.12016. URL: https://onlinelibrary.wiley.com/doi/10.1111/insr.12016 (visited on 01/11/2024).

[155] L. Gordon and R. A. Olshen. "Tree-structured survival analysis". In: *Cancer Treatment Reports* 69.10 (Oct. 1985), pp. 1065–1069. ISSN: 0361-5960.

[156] M. LeBlanc and J. Crowley. "Relative risk trees for censored survival data". In: *Biometrics* 48.2 (June 1992), pp. 411–425. ISSN: 0006-341X.

[157] Martin Krzywinski and Naomi Altman. "Classification and regression trees". In: *Nature Methods* 14.8 (Aug. 1, 2017), pp. 757–758. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4370. URL: https://www.nature.com/articles/nmeth.4370 (visited on 01/14/2024).

[158] T. M. Therneau and Atkinson E. J. "An Introduction to Recursive Partitioning Using the RPART Routines". In: *Mayo Foundation* (2023).

[159] Tuleau-Malot C. Genuer R. Poggi J-M. "VSURF: An R Package for Variable Selection Using Random Forests". In: *The R Journal* 7/2 (2015), pp. 19–33. ISSN: 2073-4859.

[160] Tuleau-Malot C. Genuer R. Poggi J.-M. "Variable selection using random forests". In: *Pattern of Recognition Letters* (2010b), pp. 19–23, 27.

[161] Leo Breiman. "[Random Forests]". In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 08856125. DOI: 10.1023/A:1010933404324. URL: http://link.springer.com/10.1023/A:1010933404324 (visited on 01/20/2024).

[162] M. A. E. Binuya et al. "Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review". In: *BMC Med Res Methodol* 22.1 (2022), p. 316. ISSN: 1471-2288 (Electronic) 1471-2288 (Linking). DOI: 10.1186/s12874-022-01801-8. URL: https://www.ncbi.nlm.nih.gov/pubmed/36510134.

[163] Lisa M. Sullivan, Joseph M. Massaro, and Ralph B. D'Agostino. "Presentation of multivariate data for clinical use: The Framingham Study risk score functions". In: *Statistics in Medicine* 23.10 (May 30, 2004), pp. 1631–1660. ISSN: 0277-6715, 1097-0258. DOI: 10.1002/sim.1742. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.1742 (visited on 11/10/2024).

[164] L. M. Sullivan, J. M. Massaro, and Sr. D'Agostino R. B. "Presentation of multivariate data for clinical use: The Framingham Study risk score functions". In: *Stat Med* 23.10 (2004), pp. 1631–60. ISSN: 0277-6715 (Print) 0277-6715 (Linking). DOI: 10.1002/sim.1742. URL: https://www.ncbi.nlm.nih.gov/pubmed/15122742.

[165] P. Bankhead et al. "QuPath: Open source software for digital pathology image analysis". In: *Sci Rep* 7.1 (2017), p. 16878. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-017-17204-5. URL: https://www.ncbi.nlm.nih.gov/pubmed/29203879.

[166] Sonja Grill et al. "Comparison of approaches for incorporating new information into existing risk prediction models". In: *Statistics in Medicine* 36.7 (Mar. 30, 2017), pp. 1134–1156. ISSN: 0277-6715, 1097-0258. DOI: 10.1002/sim.7190. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.7190 (visited on 11/17/2024).

[167] Ewout W. Steyerberg et al. "Validation and updating of predictive logistic regression models: a study on sample size and shrinkage". In: *Statistics in Medicine* 23.16 (Aug. 30, 2004), pp. 2567–2586. ISSN: 0277-6715, 1097-0258.

DOI: 10.1002/sim.1844. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.1844 (visited on 11/17/2024).

[168] Theresa Dankowski and Andreas Ziegler. "Calibrating random forests for probability estimation". In: *Statistics in Medicine* 35.22 (Sept. 30, 2016), pp. 3949–3960. ISSN: 1097-0258. DOI: 10.1002/sim.6959.

[169] Michael J. Pencina et al. "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond". In: *Statistics in Medicine* 27.2 (Jan. 30, 2008), pp. 157–172. ISSN: 0277-6715, 1097-0258. DOI: 10.1002/sim.2929. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.2929 (visited on 11/17/2024).

[170] Ewout W. Steyerberg and Frank E. Harrell. "Prediction models need appropriate internal, internal–external, and external validation". In: *Journal of Clinical Epidemiology* 69 (Jan. 2016), pp. 245–247. ISSN: 08954356. DOI: 10.1016/j.jclinepi.2015.04.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0895435615001754 (visited on 09/23/2023).

[171] Ewout W Steyerberg et al. "Internal validation of predictive models". In: *Journal of Clinical Epidemiology* 54.8 (Aug. 2001), pp. 774–781. ISSN: 08954356. DOI: 10.1016/S0895-4356(01)00341-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0895435601003419 (visited on 09/23/2023).

[172] Ewout W Steyerberg et al. "Internal and external validation of predictive models: A simulation study of bias and precision in small samples". In: *Journal of Clinical Epidemiology* 56.5 (May 2003), pp. 441–447. ISSN: 08954356. DOI: 10.1016/S0895-4356(03)00047-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0895435603000477 (visited on 09/23/2023).

[173] Jesper Kers et al. "Prediction models for delayed graft function: external validation on The Dutch Prospective Renal Transplantation Registry". In: *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association* 33.7 (July 1, 2018), pp. 1259–1268. ISSN: 1460-2385. DOI: 10.1093/ndt/gfy019.

[174] Huanxi Zhang et al. "Evaluation of predictive models for delayed graft function of deceased kidney transplantation". In: *Oncotarget* 9.2 (Jan. 5, 2018), pp. 1735–1744. ISSN: 1949-2553. DOI: 10.18632/oncotarget.22711.

[175] Jonathan A. Grossberg et al. "Utility of a Mathematical Nomogram to Predict Delayed Graft Function: A Single-Center Experience:" in: *Transplantation* 81.2 (Jan. 2006), pp. 155–159. ISSN: 0041-1337. DOI: 10.1097/01.tp.0000188621.54448.c8. URL: http://journals.lww.com/00007890-200601270-00003 (visited on 08/22/2024).

[176] Mohammed O Kaisar and David W Johnson. "Validation of a nomogram for predicting the likelihood of delayed graft function in Australian adult deceased donor renal transplant recipients". In: *Nephrology* 11.1 (Feb. 2006), pp. 78–79. ISSN: 1320-5358, 1440-1797. DOI: 10.1111/j.1440-1797.2006.00532.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1440-1797.2006.00532.x (visited on 08/22/2024).

[177] Magda Michalak et al. "Prediction of delayed graft function using different scoring algorithms: A single-center experience". In: *World Journal of Transplantation* 7.5 (Oct. 24, 2017), pp. 260–268. ISSN: 2220-3230. DOI: 10.5500/wjt.v7.i5.260. URL: http://www.wjgnet.com/2220-3230/full/v7/i5/260.htm (visited on 08/22/2024).

[178] Giuseppe Ietto et al. "Predictive Models for the Functional Recovery of Transplanted Kidney". In: *Transplantation Proceedings* 53.10 (Dec. 2021), pp. 2873–2878. ISSN: 1873-2623. DOI: 10.1016/j.transproceed.2021.08.053.

[179] E. Rodrigo et al. "Prediction of Delayed Graft Function by Means of a Novel Web-Based Calculator: A Single-Center Experience". In: *American Journal of Transplantation* 12.1 (Jan. 2012), pp. 240–244. ISSN: 16006135. DOI: 10.1111/j.1600-6143.2011.03810.x. URL: https://linkinghub.elsevier.com/retrieve/pii/S1600613522272852 (visited on 08/22/2024).

[180] Jeroen Hoogland et al. "Handling missing predictor values when validating and applying a prediction model to new patients". In: *Statistics in Medicine* 39.25 (Nov. 10, 2020), pp. 3591–3607. ISSN: 0277-6715, 1097-0258. DOI: 10.1002/sim.8682. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.8682 (visited on 09/25/2023).

[181] Rémi Kaboré et al. "Risk prediction models for graft failure in kidney transplantation: a systematic review". In: *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association* 32 (suppl_2 Apr. 1, 2017), pp. ii68–ii76. ISSN: 1460-2385. DOI: 10.1093/ndt/gfw405.

[182] Farid Foroutan et al. "Risk Factors for 1-Year Graft Loss After Kidney Transplantation: Systematic Review and Meta-Analysis". In: *Clinical journal of the American Society of Nephrology: CJASN* 14.11 (Nov. 7, 2019), pp. 1642–1650. ISSN: 1555-905X. DOI: 10.2215/CJN.05560519.

[183] Fateme Moghbeli, Majid Jangi, and Zahra Ebnehoseini. "Scoring systems of kidney donation from deceased donors: A systematic review". In: *Journal of Education and Health Promotion* 10 (2021), p. 456. ISSN: 2277-9531. DOI: 10.4103/jehp.jehp_1657_20.

[184] Stephanie Riley et al. "Using Information Available at the Time of Donor Offer to Predict Kidney Transplant Survival Outcomes: A Systematic Review of Prediction Models". In: *Transplant International: Official Journal of the European Society for Organ Transplantation* 35 (2022), p. 10397. ISSN: 1432-2277. DOI: 10.3389/ti.2022.10397.

[185] Peter Peduzzi et al. "A simulation study of the number of events per variable in logistic regression analysis". In: *Journal of Clinical Epidemiology* 49.12 (Dec. 1996), pp. 1373–1379. ISSN: 08954356. DOI: 10.1016/S0895-4356(96)00236-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0895435696002363 (visited on 09/07/2023).

[186] Florian G. Scurt et al. "Histological and clinical evaluation of discarded kidneys in a European cohort of deceased brain death donor kidneys of marginal quality". In: *Journal of Nephrology* (Oct. 19, 2023). ISSN: 1724-6059. DOI: 10.1007/s40620-023-01785-8. URL: https://link.springer.com/10.1007/s40620-023-01785-8 (visited on 10/23/2023).

[187] Fox J. Weisberg S. Freindly M. Hong J. Andersen R. Firth D. et al. *CRAN-package effects, 2022-07-13, version 4.2-2*. https://cran.r-project.org/web/packages/effects/index.html. Accessed: 01.04.2024. 2022.

[188] Chris Sutherland et al. "Practical advice on variable selection and reporting using Akaike information criterion". In: *Proceedings. Biological Sciences* 290.2007 (Sept. 27, 2023), p. 20231261. ISSN: 1471-2954. DOI: 10.1098/rspb.2023.1261.

[189] Hassan Awada et al. "A Focus on Intermediate-Risk Acute Myeloid Leukemia: Sub-Classification Updates and Therapeutic Challenges". In: *Cancers* 14.17 (Aug. 28, 2022), p. 4166. ISSN: 2072-6694. DOI: 10.3390/cancers14174166.

[190] D. S. Cross et al. "Coronary risk assessment among intermediate risk patients using a clinical and biomarker based algorithm developed and validated in two population cohorts". In: *Current Medical Research and Opinion* 28.11 (Nov. 2012), pp. 1819–1830. ISSN: 1473-4877. DOI: 10.1185/03007995.2012.742878.

[191] P. Cornford D. Tilki R.C.N. van den Bergh E. Briers D. Eberli G. De Meerleer et al. *EAU - EANM - ESTRO - ESUR - ISUP - SIOG Guidelines on Prostate Cancer*. https://uroweb.org/guidelines/prostate-cancer/chapter/diagnostic-evaluation. Accessed: 01.11.2024. 2024.

[192] Swarnalata Gowrishankar. "Banff Classification from 1991 to 2019. A Significant Contribution to Our Understanding and Reporting of Allograft Renal Biopsies". In: *Indian Journal of Nephrology* 32.1 (2022), pp. 1–7. ISSN: 0971-4065. DOI: 10.4103/ijn.IJN_270_20.

[193] H. Liapis et al. "Banff Histopathological Consensus Criteria for Preimplantation Kidney Biopsies". In: *American Journal of Transplantation* 17.1 (Jan. 2017), pp. 140–150. ISSN: 16006135. DOI: 10.1111/ajt.13929. URL: https://linkinghub.elsevier.com/retrieve/pii/S1600613522248163 (visited on 09/07/2023).

[194] Peter N. Furness et al. "International variation in histologic grading is large, and persistent feedback does not improve reproducibility". In: *The American Journal of Surgical Pathology* 27.6 (June 2003), pp. 805–810. ISSN: 0147-5185. DOI: 10.1097/00000478-200306000-00012.

[195] John O.O. Ayorinde et al. "Measurement Matters: A Metrological Approach to Renal Preimplantation Biopsy Evaluation to Address Uncertainty in Organ Selection". In: *Transplantation Direct* 10.11 (Oct. 10, 2024), e1708. ISSN: 2373-8731. DOI: 10.1097/TXD.0000000000001708. URL: https://journals.lww.com/10.1097/TXD.0000000000001708 (visited on 11/11/2024).

[196] Alton B. Farris et al. "Banff Digital Pathology Working Group: Image Bank, Artificial Intelligence Algorithm, and Challenge Trial Developments". In: *Transplant International* 36 (Oct. 16, 2023), p. 11783. ISSN: 1432-2277. DOI: 10.3389/ti.2023.11783. URL: https://www.frontierspartnerships.org/articles/10.3389/ti.2023.11783/full (visited on 11/11/2024).

[197] Jan U. Becker et al. "Artificial intelligence and machine learning in nephropathology". In: *Kidney International* 98.1 (July 2020), pp. 65–75. ISSN: 00852538. DOI: 10.1016/j.kint.2020.02.027. URL: https:

//linkinghub.elsevier.com/retrieve/pii/S0085253820303422 (visited on 03/02/2025).

[198] Elsaline Rijkse, Jan Nm IJzermans, and Robert C Minnee. "Machine perfusion in abdominal organ transplantation: Current use in the Netherlands". In: *World Journal of Transplantation* 10.1 (Jan. 18, 2020), pp. 15–28. ISSN: 2220-3230. DOI: 10.5500/wjt.v10.i1.15. URL: https://www.wjgnet.com/2220-3230/full/v10/i1/15.htm (visited on 03/02/2025).

[199] Cyril Moers et al. "Machine Perfusion or Cold Storage in Deceased-Donor Kidney Transplantation". In: *New England Journal of Medicine* 360.1 (Jan. 2009), pp. 7–19. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa0802289. URL: http://www.nejm.org/doi/abs/10.1056/NEJMoa0802289 (visited on 10/31/2024).

[200] Samuel J Tingle et al. "Normothermic and hypothermic machine perfusion preservation versus static cold storage for deceased donor kidney transplantation". In: *Cochrane Database of Systematic Reviews* 2024.7 (July 9, 2024). Ed. by Cochrane Kidney and Transplant Group. ISSN: 14651858. DOI: 10.1002/14651858.CD011671.pub3. URL: http://doi.wiley.com/10.1002/14651858.CD011671.pub3 (visited on 10/31/2024).

[201] Azita Mellati et al. "Kidney Normothermic Machine Perfusion Can Be Used as a Preservation Technique and a Model of Reperfusion to Deliver Novel Therapies and Assess Inflammation and Immune Activation". In: *Frontiers in Immunology* 13 (June 1, 2022), p. 850271. ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.850271. URL: https://www.frontiersin.org/articles/10.3389/fimmu.2022.850271/full (visited on 10/31/2024).

[202] Tim L. Hamelink et al. "Renal Normothermic Machine Perfusion: The Road Toward Clinical Implementation of a Promising Pretransplant Organ Assessment Tool". In: *Transplantation* 106.2 (Feb. 2022), pp. 268–279. ISSN: 0041-1337. DOI: 10.1097/TP.0000000000003817. URL: https://journals.lww.com/10.1097/TP.0000000000003817 (visited on 11/01/2024).

[203] International Summit on Transplant Tourism and Organ Trafficking. "The Declaration of Istanbul on Organ Trafficking and Transplant Tourism". In: *Clinical journal of the American Society of Nephrology: CJASN* 3.5 (Sept. 2008), pp. 1227–1231. ISSN: 1555-905X. DOI: 10.2215/CJN.03320708.

[204] Jordan Potter. "Imminent Death Donation: Ethical and Practical Policy Considerations". In: *Journal of Law, Medicine & Ethics* 46.2 (2018), pp. 524–537. ISSN: 1073-1105, 1748-720X. DOI: 10.1177/1073110518782961.

URL: https://www.cambridge.org/core/product/identifier/S1073110500025560/type/journal_article (visited on 10/31/2024).

[205] Mar Lomero et al. "Donation after circulatory death today: an updated overview of the European landscape". In: *Transplant International* 33.1 (Jan. 2020), pp. 76–88. ISSN: 0934-0874, 1432-2277. DOI: 10.1111/tri.13506. URL: https://onlinelibrary.wiley.com/doi/10.1111/tri.13506 (visited on 10/31/2024).

[206] Maarten G.J. Snoeijs et al. "Controlled donation after cardiac death: a European perspective". In: *Transplantation Reviews* 21.4 (Oct. 2007), pp. 219–229. ISSN: 0955470X. DOI: 10.1016/j.trre.2007.07.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0955470X07000596 (visited on 10/31/2024).

[207] Jentina Wind et al. "Variability in protocols on donation after circulatory death in Europe". In: *Critical Care* 17.5 (Oct. 3, 2013), R217. ISSN: 1364-8535. DOI: 10.1186/cc13034. URL: https://ccforum.biomedcentral.com/articles/10.1186/cc13034 (visited on 10/31/2024).

[208] Geoffrey Stewart Morrison. "Tutorial on logistic-regression calibration and fusion:converting a score to a likelihood ratio". In: *Australian Journal of Forensic Sciences* 45.2 (June 2013), pp. 173–197. ISSN: 0045-0618, 1834-562X. DOI: 10.1080/00450618.2012.733025. URL: http://www.tandfonline.com/doi/abs/10.1080/00450618.2012.733025 (visited on 11/01/2024).

[209] Daniel A. Solomon et al. "Accepting a donor kidney: an evaluation of patients' and transplant surgeons' priorities". In: *Clinical Transplantation* 25.5 (Sept. 2011), pp. 786–793. ISSN: 0902-0063, 1399-0012. DOI: 10.1111/j.1399-0012.2010.01342.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1399-0012.2010.01342.x (visited on 11/01/2024).

[210] Rachel Dzieran et al. "System-of-Systems Approach for Improving Quality of Kidney Transplant Decision-Making Support for Transplant Surgeons". In: *The Proceedings of the 2024 Conference on Systems Engineering Research*. Ed. by Alejandro Salado et al. Cham: Springer Nature Switzerland, 2024, pp. 327–340. ISBN: 9783031625534 9783031625541. DOI: 10.1007/978-3-031-62554-1_20. URL: https://link.springer.com/10.1007/978-3-031-62554-1_20 (visited on 11/01/2024).

[211] Gordon K. C. Chen. "What Is the Systems Approach?" In: *Interfaces* 6.1 (1975), pp. 32–37. ISSN: 00922102, 1526551X. URL: http://www.jstor.org/stable/25059250 (visited on 11/01/2024).

[212] Michelle J. Irving et al. "Nephrologists' Perspectives on the Effect of Guidelines on Clinical Practice: A Semistructured Interview Study". In: *American Journal of Kidney Diseases* 55.2 (Feb. 2010), pp. 241–249. ISSN: 02726386. DOI: 10.1053/j.ajkd.2009.09.032. URL: https://linkinghub.elsevier.com/retrieve/pii/S0272638609013869 (visited on 11/14/2024).

[213] Australia Caring for AustraliansWith Renal Impairment (CARI) Guidelines Adelaide. https://www.cariguidelines.org/guidelines/. Accessed: 01.04.2024. 2024.

[214] Karolina Schantz et al. "The Role of Procurement Biopsies in Kidney Acceptance Decision Making and Kidney Discard: Perceptions of Physicians, Nurse Coordinators, and OPO Staff and Directors". In: *Transplantation Direct* 8.4 (Mar. 10, 2022), e1299. ISSN: 2373-8731. DOI: 10.1097/TXD.0000000000001299. URL: https://journals.lww.com/10.1097/TXD.0000000000001299 (visited on 11/01/2024).

[215] EDQM. *Guide to the quality and safety of organs for transplantation*. ERA. URL: https://www.era-online.org/publications/ (visited on 09/06/2023).

[216] Sherry G. Mansour et al. "Reliability of deceased-donor procurement kidney biopsy images uploaded in United Network for Organ Sharing". In: *Clinical Transplantation* 32.12 (Dec. 2018), e13441. ISSN: 0902-0063, 1399-0012. DOI: 10.1111/ctr.13441. URL: https://onlinelibrary.wiley.com/doi/10.1111/ctr.13441 (visited on 11/01/2024).

[217] Darren Stewart et al. "Can Behavioral Research Improve Transplant Decision-Making? A Mock Offer Study on the Role of Kidney Procurement Biopsies". In: *Kidney360* 1.1 (Jan. 2020), pp. 36–47. ISSN: 2641-7650. DOI: 10.34067/KID.0000212019. URL: https://journals.lww.com/10.34067/KID.0000212019 (visited on 11/01/2024).

[218] Krista L. Lentine et al. "Deceased Donor Procurement Biopsy Practices, Interpretation, and Histology-Based Decision-Making: A Survey of US Kidney Transplant Centers". In: *Kidney International Reports* 7.6 (June 2022), pp. 1268–1277. ISSN: 24680249. DOI: 10.1016/j.ekir.2022.03.021. URL: https://linkinghub.elsevier.com/retrieve/pii/S2468024922012414 (visited on 11/01/2024).

[219] UNOS Policy Department Kelley Poff. *OPTN Public Comment Proposal, Refit Kidney Donor Profile Index without Race and Hepatitis C Virus*. https://hrsa.unos.org/media/ekkfxy4t/mac$_k dpiracehcv_p cjan24.pdf$. Accessed: 01.11.2024. 2024.

[220] Olivier Aubert et al. "Disparities in Acceptance of Deceased Donor Kidneys Between the United States and France and Estimated Effects of Increased US Acceptance". In: *JAMA Internal Medicine* 179.10 (Oct. 1, 2019), p. 1365. ISSN: 2168-6106. DOI: 10.1001/jamainternmed.2019.2322. URL: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2748452 (visited on 11/01/2024).

[221] Antony Attia et al. "Implausible algorithm output in UK liver transplantation allocation scheme: importance of transparency". In: *The Lancet* 401.10380 (Mar. 2023), pp. 911–912. ISSN: 01406736. DOI: 10.1016/S0140-6736(23)00114-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673623001149 (visited on 11/18/2024).

[222] Associated Press Lauran Neergaard. *A racially biased test kept thousands of Black people from getting a kidney transplant*. https://www.pbs.org/newshour/nation/a-racially-biased-test-kept-thousands-of-black-people-from-getting-a-kidney-transplant. Accessed: 01.11.2024. 2024.

[223] Mohammadali M. Shoja et al. "Marie-François Xavier Bichat (1771–1802) and his contributions to the foundations of pathological anatomy and modern medicine". In: *Annals of Anatomy - Anatomischer Anzeiger* 190.5 (Nov. 2008), pp. 413–420. ISSN: 09409602. DOI: 10.1016/j.aanat.2008.07.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0940960208000976 (visited on 11/18/2024).

Chapter 7

Supplemental material

List of preliminary published results

Peer-reviewed articles

- **Ernst A**, Regele H, Chatzikyrkou C, Dendooven A, Turkevi-Nagy S, Tieken I, Oberbauer R, Reindl-Schwaighofer R, Abramowicz D, Hellemans R, Massart A, Ljubanovic DG, Senjug P, Maksimovic B, Aßfalg V, Neretljak I, Schleicher C, Clahsen-van Groningen M, Kojc N, Ellis CL, Kurschat CE, Lukomski L, Stippel D, Ströhlein M, Scurt FG, Roelofs JJ, Kers J, Harth A, Jungck C, Eccher A, Prütz I, Hellmich M, Vasuri F, Malvi D, Arns W, Becker JU. 2-Step-Scores with optional nephropathology for the prediction of adverse outcomes for brain-dead donor kidneys in Eurotransplant. Nephrol Dial Transplant. 2024 Apr 17:gfae093. doi: 10.1093/ndt/gfae093. Epub ahead of print. PMID: 38632055.

Congress presentations

- **5th International Renal Pathology Conference**, 18.05 - 20.05.2023, Zagreb (Croatia). Presentation by Dr. med. Jan Ulrich Becker: *Clinicopathological Scores with Optional Nephropathology for the Prognostication of Outcome for Deceased Donor Kidneys in Eurotransplant.*

- **15. Jahrestagung der Deutschen Gesellschaft für Nephrologie**, 05.10 - 08.10.2023, Berlin. Presentation by Angela Ernst: *Klinisch-pathologische Scores mit optionaler Nephropathologie für die Prognose der Ergebnisse für verstorbene Spendernieren in Eurotransplant.*

- **32. Jahrestagung der Deutschen Transplantationsgesellschaft**, 26.10 - 28.10.2023, Jena. Poster by Angela Ernst: *P01-02, Clinicopathological scores with optional nephropathology for the prognostication of outcome for deceased donor kidneys in Eurotransplant.*

Search strategy and PICOS: systematic literature review

For outcome = delayed graft function (DGF)
(((((("kidney"[MeSH Terms] OR "kidney"[All Fields] OR "renal"[All Fields]) AND "transplant*"[Title/Abstract]) OR (("transplants"[MeSH Terms] OR "transplantation"[MeSH Terms] OR "graft*"[Title/Abstract] OR "allograft*"[Title/Abstract] OR "allotransplant*"[Title/Abstract] OR "deceased don*"[Title/Abstract] OR "grafting"[All Fields]) AND "kidney*"[MeSH Terms])) AND 2000/01/01:3000/12/12[Date - Publication] AND "delayed graft function"[MeSH Terms] AND ((("predict*"[Title/Abstract] OR "prognos*"[Title/Abstract] OR "risk"[Title/Abstract] OR "probabilit*"[Title/Abstract]) AND ("tool*"[Title/Abstract] OR "calculat*"[Title/Abstract] OR "model*"[Title/Abstract] OR "scor*"[Title/Abstract])) OR "nomogram"[Title/Abstract])) NOT (("liver"[Title/Abstract] OR "heart"[Title/Abstract] OR "pancreas"[Title/Abstract] OR "lung"[Title/Abstract]) AND "transplant*"[Title/Abstract])) NOT ("living donors"[MeSH Terms] OR "genes"[Title/Abstract] OR "living donor*"[Title/Abstract] OR "pediatrics"[MeSH Terms] OR "living donation"[Title/Abstract] OR "child"[MeSH Terms])

For outcome = transplant loss
(((((("kidney"[MeSH Terms] OR "kidney"[All Fields] OR "renal"[All Fields]) AND "transplant*"[Title/Abstract]) OR (("transplants"[MeSH Terms] OR "transplantation"[MeSH Terms] OR "graft*"[Title/Abstract] OR "allograft*"[Title/Abstract] OR "allotransplant*"[Title/Abstract] OR "deceased don*"[Title/Abstract] OR "grafting"[All Fields]) AND "kidney*"[MeSH Terms])) AND ("graft survival"[MeSH Terms] OR (("graft"[Title/Abstract] OR "allograft"[Title/Abstract] OR "transplant"[Title/Abstract]) AND ("failure"[Title/Abstract] OR "loss"[Title/Abstract] OR "survival"[Title/Abstract] OR "outcome"[Title/Abstract])))AND ((("predict*"[Title/Abstract] OR "prognos*"[Title/Abstract] OR "risk"[Title/Abstract] OR "probabilit*"[Title/Abstract]) AND ("tool*"[Title/Abstract] OR "calculat*"[Title/Abstract] OR "model*"[Title/Abstract] OR "scor*"[Title/Abstract])) OR "nomogram"[Title/Abstract] OR "scoring system"[Title/Abstract]) AND 2000/01/01 : 3000/12/12[Date - Publication])NOT ("living donors"[MeSH Terms] OR "living donor*"[Title/Abstract] OR "living donation"[Title/Abstract] OR "child"[MeSH Terms] OR "children"[Title/Abstract] OR "pediatrics"[MeSH Terms] OR "gene"[Title/Abstract] OR "genes"[Title/Abstract] OR "animals"[MeSH Terms] OR "pediatr*"[Title/Abstract] OR "antibodies"[MeSH Terms] OR "treatment failure"[MeSH Terms])) NOT (("liver"[Title/Abstract] OR "heart"[Title/Abstract] OR "pancreas"[Title/Abstract] OR "lung"[Title/Abstract]) AND "transplant*"[Title/Abstract])

	Participants	Intervention	Comparison	Outcome
Terms	- kidney transplant recipients	- Kidney transplantation - transplantation of kidneys - grafting - renal transplantation	- Risk prediction - Risk score - Score - Scoring system - Model - Prognosis	- DGF - graft survival - graft failure - graft loss - transplant loss - Transplant survival - transplant failure
MeSH (Entry Terms)	"transplant recipients"[MeSH Terms] **OR** transplant recipient[Text Word]	"kidney transplantation"[MeSH Terms] **OR** renal transplantation[Text Word] **OR** renal transplantations[Text Word] **OR** kidney grafting[Text Word] **OR** Kidney Transplantations[Text Word] **OR** Transplant* ***AND*** Kidney	"prognosis"[MeSH Terms] **OR** Prognoses[Text Word] **OR** Prognostic Factors[Text Word] **OR** Prognostic Factor[Text Word] **OR** "clinical decision rules"[MeSH Terms] **OR** clinical decision rule[Text Word] "clinical decision rules"[MeSH Terms] **OR** Clinical Prediction Rule[Text Word] **OR** risk score [Text Word] **OR** "probability learning"[MeSH Terms] **OR** Probability Learnings [Text Word]	"delayed graft function"[MeSH Terms] **OR** "graft survival"[MeSH Terms] **OR** "graft rejection"[MeSH Terms] **OR** graft rejections[Text Word] **OR** transplant rejection[Text Word] **OR** transplantation rejections[Text Word] **OR** transplant rejections[Text Word] **OR** failure ***AND*** transplant* **OR** survival ***AND*** Transplant* **OR** survival ***AND*** graft **OR** los* ***AND*** transplant* **OR** Los* ***AND*** graft

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen - einschließlich Tabellen, Karten und Abbildungen -, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind in jedem Einzelfall als Entlehnung kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertationsschrift noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentli-chung vor Abschluss der Promotion nicht ohne Genehmigung der / des Vorsitzenden des IPHS-Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. rer. medic. Martin Hellmich (Erstbetreuer) und PD Dr. med. Jan Becker (Zweitbetreuer) betreut worden. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.
Übersicht der Publikationen (S. 151):

Ernst A, Regele H, Chatzikyrkou C, Dendooven A, Turkevi-Nagy S, Tieken I, Oberbauer R, Reindl-Schwaighofer R, Abramowicz D, Hellemans R, Massart A, Ljubanovic DG, Senjug P, Maksimovic B, Aßfalg V, Neretljak I, Schleicher C, Clahsen-van Groningen M, Kojc N, Ellis CL, Kurschat CE, Lukomski L, Stippel D, Ströhlein M, Scurt FG, Roelofs JJ, Kers J, Harth A, Jungck C, Eccher A, Prütz I, Hellmich M, Vasuri F, Malvi D, Arns W, Becker JU. 2-Step-Scores with optional nephropathology for the prediction of adverse outcomes for brain-dead donor kidneys in Eurotransplant. Nephrol Dial Transplant. 2024 Apr 17:gfae093. doi: 10.1093/ndt/gfae093. Epub ahead of print. PMID: 38632055

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Verän-derungen, dem IPHS-Promotionsausschuss unverzüglich mitzuteilen.

Ernst Angela
Digital unterschrieben von Ernst Angela
Datum: 2025.05.19 15:14:11 +02'00'

. . . . . .

Datum

. . . . . .

Unterschrift

Curriculum Vitae

Mein Lebenslauf wird aus Gründen des Datenschutzes in der elektronischen Fassung meiner Arbeit nicht veröffentlicht.