Predicting outcomes in patients with kidney disease: methodology and clinical applications

Ramspek, C.L.

9

Combining existing models and recent data into an up-to-date prediction model for evaluating kidneys from older deceased donors: development and external validation study

Chava L. Ramspek, Mostafa El Moumni, Eelaha Wali, Martin B. A. Heemskerk, Robert A. Pol, Meindert J. Crop, Nichon E. Jansen, Andries Hoitsma, Friedo W. Dekker, Merel van Diepen, Cyril Moers

Kidney Int. 2021;99(6):1459-69

Abstract

Background. With a rising demand for kidney transplantation, reliable pre-transplant assessment of organ quality becomes top priority. In clinical practice, physicians are regularly in doubt whether suboptimal kidney offers from older donors should be accepted. The aim of the current study was to externally validate existing prediction models in a European population of older deceased donors, as well as develop and externally validate an adverse outcome (AO) prediction tool.

Methods. Recipients of kidney grafts from deceased donors ≥50 years were included from the Dutch and United States organ transplant registry (NOTR and OPTN database) from 2006-2018. The predicted AO was a composite of graft failure, death or CKD stage 4+ within 1 year after transplantation, modelled using logistic regression. Discrimination and calibration were assessed in internal, temporal and external validation. Seven existing models were validated in the same cohorts.

Results. The NOTR development cohort contained 2510 patients and 823 events. The temporal validation NOTR had 837 patients and the external validation OPTN 31987 patients. Discrimination of our full AO model was moderate in external validation (C-statistics=0.63), though somewhat better than discrimination of the 7 existing prediction models (average C-statistic=0.57). The AO model's calibration was highly accurate.

Conclusion. Existing kidney graft survival models performed poorly in a population of older deceased donors. Novel AO models were developed and externally validated, with maximum achievable performance in a population of older deceased kidney donors. These models could assist transplant clinicians in deciding whether to accept a kidney from an older donor.

Introduction

Kidney transplantation is the treatment of choice for patients with end-stage renal disease, in terms of survival and quality of life.[1 2] With rising demand for kidney transplantation and the kidney donor pool lagging behind, the acceptance criteria for donor kidneys continue to expand.[3 4] Grafts recovered from sub-optimal donors, who are on average older with more comorbidities, come with higher rates of early graft dysfunction and recipient mortality.[5 6] The decision whether to accept or decline a kidney offer is largely subjective and depends on donor-, organ preservation- and recipient-related characteristics. Discard rates vary widely between individual physicians and across geographic areas.[7-9] Reliable pre-transplant assessment of organ quality and selection of the best recipient-to-donor match in order to minimize unjust discard and maximize graft and patient survival has thus become increasingly important.

Various regression-based mathematical models have been developed that aim to predict outcomes after kidney transplantation.[10] As reliably predicting the risk of post-transplant graft failure *prior* to transplantation has proven to be challenging, several models have included predictors measured during transplant surgery or shortly after transplantation, such as the iBox risk score.[11 12] Though these models might be useful for monitoring patients, they cannot be used to guide physicians to accept or decline a kidney offer. One of the most widely used models predicting graft survival (combined graft failure and mortality) is the kidney donor risk index (KDRI).[13] The Kidney Donor Profile Index (KDPI), derived from this KDRI, has been implemented in the new US kidney allocation system in effect since 2014.[14] Long-term consequences of this implementation are still unknown. Nevertheless, the KDPI has been criticized as delayed graft function rates have increased, the score is highly dependent on donor age and KDPI labelling may cause unjust and almost automatic discard of kidneys with a high KDPI.[15-18]

In most European transplant systems, allocation prediction models have yet to be implemented. Similar models to the KDRI have been developed, but the vast majority of these models has been constructed on transplant data from the United States.[19-21] As patient populations, kidney transplant procedures and policies differ considerably between Europe and the US, there is a need to develop and validate such prediction models on European patients. Furthermore, a prediction tool specifically tailored to older deceased donors, might allow for improved decision-making regarding the transplantation of these suboptimal grafts for which there is little consensus whether to accept or decline. Therefore, we have externally validated existing prediction models that can be used prior to transplantation and predict graft survival, in a European and Northern American population of kidney transplant recipients who received organs from deceased donors aged 50 years or older. Subsequently, our aim was to improve upon these existing prediction models by developing and externally validating new prediction models of adverse outcome (AO) within 1 year after kidney transplantation from older deceased donors.

Methods

This study was conducted in accordance with the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement.[22 23]

Existing models

To identify existing prediction models which were suitable for external validation, a systematic search was performed, of which the details are reported in the supplemental material. The identified models were validated on both the Dutch and US transplant registry for the combined outcome of graft failure and recipient death at 1 year post transplantation.

Dutch Transplantation cohort.

The Netherlands Organ Transplantation Registry (NOTR) prospectively collects data from all 8 transplant centres in the Netherlands and contains post-transplantation follow-up information at 3 months, 1 year and yearly thereafter. It is also linked to the Dutch national dialysis registry. Recipients of a single kidney transplant from a deceased donor ≥50 years were included. Recipients younger than 18 years were excluded as well as recipients of multiple-organ transplants. For the development and temporal validation of our AO models, the NOTR dataset was split based on transplant date. The AO model development took place on NOTR patients transplanted between 1 January 2006 and 31 December 2016, these models were temporally validated on patients transplanted between 1 January 2017 and November 2018. Follow-up data were available up to November 2019.

United States Transplantation cohort

Patients who received a solitary deceased donor kidney transplant in the US between 1 January 2006 and 1 January 2017 were included as validation cohort. Data from a UNOS Standard Transplant Analysis and Research File from the Organ Procurement and Transplantation Network (OPTN), as of 1 March 2018 were used. Deceased donors younger than 50 years old were excluded, as were recipients aged younger than 18 and recipients waitlisted for a multiple-organ transplant.

Selected predictors

A priori, a list of candidate predictors was compiled by the research group based on existing literature, identified prediction models and clinical experience.[10 24-27] This list was presented to an expert panel of 10 nephrologists working at 4 different transplant centres in the Netherlands. These nephrologists were asked to add any missing potential predictors and to rank the list of candidate predictors from most important to least important. A full model was developed using all predictors and two approaches were explored to shorten

this model to a more parsimonious one. Besides the full model, a data-driven model was developed using backward elimination. Finally, an expert model was developed based on the 14 expert top-ranked predictors. The experts were only involved in selecting the predictors, the regression coefficient (weight) given to these predictors was based on the data. Interaction terms between recipient and donor age, height and weight, and donation after circulatory death and cold ischemic time (CIT) were added based on clinical expertise and literature.[20 28 29]

Predicted outcome

For the newly developed AO models the predicted outcome was a combined endpoint including at least one of the following within one year after transplantation: graft failure, recipient death, or chronic kidney disease (CKD) stage 4+. This composite outcome was defined by an expert panel of nephrologists, transplant surgeons and epidemiologists. The estimated glomerular filtration rate (eGFR) was calculated using the CKD-EPI formula and CKD stage 4+ was defined as an unrecovered eGFR < 30 ml/min/1.73m^2.[30] In the NOTR creatinine was measured at the 1 year follow-up visit for all patients. In the OPTN creatinine is not recorded at a set time-point; therefore all serum creatinine registrations between 90 days and 1 year after transplantation were considered. For the validation of the existing prediction models, their original outcome was used. This was graft survival (graft failure and recipient death combined) for all included models. As the outcome between our AO models and the existing models differs, we performed a sensitivity analysis in which we changed the predicted outcome to graft survival.

Statistical analysis

Baseline characteristics are presented as percentages, means with standard deviation or median with interquartile range. Missing data was assumed to be largely missing at random and a 10-fold multiple imputation, including all predictors and outcome, was performed.[31 32]

Validation of existing models

The existing models were validated by calculating a risk score based on the reported regression coefficients per predictor for each included donor-recipient pair. As these models were Cox models, they were validated as such and it was assessed how well the risk scores (prognostic index) corresponded to the observed time till graft failure/recipient death within 1 year. Discrimination was calculated by Harrel's C-statistic. The C-statistic is a relative measure and examines if patients with the outcome had a higher risk score than patients without the outcome. A C-statistic of 0.5 is equivalent to chance and 1 is perfect discrimination.[33 34] Most reports on existing models did not publish the full model formula, meaning that it was not possible to calculate individuals' probabilities or assess calibration. Calibration is the agreement between the absolute predicted risk and observed risk.[35] Therefore, we recalibrated all existing models in a conservative manner; by updating

the baseline hazard of the outcome (updating results and model formulas are given in the supplement). Additionally, the two models for which full formulas were available, were validated without updating.

AO model development and validation

To develop the AO full model, all candidate predictors and interaction terms were entered in a logistic regression model.[33] Non-linear continuous predictors were modelled using restricted cubic splines[36] For the data-driven models we used a backward elimination procedure with P-value <0.157 as stopping criterion.[37] For the AO expert model the top-ranked predictors were entered in a logistic regression model. The three developed AO models were first internally validated. This internal validation was done by a 250-fold bootstrapping analysis, as recommended by the TRIPOD guidelines. Based on the bootstrapped results the models were adjusted for overfitting by multiplying each coefficient by a shrinkage factor (the bootstrapped slope).[38] These optimism-corrected models were subsequently validated in the Dutch temporal and US external validation cohort. Temporal validation can be seen as midway between internal and external validation, the patients are from the same region and included in the same manner but do not overlap with the development population. Additionally, certain practices may change over time, which differentiates the temporal validation cohort from the development cohort. As the incidence of adverse outcome differs between the Netherlands and the US, the models were conservatively recalibrated for the US outcome incidence, by adding a correction factor to the model formula.[39] This improves calibration but does not affect discrimination. Discrimination was assessed in the development, internal validation, temporal validation and external validation cohort by calculating the C-statistic. Calibration was assessed by plotting the predicted risks against the observed risks in calibration plots. Additionally, the calibration in-the-large, which is the average predicted risk in the entire population compared to the proportion of patients who actually experience the outcome, was computed. Finally, the calibration slope and intercept were calculated by fitting the prognostic index in a new regression model in the validation cohorts. In development the calibration slope is 1 per definition, a slope smaller than 1 indicates the predicted risks are too extreme, which is generally seen in overfitted models. As a sensitivity analysis, the AO models were validated in the US population without recalibration. The full model formula of all final models is given in the supplemental material. Model performance measures and coefficients were pooled over the ten imputation datasets according to Rubin's Rules.[40] All analyses were performed in R version 3.6.1.

Results

Existing prediction models

Following the systematic screening, 6 studies, presenting 7 prediction models, were considered appropriate for validation (flowchart in Figure S1). Characteristics of the validated models are shown in Table 1. All models but one were developed on US transplant data and showed similar C-statistics of around 0.63 in previous internal validation. Most existing models had a high risk of bias when assessed with the PROBAST tool (see Table S1).[41] Only two of the models provided the full model formula. Included predictors vary considerably per study and donor age is the only predictor included in each model (see Table 2). The majority of existing models only included donor characteristics.

Baseline characteristics

In total, 3333 transplant recipients of kidneys from deceased donors ≥50 years were included from the Netherlands Organ Transplant Registry (NOTR). These patients were split into a development cohort (2510 patients) and temporal validation cohort (837 patients). From the US transplant registry (OPTN) 31987 recipients were included as external validation cohort. At baseline the OPTN dataset had slightly younger donors with more diabetes and hypertension and substantially fewer donations after circulatory death (DCD) (see Table 3). More extensive baseline tables including percentage of missing data and stratified by outcome are given in the supplemental material (Table S2-S4). In the NOTR development cohort a total of 10.2% (n=257) experienced graft failure, 6.9% (n=172) death and 17.8% (n=446) CKD stage 4+ within 1 year, 9 patients (<1%) were lost to follow up. In the NOTR temporal validation cohort 8% (n=67) experienced graft failure, 3.6% (n=30) death, 17.4% (n=146) CKD stage 4+ and 4% (n=35) were lost to follow up. In the OPTN validation cohort 6.2% (n=1992) experienced graft failure, 5,3% (n=1711) death and 12.8% (n=4094) CKD stage 4+. In total, 200 patients (<1%) were lost to follow up. For the AO models, patients lost to follow up were assumed not to have experienced the outcome.

Table 1: characteristics of externally validated models. Each of these models was developed for use at the time of kidney allocation. The predicted outcome was combined graft failure and recipient death for each of these models.

Prediction model	Time horizon	Outcome	Development cohort	Population	Mean donor age	Reported C-statistic	Overall risk of bias	Model
Donor risk score by Schold et al. (2005)[45]	-	Death/ GF	US population 1996-2020 (OPTN: SRTR data)	First-time, single kidney only, adult recipients. Deceased donors.	-	-	High	Regression coefficients given (unclear what statistical model was used)
Rao et al. KDRI (2009)[13], full and donor-only model	-	Death/ GF	US population 1995-2005 (OPTN: SRTR data)	First-time, kidney only, adult recipients. Deceased donors.	-	0.62 (IV)	High	Cox model, hazard ratios given.
Kasiske et al. pretransplant model (2010)[19]	5 years	Death/ GF	US population 2000-2006 (OPTN: USRDS data)	Single kidney only, adult recipients. Deceased donor	38	0.64 (IV)	High	Cox model, full model formula given.
Watson et al. UKKDRI (2012)[46]	9 years	Death/ GF	European population 2000-2007 (UK Transplant Registry)	Adult kidney transplant recipients. Adult deceased donors.	49[a]	0.62 (IV)	High	Cox model, hazard ratios given.
Molnar et al. (2017)[21]	5 years	Death/ GF	US population 2001-2006 (OPTN[b]: SRTR data)	First-time, adult recipients on dialysis. Deceased donors.	39	0.63 (IV)	Low	Cox model, full model formula given
Vinson et al. model 3 (2018)[20]	-	Death/ GF	US population 2000-2014 (OPTN: SRTR data)	Single kidney only, adult recipients. Deceased donors.	39	0.63 (IV)	High	Cox model, only HRs given.

-: unknown/not reported [a]median age [b]OPTN data linked to dialysis facility data

Abbreviations: DDS: deceased donor score, GF: graft failure, OPTN: organ procurement and transplantation network, SRTR: scientific registry of transplant recipients, this dataset OPTN data which is supplemented with data from various secondary sources. USRDS: US Renal Data system, OPTN data supplemented with data from centers for Medicare & Medicaid services, HRs: hazard ratios, IV: internal validation

Table 2: Final predictors in developed and validated models

Predictors	AO full model	AO data-driven	AO expert model	Schold model	KDRI Full	KDRI donor-only	Kasiske model	UKKDRI	Molnar model	Vinson model
DONOR CHARACTERISTICS										
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BMI	✓	✓								
Cause of death	✓	✓	✓	✓	✓	✓	✓			
Cold ischemic time	✓		✓	✓	✓					✓
CPR performed	✓									
Days in hospital								✓		
DCD * CIT	✓	✓	✓							
Diabetes mellitus	✓		✓	✓	✓	✓			✓	
Donor after cardiac death	✓	✓	✓		✓	✓				
Double Tx					✓					
ECD									✓	
En bloc Tx					✓					
Ethnicity				✓	✓	✓				
HCV status					✓	✓				✓
Height					✓	✓				
Hypertension history	✓		✓	✓	✓	✓	✓	✓		✓
Hypotension	✓		✓							✓
Inotropes use	✓	✓						✓		
Last serum creatinine	✓	✓	✓		✓	✓				
Left/right kidney	✓									
Proteinuria	✓									
Sex	✓									
Smoking	✓									
Warm ischemic time	✓	✓	✓							
Weight					✓	✓		✓		
RECIPIENT CHARACTERISTICS										
Age	✓	✓	✓				✓		✓	✓
Blood hemoglobin									✓	
BMI	✓									
Cardiovascular disease	✓	✓								
Coronary artery disease									✓	✓
Diabetes mellitus	✓		✓						✓	✓
Dialysis duration	✓	✓	✓				✓		✓	✓
Ethnicity							✓		✓	
HCV status							✓			

Table 2: Continued

Predictors	**AO full model**	**AO data-driven**	**AO expert model**	**Schold model**	**KDRI** Full	**KDRI** donor-only	**Kasiske model**	**UKKDRI**	**Molnar model**	**Vinson model**
Medical insurance							✓		✓	
N previous kidney Tx	✓		✓				✓			
Peripheral vascular disease										✓
Primary kidney disease	✓						✓		✓	✓
Serum albumin									✓	
Sex	✓	✓								
DONOR-RECIPIENT										
Donor age * Recipient age	✓	✓	✓				✓			
D-R CMV match				✓						
D-R ethnicity difference										✓
D-R height difference	✓									✓
D-R weight difference	✓	✓								✓
HLA mismatches	✓	✓	✓	✓	✓		✓		✓	✓
Peak PRA	✓									✓

Abbreviations: BMI: body mass index, CPR: cardiopulmonary resuscitation, Tx: transplantation, HCV: hepatitis C virus, DCD: donor after cardiac death, ECD: expanded criteria donor, CIT: cold ischemic time, N: number, HLA: human leukocyte antigen, PRA: panel-reactive antibody. D-R: donor-recipient, CMV: cytomegalovirus

Table 3: Baseline characteristics stratified by cohort.

	Development cohort (NOTR 2006-2017) n = 2510	Temporal validation cohort (NOTR 2017-2018) n = 837	External validation cohort (OPTN 2006-2017) n=31987
Donor characteristics			
Age (years)	60 (55-65)	61 (55-66)	56 (53-60)
Sex (% male)	51.4%	56.2%	53.5%
Cause of death (%)			
Trauma	14.4%	16.9%	21.1%
Cerebrovascular accident	64.4%	56.2%	56.5%
Anoxia	18.1%	24.6%	20.0%
Other	3.1%	2.2%	2.4%
DCD donor (%)	46.1%	58.8%	13.8%
Serum creatinine (µmol/L)	66 (53-83)	64 (52-82)	80 (62-106)
Proteinuria (%)	44.4%	49.4%	41.9%
BMI (kg/m^2)	26 (4.7)	26 (4.4)	29 (6.4)
History of diabetes mellitus (%)	8.1%	9.3%	12.7%
History of hypertension (%)	37.5%	38.1%	50.6%
Hypotension (%)	31.5%	21.9%	-
Use of inotropic medication (%)	71.7%	69.9%	51.8%
Left kidney (%)	50.4%	49.8%	49.6%
WIT in DCD donors (minutes)	17 (14-21)	15 (13-18)	18 (11-27)
Cold ischemic time (hours)	15.8 (5.8)	13.3 (5.7)	18.2 (9.2)
Recipient characteristics			
Age (years)	60 (49-67)	62 (51-69)	60 (51-66)
Sex (% male)	60.6%	63.6%	62.2%
BMI (kg/m^2)	26 (4.7)	27 (4.4)	27 (4.8)
Primary kidney disease (%)			
Diabetes mellitus	14.0%	18.3%	32.6%
Hypertension	20.6%	22.4%	25.9%
Glomerular nephritis	16.6%	17.8%	11.4%
Cystic kidney disease	14.7%	9.9%	7.7%
Other	34.2%	31.5%	22.4%
Diabetes mellitus (%)	21.5%	26.8%	43.0%
Time on dialysis (months)	39 (25-57)	25 (15-42)	40 (13-66)
≥1 previous kidney transplant (%)	12.9%	15.2%	8.9%
Donor-recipient characteristics			
Total number of HLA mismatches	3 (2-4)	3 (2-4)	5 (4-5)
Peak PRA	0 (0-0)	0 (0-0)	0 (0-13)

Abbreviations: BMI:body mass index; PRA:panel reactive antibody; HLA:human leukocyte antigen. Lab values are shown in SI units and can be converted to conventional units as follows, serum creatinine in mg/dL: multiply by 0.011

Validation results of existing models

In total, 7 existing prediction models were validated. All these models predicted graft survival (graft failure and recipient death combined) and were therefore validated for this outcome. In the Dutch study population of 50+ donors (NOTR), predictive performance ranged from poor to mediocre. The C-statistics ranged from 0.538 (UKKDRI) to 0.611 (Vinson model), and the average C-statistic was 0.565 (see Table 4). The models' discrimination was slightly better in the OPTN data of 50+ donor kidneys (average C-statistic: 0.587), which is unsurprising considering that most models were developed on OPTN data. Overall, the best discrimination was seen for the most recent model by Vinson et al.[20] Models were conservatively updated in order to assess the calibration (calibration results shown in Table S6, Figure S2-S5), which was generally reasonable. The best calibration was seen for the Schold model and KDRI's.

Table 4: external validation results: Harrel's C-statistics for 1year risk of the combined endpoint: graft failure and recipient death.

	NOTR 2006-2017 C-statistic (95% CI)	**NOTR 2017-2018** C-statistic (95% CI)	**OPTN 2006-2017** C-statistic (95% CI)
Schold	0.562 (0.532-0.591)	0.555 (0.495-0.615)	0.577 (0.567-0.586)
KDRI$_{\text{full model}}$ (Rao)	0.572 (0.542-0.601)	0.560 (0.495-0.625)	0.592 (0.582-0.601)
KDRI$_{\text{donor-only}}$ (Rao)	0.571 (0.541-0.600)	0.559 (0.495-0.623)	0.590 (0.581-0.600)
Kasiske	0.584 (0.556-0.612)	0.547 (0.484-0.610)	0.609 (0.599-0.618)
UKKDRI (Watson)	0.544 (0.515-0.574)	0.538 (0.473-0.603)	0.552 (0.542-0.562)
Molnar	0.566 (0.537-0.596)	0.575 (0.515-0.636)	0.578 (0.569-0.588)
Vinson	0.598 (0.569-0.626)	0.573 (0.510-0.636)	0.611 (0.601-0.620)

CI: confidence interval

AO models

For the newly developed AO models the predicted outcome was a combined endpoint including at least one of the following within one year after transplantation: graft failure, recipient death, or chronic kidney disease (CKD) stage 4+. In the AO full model all candidate predictors, pre-defined by the research team as well as additionally suggested predictors from a nephrologist panel, were included. This resulted in a model with 28 predictors. In the AO data-driven model, logistic regression with backward selection resulted in the inclusion of 14 predictors. In the AO expert model, the 14 expert top-ranked predictors were included. The predictors included per AO model are shown in Table 2. The ranking results from the expert panel of 10 nephrologists is shown in Table S5. In general, there was a lot of variation in ranking between individual nephrologists, though all agreed that donor age and donor serum creatinine were the most important predictors.

Discrimination of our AO models was moderate, but nevertheless substantially better than for the existing models. The C-statistics of the AO full model were 0.635 and 0.630 in temporal and external validation, respectively. The AO data-driven model showed similar C-statistics of 0.628 and 0.624, and the AO expert model had slightly lower C-statistics of 0.609 and 0.619 (see Table 5). Calibration was generally good, although the models tended to over-predict in higher risk patients (see Table 5 and Figure 1). Without recalibration the AO models generally over-predicted risks in the OPTN dataset (supplement, Table S8 & Figure S6). In a sensitivity analyses we also built a model that predicts the more conventional graft survival outcome. The performance of this model was poorer than of the AO models, though slightly better than that of most existing models (data not shown).

Table 5: AO models: development, internal and external validation model performance results.

		Development NOTR 2006-2017	**Internal validation**	**Temporal validation NOTR 2017-2018**	**External validation[b] OPTN 2006-2017**
AO full model	C-statistic (95% CI)	0.680 (0.657-0.703)	0.646[a]	0.635 (0.593-0.678)	0.630 (0.622-0.637)
	Calibration slope	1	0.809	0.885	0.739
	Calibration intercept	0	-0.125	-0.319	-0.366
	Calibration-in-the-large[c]	32.8% vs. 32.8%	32.9% vs 32.8%	32.2% vs 27.5%	21.9% vs 21.1%
AO data-driven model	C-statistic	0.667 (0.644-0.690)	0.637[a]	0.628 (0.586-0.669)	0.624 (0.617-0.631)
	Calibration slope	1	0.813	0.909	0.796
	Calibration intercept	0	-0.122	-0.284	-0.286
	Calibration-in-the-large[c]	32.8% vs. 32.8%	33.0% vs. 32.8%	31.8% vs. 27.5%	21.7% vs 21.1%
AO expert model	C-statistic	0.658 (0.634-0.682)	0.638[a]	0.609 (0.566-0.653)	0.619 (0.612-0.627)
	Calibration slope	1	0.869	0.776	0.761
	Calibration intercept	0	-0.087	-0.391	-0.327
	Calibration-in-the-large[c]	32.8% vs. 32.8%	32.9% vs 32.8%	32.2% vs 27.5%	21.7% vs 21.1%

[a]no confidence intervals computed as it concerns a bootstrap shrinkage corrected C-statistic. [b]correction factor was added to the model, to recalibrate to the US outcome incidence. The results without recalibration are shown in the supplemental material. [c]Calibration-in-the-large is given as predicted versus observed.

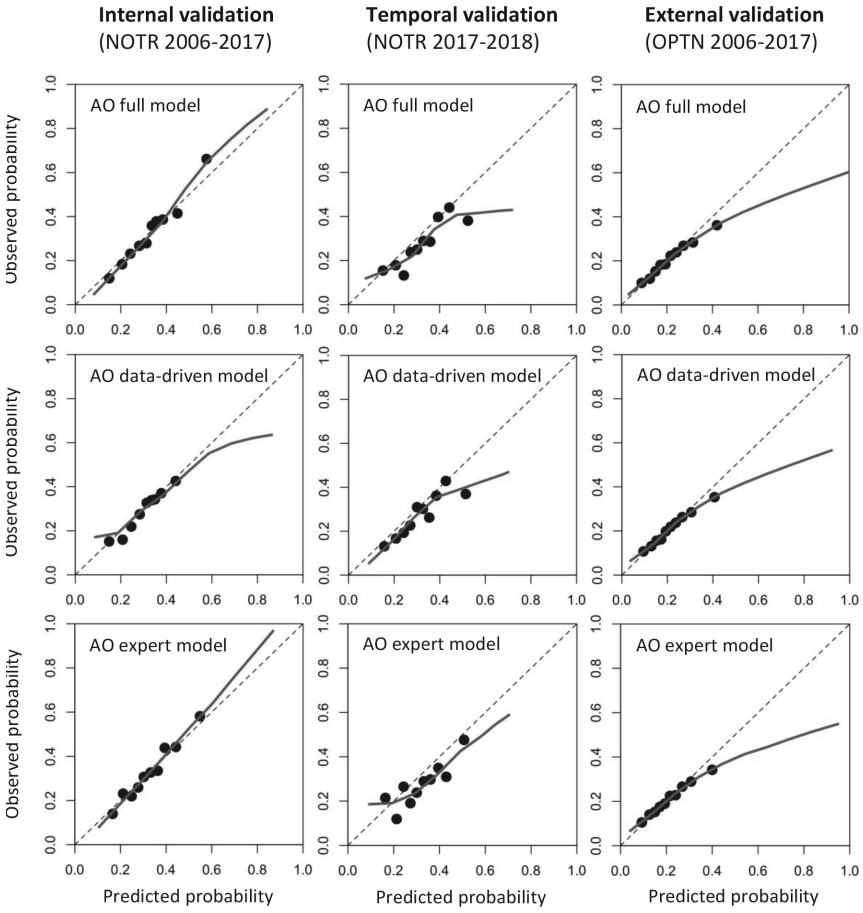

Figure 1: Calibration plots of AO models in internal, temporal and external validation. The external validation plots are recalibrated to the U.S. outcome incidence. Predicted risk on the x-axis and observed risk on the y-axis per decile of predicted probability, augmented by a smoothed (lowess) regression line. The 45° dotted line indicates perfect agreement between predicted and observed risks.

Clinical applicability of the AO models

An individual's probability of having an AO in the first year after receiving a kidney transplant from a deceased donor aged 50 years or older, can be calculated using the formulas provided in the supplemental material or R-script. Both a European formula and a North American formula are provided. Risk predictions for four hypothetical patients are shown in Table S7. The characteristics of these example patients were defined by independent nephrologists, to exemplify 4 realistic organ offer scenarios ranging from ideal to poor. As these models might be of use to aid in the clinical decision whether

to accept or reject a kidney prior to transplantation, diagnostic properties of various decision thresholds are shown in Table 6. The specificity is generally high; the prediction model correctly generates a low predicted risk for recipients who do not experience AO. However, the sensitivity is very low, meaning there are many AO cases that are missed by the prediction model. Of the recipient-donor pairs with a high predicted risk, less than half will get an AO (the positive predictive value is lower than 50%). However, a low predicted risk will usually mean that the recipient will not experience an AO within 1 year after transplantation (high negative predictive value). These thresholds are solely given as examples. The models should not be used to *fully* determine the acceptance or decline of donor kidneys but may enhance the physicians' decision process.

Table 6: Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for various hypothetical risk thresholds based on the developed AO full prediction model. Calculated on the temporal NOTR validation cohort. The 'N risk < threshold' would be the number of accepted kidneys. The false negatives are the number of these transplanted recipients whom would experience graft failure, death or CKD stage 4+ within 1 year of transplantation. The false positives are the number of donor-recipient pairs whose predicted risk was above the threshold, yet did not experience the outcome within 1 year. ≥

Thresholds AO full model	Sensitivity	Specificity	PPV	NPV	N risk < threshold	N false negatives	N risk ≥ threshold	N false positives
P ≥ 70%	0.4%	99.8%	50.0%	72.6%	835	229	2	1
P ≥ 65%	0.9%	99.3%	33.3%	72.6%	831	228	6	4
P ≥ 60%	2.2%	99.0%	45.5%	72.8%	826	225	11	6
P ≥ 55%	4.3%	98.0%	45.5%	73.0%	815	220	22	12
P ≥ 50%	8.7%	94.9%	39.2%	73.3%	786	210	51	31
P ≥ 45%	19.1%	89.1%	40.0%	74.5%	728	186	110	66
P ≥ 40%	35.7%	81.2%	41.8%	76.9%	641	148	196	114

Discussion

In the current study we developed and validated prediction models of AO (graft failure, recipient death or CKD stage 4+) after kidney transplantation from older deceased donors, using pre-transplant donor and recipient characteristics. Additionally, 7 existing prediction models of graft survival were validated in the same cohorts. The current study improved on existing studies by employing advanced statistical methods and choosing a broader outcome definition with a shorter prediction horizon. In addition, we selected a clinically relevant population of older deceased donors, developed models in a European population and subsequently updated these models for a North American population, making them applicable to patients in both regions. The discrimination of existing models was poor compared to a moderate discrimination of our new AO models in external validation. Overall, the developed AO models display a good calibration. When investigating various decision thresholds for kidney acceptance, the AO full model showed high specificity; the model can accurately classify donor-recipient pairs as low-risk.

To the best of our knowledge, this study presents the most comprehensive external validation of existing kidney graft survival models so far. Previously, the KDRI has been externally validated numerous times and consistently showed a moderate discrimination with a C-statistic around 0.62.[10 13 42] When considering that these validation cohorts also included many young donors, where the transplant clinicians are not in doubt whether to accept or decline the kidney offer, the added value of these models over physicians' judgement is unsure. Regrettably, most existing studies presenting prediction models, including the KDRI, do not report the full prediction model and therefore cannot be used to calculate absolute risks for individuals. The KDRI is rescaled to the KDPI which is a relative measure and relates to the average kidney transplanted in the year prior in the US. By doing so, the same KDPI score may translate to very different absolute risks of graft failure in different years.[43] If over time more high-risk kidneys are discarded based on their KDRI score, the donor pool will continue to decrease each year. So far, no prediction model has been developed which is also recommended for use in European populations. Existing models show a substantial variety in included predictors, which exemplifies the difficulty of predicting future transplantation outcomes.[10] Our conducted expert ranking shows that nephrologists also have different opinions on which factors have prognostic value when appraising a kidney offer.

Our present study has a number of limitations. The most important limitation is that, although we aimed to provide prediction models that are applicable prior to transplantation, we only have outcome data for those kidneys that were actually transplanted. This means that the included donor-recipient pairs were subjected to the current allocation system and the healthcare professionals' judgement. Declined kidneys were not included in the current analyses, although these models would be used on such kidneys in clinical practice. This may bias the model's predictions and is a limitation of all studies similar to the present one. Ideally, we would provide a recipient's risk of adverse outcome if this recipient remains on the waiting list (dialysis) versus if the patient receives a particular kidney offered for transplantation. The comparison of these risks could truly guide decision making. Such a model would require assumptions on what would happen to an individual had he or she not received a particular kidney that was offered for transplantation. This is extremely difficult to assess without a randomized controlled trial but could be attempted in future studies using inverse probability weighting. Additionally, by using a combined outcome measure, some interpretability of the predicted risks is lost. This may be a limiting factor when trying to reduce mismatches between a donor kidney's longevity and a recipient's life expectancy.[44] Furthermore, the 7 existing models were not specifically developed for an older donor population. By validating these in such a sub-population, the discrimination will invariably be lower. Although calibration was reasonably good, the discrimination of our developed AO models was moderate at best. It remains to be seen if models with such moderate discrimination can actually improve clinical practice and patient outcomes. Although our models are clearly better than chance, their effect on patients should be evaluated in impact studies. Finally, over the span of the current study (2006-2018) some gradual, but nevertheless important new developments took place, such as the increasing

use of older donors, acceptance of older recipients, change in immunosuppressive regimens and increasing use of hypothermic machine perfusion to improve preservation of grafts. However, the AO models showed consistent performance in the recent 2017-2018 NOTR temporal validation cohort, indicating model robustness.

One of the strengths of this study is that a large number of existing models and a newly developed model were compared in external validation on independent data, allowing for a fair comparison of predictive performance. Furthermore, as most existing models were exclusively developed and validated on US data, the use of a large contemporary European cohort improves generalizability. As the US population is structurally different from the European cohort, using the OPTN as external validation cohort allowed us to adequately test transportability of the newly developed AO models. To increase clinical relevance, only older deceased donors in whom there is relatively little consensus amongst physicians whether to accept or decline were included. Lastly, by including a multitude of donor and recipient characteristics, interaction terms, non-linear associations, and multiple validation steps, we strove to create the most optimal prediction models that could be obtained with the available data, working in accordance with the most recent methodological and statistical recommendations on model development and validation.

The developed AO models can aid clinicians' decision-making surrounding acceptance or decline of kidneys especially from older deceased donors. Considering the high specificity of the AO model, it can accurately classify recipients as low risk. Besides augmenting individualized decision-making, these models could be a useful learning tool for more inexperienced physicians, as they help develop a feeling for risks associated with various patient characteristic profiles and donor-recipient combinations. For clinical use, we would recommend either the AO full model or AO data-driven model. Although the full model has a slightly higher discrimination, the data-driven model is more convenient as it contains less predictors. These models cannot replace a physician's judgment, but can provide added value to a clinician's decision of donor kidney acceptance and a more objective assessment which may improve uniformity between transplant centers. As our predictive models remain moderately precise at best, the transportability to different settings, without recalibration, is questionable.

Predicting *post*-transplantation outcomes *prior to* transplantation remains difficult and future research might explore novel biomarkers or ex vivo perfusion parameters that could improve predictions. However, due to the large number of unpredictable and dynamic post-transplantation factors that affect transplant outcomes, there may only be limited room for improvement. Future studies may also look into using competing risk prediction models to predict various outcomes separately. In addition, the prediction of other outcome parameters related to patient reported outcomes (PRO's) and quality of life deserves more attention. Such outcomes are often overlooked although they are extremely important to many patients and can play a valuable role in shared decision making. A comprehensive allocation scheme including donor and recipient characteristics, various outcomes, an individual's prognoses for various treatment options and longevity matching is difficult to capture within a single or even multiple prediction models. More research is needed

to further elucidate these important prognoses and ultimately design the most optimal organ allocation policy.

In conclusion, we externally validated 7 existing prediction models and developed as well as externally validated new AO prediction models for kidneys from deceased donors ≥50 years. These AO models may be used to aid decisions on acceptance of kidneys from older deceased donors and were developed on European transplant data. Although their predictive performance was moderate, it was superior to existing models. Therefore, given current registry data and known predictors, our study provides new adverse outcome prediction models with maximum achievable performance for renal grafts recovered from older deceased donors.

References

1. Wolfe RA, Ashby VB, Milford EL, et al. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N Engl J Med* 1999;341(23):1725-30. doi: 10.1056/nejm199912023412303 [published Online First: 1999/12/02]
2. Liem YS, Bosch JL, Arends LR, et al. Quality of life assessed with the Medical Outcomes Study Short Form 36-Item Health Survey of patients on renal replacement therapy: a systematic review and meta-analysis. *Value Health* 2007;10(5):390-7. doi: 10.1111/j.1524-4733.2007.00193.x [published Online First: 2007/09/25]
3. Port FK, Bragg-Gresham JL, Metzger RA, et al. Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors. *Transplantation* 2002;74(9):1281-6. doi: 10.1097/00007890-200211150-00014 [published Online First: 2002/11/27]
4. Tullius SG, Rabb H. Improving the Supply and Quality of Deceased-Donor Organs for Transplantation. *N Engl J Med* 2018;378(20):1920-29. doi: 10.1056/NEJMra1507080 [published Online First: 2018/05/17]
5. Querard AH, Foucher Y, Combescure C, et al. Comparison of survival outcomes between Expanded Criteria Donor and Standard Criteria Donor kidney transplant recipients: a systematic review and meta-analysis. *Transpl Int* 2016;29(4):403-15. doi: 10.1111/tri.12736 [published Online First: 2016/01/13]
6. van Ittersum FJ, Hemke AC, Dekker FW, et al. Increased risk of graft failure and mortality in Dutch recipients receiving an expanded criteria donor kidney transplant. *Transpl Int* 2017;30(1):14-28. doi: 10.1111/tri.12863 [published Online First: 2016/09/21]
7. Cooper M, Formica R, Friedewald J, et al. Report of National Kidney Foundation Consensus Conference to Decrease Kidney Discards. *Clin Transplant* 2019;33(1):e13419. doi: 10.1111/ctr.13419 [published Online First: 2018/10/23]
8. Mohan S, Chiles MC, Patzer RE, et al. Factors leading to the discard of deceased donor kidneys in the United States. *Kidney Int* 2018;94(1):187-98. doi: 10.1016/j.kint.2018.02.016 [published Online First: 2018/05/08]
9. Mathur AK, Ashby VB, Sands RL, et al. Geographic variation in end-stage renal disease incidence and access to deceased donor kidney transplantation. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2010;10(4 Pt 2):1069-80. doi: 10.1111/j.1600-6143.2010.03043.x [published Online First: 2010/04/28]
10. Kabore R, Haller MC, Harambat J, et al. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2017;32(suppl_2):ii68-ii76. doi: 10.1093/ndt/gfw405 [published Online First: 2017/02/17]
11. Loupy A, Aubert O, Orandi BJ, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. 2019;366:l4923. doi: 10.1136/bmj.l4923 %J BMJ
12. Loupy A, Aubert O, Orandi BJ, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. *BMJ (Clinical research ed)* 2019;366:l4923. doi: 10.1136/bmj.l4923
13. Rao PS, Schaubel DE, Guidinger MK, et al. A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. *Transplantation* 2009;88(2):231-6. doi: 10.1097/TP.0b013e3181ac620b [published Online First: 2009/07/23]
14. Stewart DE, Kucheryavaya AY, Klassen DK, et al. Changes in Deceased Donor Kidney Transplantation One Year After KAS Implementation. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2016;16(6):1834-47. doi: 10.1111/ajt.13770 [published Online First: 2016/03/05]
15. Massie AB, Luo X, Lonze BE, et al. Early Changes in Kidney Distribution under the New Allocation System. *J Am Soc Nephrol* 2016;27(8):2495-501. doi: 10.1681/asn.2015080934 [published Online First: 2015/12/19]

16. Ruggenenti P, Remuzzi G. Invited letter in response to: "Is the kidney donor profile index (KDPI) universal or UNOS-specific?". *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2018;18(4):1033-34. doi: 10.1111/ajt.14652 [published Online First: 2018/01/10]
17. Stallone G, Grandaliano G. To discard or not to discard: transplantation and the art of scoring. *Clinical Kidney Journal* 2019;12(4):564-68. doi: 10.1093/ckj/sfz032 %J Clinical Kidney Journal
18. Bae S, Massie AB, Luo X, et al. Changes in Discard Rate After the Introduction of the Kidney Donor Profile Index (KDPI). *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2016;16(7):2202-7. doi: 10.1111/ajt.13769 [published Online First: 2016/03/05]
19. Kasiske BL, Israni AK, Snyder JJ, et al. A simple tool to predict outcomes after kidney transplant. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2010;56(5):947-60. doi: 10.1053/j.ajkd.2010.06.020 [published Online First: 2010/08/31]
20. Vinson AJ, Kiberd BA, Davis RB, et al. Nonimmunologic Donor-Recipient Pairing, HLA Matching, and Graft Loss in Deceased Donor Kidney Transplantation. *Transplantation direct* 2019;5(1):e414. doi: 10.1097/txd.0000000000000856 [published Online First: 2019/01/19]
21. Molnar MZ, Nguyen DV, Chen Y, et al. Predictive Score for Posttransplantation Outcomes. 2017;101(6):1353-64. doi: 10.1097/tp.0000000000001326
22. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed)* 2015;350:g7594. doi: 10.1136/bmj.g7594 [published Online First: 2015/01/09]
23. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162(1):W1-73. doi: 10.7326/m14-0698 [published Online First: 2015/01/07]
24. Jeldres C, Cardinal H, Duclos A, et al. Prediction of delayed graft function after renal transplantation. *Can Urol Assoc J* 2009;3(5):377-82. doi: 10.5489/cuaj.1147 [published Online First: 2009/10/16]
25. Zaza G, Ferraro PM, Tessari G, et al. Predictive model for delayed graft function based on easily available pre-renal transplant variables. *Intern Emerg Med* 2015;10(2):135-41. doi: 10.1007/s11739-014-1119-y [published Online First: 2014/08/29]
26. Irish WD, Ilsley JN, Schnitzler MA, et al. A risk prediction model for delayed graft function in the current era of deceased donor renal transplantation. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2010;10(10):2279-86. doi: 10.1111/j.1600-6143.2010.03179.x [published Online First: 2010/10/05]
27. Chapal M, Le Borgne F, Legendre C, et al. A useful scoring system for the prediction and management of delayed graft function following kidney transplantation from cadaveric donors. *Kidney Int* 2014;86(6):1130-9. doi: 10.1038/ki.2014.188 [published Online First: 2014/06/05]
28. Summers DM, Johnson RJ, Hudson A, et al. Effect of donor age and cold storage time on outcome in recipients of kidneys donated after circulatory death in the UK: a cohort study. *Lancet* 2013;381(9868):727-34. doi: 10.1016/S0140-6736(12)61685-7 [published Online First: 2012/12/20]
29. Miller AJ, Kiberd BA, Alwayn IP, et al. Donor-Recipient Weight and Sex Mismatch and the Risk of Graft Loss in Renal Transplantation. 2017;12(4):669-76. doi: 10.2215/CJN.07660716 %J Clinical Journal of the American Society of Nephrology
30. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150(9):604-12. doi: 10.7326/0003-4819-150-9-200905050-00006 [published Online First: 2009/05/06]
31. de Goeij MC, van Diepen M, Jager KJ, et al. Multiple imputation: dealing with missing data. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2013;28(10):2415-20. doi: 10.1093/ndt/gft221 [published Online First: 2013/06/05]
32. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research* 2007;16(3):219-42. doi: 10.1177/0962280206074463 [published Online First: 2007/07/11]

33. Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clinical research ed)* 2009;338:b604. doi: 10.1136/bmj.b604 [published Online First: 2009/04/02]
34. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry* 2008;54(1):17-23. doi: 10.1373/clinchem.2007.096529 [published Online First: 2007/11/21]
35. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Statistics in medicine* 2014;33(3):517-35. doi: 10.1002/sim.5941 [published Online First: 2013/09/05]
36. Harrell F. REGRESSION MODELING STRATEGIES : with applications to linear models, logistic and ordinal ... regression, and survival analysis. [Place of publication not identified]: SPRINGER 2016.
37. Steyerberg EW, Eijkemans MJC, Harrell Jr FE, et al. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. 2000;19(8):1059-79. doi: 10.1002/(sici)1097-0258(20000430)19:8<1059::Aid-sim412>3.0.Co;2-0
38. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart (British Cardiac Society)* 2012;98(9):683-90. doi: 10.1136/heartjnl-2011-301246 [published Online First: 2012/03/09]
39. Janssen KJ, Vergouwe Y, Kalkman CJ, et al. A simple method to adjust clinical prediction models to local circumstances. *Canadian journal of anaesthesia = Journal canadien d'anesthesie* 2009;56(3):194-201. doi: 10.1007/s12630-009-9041-x [published Online First: 2009/02/28]
40. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley 1987.
41. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine* 2019;170(1):51-58. doi: 10.7326/M18-1376 %J Annals of Internal Medicine
42. Peters-Sengers H, Heemskerk MBA, Geskus RB, et al. Validation of the Prognostic Kidney Donor Risk Index Scoring System of Deceased Donors for Renal Transplantation in the Netherlands. *Transplantation* 2018;102(1):162-70. doi: 10.1097/tp.0000000000001889 [published Online First: 2017/07/22]
43. Salkowski N GS, Wey A, Snyder J. KDPI Obscures Trends in Absolute Donor Risk. Am J Transplant, 2017.
44. Friedewald JJ, Samana CJ, Kasiske BL, et al. The kidney allocation system. *Surg Clin North Am* 2013;93(6):1395-406. doi: 10.1016/j.suc.2013.08.007 [published Online First: 2013/11/12]
45. Schold JD, Kaplan B, Baliga RS, et al. The broad spectrum of quality in deceased donor kidneys. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2005;5(4 Pt 1):757-65. doi: 10.1111/j.1600-6143.2005.00770.x [published Online First: 2005/03/12]
46. Watson CJE, Johnson RJ, Birch R, et al. A simplified donor risk index for predicting outcome after deceased donor kidney transplantation. *Transplantation* 2012;93(3):314-18. doi: 10.1097/TP.0b013e31823f14d4

Supplemental Material for Chapter 9

Systematic search methods

To identify existing prediction models which were suitable for external validation, a recently published systematic review was screened and their search was updated in Pubmed to include articles published up to February 2019 (see supplement for search strategy).[1] Studies presenting prediction models for kidneys retrieved from deceased donors were included if the development population included adult kidney recipients, the outcome predicted contained graft failure, the competing risk of recipient death was handled appropriately (either by using a combined end point or competing risk model) and the moment of prediction was prior to transplantation. Additionally, sufficient model information (regression coefficients or hazard ratios for each predictor) had to be reported and the included predictors or suitable proxies had to be available in at least one of our validation datasets. Screening of articles and data extraction was performed by two independent reviewers (CLR & EW).

Search strategy Pubmed (search performed on 22 February 2019)

(((”graft failure”[tw] OR ”graft failures”[tw] OR graft fail*[tw] OR ”Graft Rejection”[Mesh] OR ”graft rejection”[tw] OR ”graft rejections”[tw] OR ”transplant rejection”[tw] OR ”transplantation rejection”[tw] OR ”transplantation rejections”[tw] OR ”transplant rejections”[tw] OR ”Graft Survival”[Mesh] OR ”graft survival”[tw] OR ”Treatment Failure”[Mesh:NoExp]) AND (”Kidney Transplantation”[mesh] OR ”kidney transplantation”[tw] OR kidney transplant*[tw] OR kidney graft*[tw] OR ”renal transplantation”[tw] OR renal transplant*[tw] OR renal graft*[tw] OR ((”kidney”[tw] OR kidney*[tw] OR ”renal”[tw] OR renal*[tw]) AND (”transplantation”[tw] OR transplant*[tw] OR graft*[tw])))) AND (”prediction model”[tw] OR ”prediction models”[tw] OR ”prediction modeling”[tw] OR ”prediction modelling”[tw] OR ”prediction model”[tw] OR prediction model*[tw] OR ”predictive model”[tw] OR ”predictive models”[tw] OR ”predictive modeling”[tw] OR ”predictive modelling”[tw] OR ”predictive model”[tw] OR predictive model*[tw] OR ”prediction score”[tw] OR ”prediction scores”[tw] OR prediction scor*[tw] OR ”predictive score”[tw] OR ”predictive scores”[tw] OR predictive score*[tw] OR ”prognostic model”[tw] OR ”prognostic models”[tw] OR ”prognostic modeling”[tw] OR ”prognostic modelling”[tw] OR prognostic model*[tw] OR ”prognostic score”[tw] OR ”prognostic scores”[tw] OR prognostic scor*[tw] OR ”prognostic risk model”[tw] OR ”prognostic risk models”[tw] OR ”prognostic risk score”[tw] OR ”prognostic risk scores”[tw])) **OR** ((”graft failure”[tw] OR ”graft failures”[tw] OR graft fail*[tw] OR ”Graft Rejection”[mesh] OR ”graft rejection”[tw] OR ”graft rejections”[tw] OR ”transplant rejection”[tw] OR ”transplantation rejection”[tw] OR ”transplantation rejections”[tw] OR ”transplant rejections”[tw] OR ”Graft Survival”[mesh] OR ”graft survival”[tw] OR ”Treatment Failure”[mesh:NoExp]) AND (”Kidney Transplantation”[majr] OR ”kidney transplantation”[ti] OR kidney transplant*[ti]

OR kidney graft*[ti] OR "renal transplantation"[ti] OR renal transplant*[ti] OR renal graft*[ti] OR (("kidney"[ti] OR kidney*[ti] OR "renal"[ti] OR renal*[ti]) AND ("transplantation"[ti] OR transplant*[ti] OR graft*[ti]))) AND ("prediction"[tw] OR "predictive"[tw] OR predict*[tw] OR "Forecasting"[mesh:NoExp] OR Forecast*[tw] OR prognos*[tw] OR "Prognosis"[Mesh:noexp]) AND ("Models, Theoretical"[mesh] OR model*[tiab]))) AND ("2015/01/01"[PDAT] : "3000/12/31"[PDAT]) NOT ("after"[ti] NOT "before"[ti])

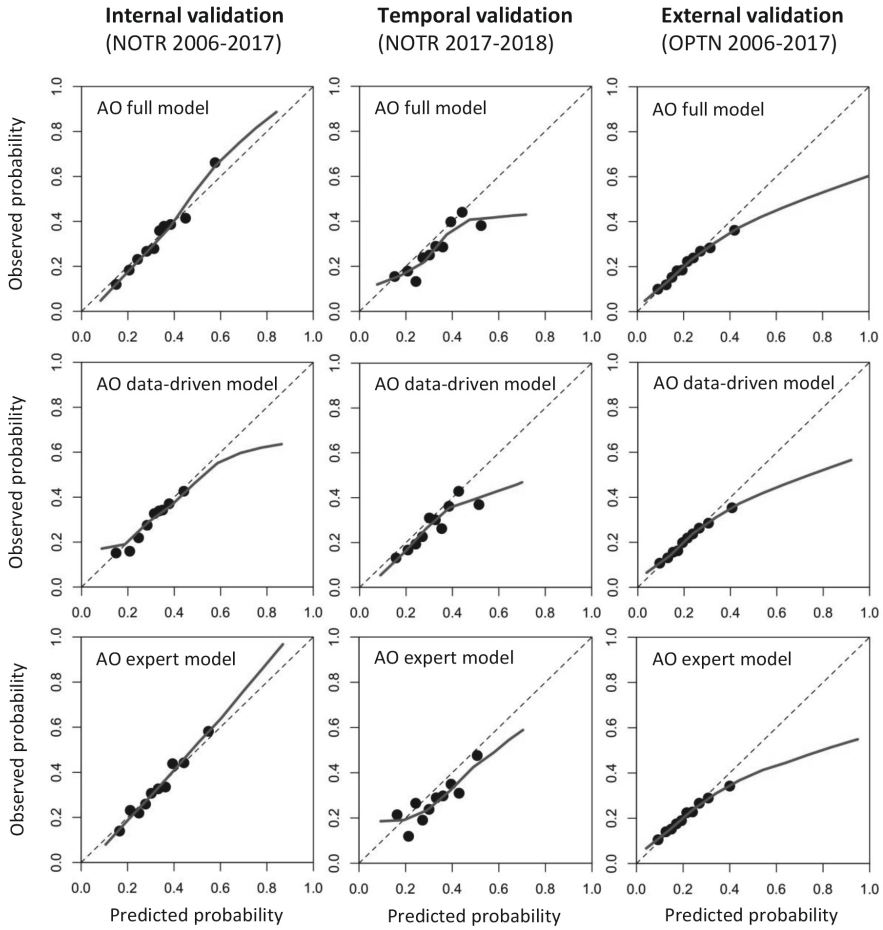

Figure S1: Study selection flowchart for external validation of existing prediction models.

Table S1: Risk of bias score, according to PROBAST tool.

	Selection of participants	Predictors or their assesment	Outcome or its determination	Analysis	Overall
Schold	Low	Low	High	High	High
Rao et al.	Low	Low	High	High	High
Kasiske et al.	Low	Low	Low	High	High
Watson	Low	Low	Low	High	High
Molnar	Unclear	Low	Low	Low	Low
Vinson et al.	Low	Low	High	High	High

Table S2: Baseline characteristics of the NOTR development cohort (January 2006 - January 2016), stratified on outcome and including number of missing values.

	Total cohort n = 2510	Missing values	No adverse events within 1 year n=1687	Adverse event within 1 year n=823
Donor characteristics				
Age (years)	60 (55-65)	0%	59 (54-64)	62 (56-67)
Sex (% male)	51.4%	0%	51.0%	52.2%
Ethnicity (%)	-	100%	-	-
Cause of death (%)		10%		
Trauma	14.4%		15.5%	12.2%
Cerebrovascular accident	64.4%		64.1%	65.0%
Anoxia	18.1%		17.1%	20.1%
Other	3.1%		3.4%	2.7%
DCD donor (%)	46.1%	0%	43.1%	52.2%
CPR performed (%)	31.2%	9%	30.6%	32.5%
Days in hospital	2 (1-5)	1%	2 (1-4)	2 (1-5)
Serum creatinine (µmol/L)	66 (53-83)	<1%	65 (52-80)	68 (54-86)
Proteinuria (%)	44.4%	29%	45.0%	43.1%
BMI (kg/m^2)	26 (4.7)	0%	26 (4.4)	26 (5.2)
History of diabetes mellitus (%)	8.1%	8%	7.4%	9.5%
History of hypertension (%)	37.5%	9%	35.9%	40.8%
HCV status positive	0.0%	0%	0.0%	0.1%
History of smoking (%)	55.0%	5%	55.9%	53.0%
Hypotension (%)	31.5%	14%	31.7%	31.1%
Use of inotropic medication (%)	71.7%	0%	74.0%	67.0%
Left kidney (%)	50.4%	0%	50.0%	51.3%

Table S2: Continued.

	Total cohort n = 2510	**Missing values**	**No adverse events within 1 year n=1687**	**Adverse event within 1 year n=823**
WIT in DCD donors (minutes)	17 (14-21)	2%	16 (13-20)	17 (14-21)
Cold ischemic time (hours)	15.8 (5.8)	32%	15.5 (5.7)	16.3 (5.9)
Recipient characteristics				
Age (years)	60 (49-67)	0%	59 (49-66)	63 (51-68)
Sex (% male)	60.6%	0%	61.5%	58.6%
Ethnicity (%)	-	100%	-	
BMI (kg/m^2)	26 (4.7)	5%	26 (4.6)	27 (4.8)
Primary kidney disease (%)		4%		
Diabetes mellitus	14.0%		13.5%	14.9%
Hypertension	20.6%		20.0%	22.0%
Glomerular nephritis	16.6%		17.1%	15.5%
Cystic kidney disease	14.7%		15.8%	12.2%
Other	34.2%		35.6%	35.4%
Diabetes mellitus (%)	21.5%	11%	19.9%	24.9%
History of CVA (%)	8.4%	4%	7.8%	9.6%
Coronary artery disease (%)	18.4%	3%	17.2%	20.8%
Peripheral vascular disease event (%)	13.5%	3%	12.4%	15.7%
HCV status positive	0.3%	0%	0.3%	0.2%
Time on dialysis (months)	39 (25-57)	1%	39 (24-57)	39 (26-57)
≥1 previous kidney transplant (%)	12.9%	0%	13.2%	12.4%
Donor-recipient characteristics				
Total number of HLA mismatches	3 (2-4)	1%	3 (2-4)	3 (2-4)
Peak PRA	0 (0-0)*	0%	0 (0-0)	0 (0-0)

Abbreviations: DCD: donation after circulatory death. BMI:body mass index; WIT: warm ischemic time, PRA:panel reactive antibody; HLA:human leukocyte antigen. Lab values are shown in SI units and can be converted to conventional units as follows, serum creatinine in mg/dL: multiply by 0.011. *mean Peak PRA is 4.3 (SD: 14.5). Coronary artery disease was defined as the occurrence of myocardial infarction, percutaneous coronary intervention or cardiac bypass surgery. Peripheral vascular disease event was defined as percutaneous angioplasty, bypass operation (non-cardiac), amputation for vascular reasons or aortic bifurcation graft.

Table S3: Baseline characteristics of the NOTR temporal validation cohort (January 2016 - November 2018), stratified on outcome and including number of missing values.

	Total cohort n = 837	Missing values	No adverse events within 1 year n=607	Adverse event within 1 year n=230
Donor characteristics				
Age (years)	61 (55-66)	0%	60 (54-65)	64 (58-69)
Sex (% male)	56.2%	0%	58.0%	51.3%
Ethnicity (%)	-	100%	-	-
Cause of death (%)		9%		
Trauma	16.9%		17.7%	14.9%
Cerebrovascular accident	56.2%		54.7%	60.0%
Anoxia	24.6%		25.0%	23.7%
Other	2.2%		2.6%	2.6%
DCD donor (%)	58.8%	0%	59.3%	57.4%
CPR performed (%)	36.5%	6%	38.2%	32.1%
Days in hospital	2 (1-5)	4%	2 (1-5)	2 (1-4)
Serum creatinine (µmol/L)	64 (52-82)	0%	64 (52-80)	65 (54-84)
Proteinuria (%)	49.4%	44%	49.0%	50.4%
BMI (kg/m^2)	26 (4.4)	0%	26 (4.4)	26 (5.2)
History of diabetes mellitus (%)	9.3%	5%	8.4%	11.6%
History of hypertension (%)	38.1%	7%	35.1%	45.5%
HCV status positive	0.0%	0%	0.0%	0.0%
History of smoking (%)	56.4%	3%	55.6%	58.6%
Hypotension (%)	21.9%	13%	21.8%	22.2%
Use of inotropic medication (%)	69.9%	0%	68.0%	74.8%
Left kidney (%)	49.8%	0%	49.8%	50.0%
WIT in DCD donors (minutes)	15 (13-18)	9%	15 (13-18)	15 (13-17)
Cold ischemic time (hours)	13.3 (5.7)	8%	13.4 (5.8)	13.1 (5.4)
Recipient characteristics				
Age (years)	62 (51-69)	0%	61 (50-68)	65 (55-70)
Sex (% male)	63.6%	0%	65.4%	58.7%
Ethnicity (%)	-	100%	-	-
BMI (kg/m^2)	27 (4.4)	3%	27 (4.3)	27 (4.6)
Primary kidney disease (%)		4%		
Diabetes mellitus	18.3%		18.4%	18.1%
Hypertension	22.4%		21.2%	25.8%
Glomerular nephritis	17.8%		16.7%	20.8%

Table S3: Continued.

	Total cohort n = 837	**Missing values**	**No adverse events within 1 year n=607**	**Adverse event within 1 year n=230**
Cystic kidney disease	9.9%		11.1%	6.8%
Other	31.5%		32.6%	28.5%
Diabetes mellitus (%)	26.8%	9%	25.8%	29.4%
History of CVA (%)	8.2%	5%	7.6%	9.7%
Coronary artery disease (%)	17.3%	5%	16.7%	18.9%
Peripheral vascular disease event (%)	10.7%	6%	10.9%	10.1%
HCV status positive	0.4%	0%	0.3%	0.4%
Time on dialysis (months)	25 (15-42)	16%	26 (15-43)	24 (18-41)
≥1 previous kidney transplant (%)	15.2%	0%	15.5%	14.3%
Donor-recipient characteristics				
Total number of HLA mismatches	3 (2-4)	<1%	3 (2-4)	3 (2-4)
Peak PRA	0 (0-0)*	0%	0 (0-0)	0 (0-0)

Abbreviations: DCD: donation after circulatory death. BMI:body mass index; WIT: warm ischemic time, PRA:panel reactive antibody; HLA:human leukocyte antigen. Lab values are shown in SI units and can be converted to conventional units as follows, serum creatinine in mg/dL: multiply by 0.011. *mean Peak PRA is 4.0 (SD: 14.7). Coronary artery disease was defined as the occurrence of myocardial infarction, percutaneous coronary intervention or cardiac bypass surgery. Peripheral vascular disease event was defined as percutaneous angioplasty, bypass operation (non-cardiac), amputation for vascular reasons or aortic bifurcation graft.

Table S4: Baseline characteristics of the OPTN external validation cohort, stratified on outcome and including number of missing values.

	Total cohort n = 31987	Missing values	No adverse events within 1 year n=25229	Adverse event within 1 year n=6758
Donor characteristics				
Age (years)	56 (53-60)	0%	55 (52-60)	57 (53-62)
Sex (% male)	53.5%	0%	54.2%	50.8%
Ethnicity (%)		0%		
White	73.4%		74.0%	71.5%
Black	11.2%		10.6%	13.4%
Hispanic	11.0%		11.2%	10.0%
Other	4.4%		4.2%	5.1%
Cause of death (%)		0%		
Trauma	21.1%		22.2%	17.0%
Cerebrovascular accident	56.5%		54.9%	62.4%
Anoxia	20.0%		20.5%	18.1%
Other	2.4%		2.0%	2.2%
DCD donor (%)	13.8%	0%	13.4%	15.1%
CPR performed (%)	5.0%	0%	5.1%	4.6%
Days in hospital	4 (3-6)	7%	4 (3-6)	4 (3-6)
Serum creatinine (µmol/L)	80 (62-106)	<1%	80 (62-106)	88 (66-115)
Proteinuria (%)	41.9%	1%	41.8%	42.2%
BMI (kg/m^2)	29 (6.4)	<1%	29 (6.3)	29 (6.8)
History of diabetes mellitus (%)	12.7%	1%	11.8%	16.3%
History of hypertension (%)	50.6%	1%	48.6%	58.2%
HCV status positive (%)	1.7%	<1%	1.8%	1.5%
History of smoking (%)	38.0%	2%	37.8%	38.5%
Hypotension (%)	-	100%	-	-
Use of inotropic medication (%)	51.8%	<1%	52.0%	51.1%
Left kidney (%)	49.6%	0%	49.5%	50.0%
WIT in DCD donors (minutes)	18 (11-27)	4%	17 (11-26)	19 (12-30)
Cold ischemic time (hours)	18.2 (9.2)	2%	17.9 (9.0)	19.6 (9.5)
Recipient characteristics				
Age (years)	60 (51-66)	0%	59 (51-66)	61 (54-68)
Sex (% male)	62.2%	0%	62.3%	62.2%
Ethnicity (%)		0%		
White	44.1%		43.3%	47.2%

Table S4: Continued.

	Total cohort n = 31987	**Missing values**	**No adverse events within 1 year n=25229**	**Adverse event within 1 year n=6758**
Black	32.0%		31.6%	33.7%
Hispanic	14.6%		15.3%	11.9%
Other	9.3%		9.8%	7.2%
BMI (kg/m^2)	28 (5.2)	<1%	28 (5.2)	29 (5.4)
Primary kidney disease (%)		<1%		
Diabetes mellitus	32.6%		32.2%	34.0%
Hypertension	25.9%		25.8%	26.1%
Glomerular nephritis	11.4%		11.7%	10.2%
Cystic kidney disease	7.7%		8.0%	6.6%
Other	22.4%		22.3%	23.2%
Diabetes mellitus (%)	43.0%	1%	42.3%	45.7%
History of CVA (%)	-	100%	-	-
Coronary artery disease (%)	-	100%	-	-
Peripheral vascular disease (%)	8.4%	2%	7.9%	10.0%
HCV status positive (%)	5.1%	4%	5.2%	4.7%
Time on dialysis (months)	40 (13-66)	0%	39 (12-44)	41 (17-66)
≥1 previous kidney transplant (%)	8.9%	0%	8.9%	8.7%
Health insurance (%)		0%		
Medicare	72.0%		71.1%	75.3%
Medicaid	4.1%		4.2%	3.7%
Private	22.3%		23.1%	19.6%
Other	1.6%		1.6%	1.4%
Donor-recipient characteristics				
Total number of HLA mismatches	5 (4-5)	0%	5 (4-5)	5 (4-5)
Peak PRA	0 (0-13)	67%	0 (0-13)	0 (0-13)

Abbreviations: DCD: donation after circulatory death. BMI:body mass index; PRA:panel reactive antibody; HLA:human leukocyte antigen. Lab values are shown in SI units and can be converted to conventional units as follows, serum creatinine in mg/dL: multiply by 0.011

Tabel S5: expert opinion ranking, median rank calculated from survey responses from 10 nephrologists.

	Predictors	Median rank (IQR)
1	Donor Age	1 (1-3)
2	Donor serum creatinine	2 (1-6)
3	Donor DM	5 (3-12)
4	Recipient Age	6 (4-7)
5	Cold ischemic time	6 (4-8)
6	Warm ischemic time	6 (3-11)
7	Recipient dialysis duration	8 (6-12)
8	Recipient DM	8 (6-15)
9	Donation after circulatory death	10 (3-13)
10	Donor hypotension	10 (6-18)
11	Donor hypertension	12 (8-15)
12	Recipient previous kidney Tx	12 (9-14)
13	Donor cause of death	14 (10-15)
14	Total number of HLA mismatches	14 (10-17)
15	Peak PRA	14 (11-17)
16	Donor inotropes use	15 (11-17)
17	Donor BMI	15 (13-17)
18	Recipient BMI	16 (14-18)
19	Donor Gender	20 (17-21)
20	Recipient Gender	20 (19-20)
21	Left or right kidney	20 (19-21)

Based on the nephrologists surveys the following predictors were added to the list mentioned above: donor proteinuria, recipient cardiovascular comorbidities, donor CPR, donor smoking and primary kidney disease of recipient.

Table S6: External validation results of existing models. Discrimination and calibration results.

		NOTR 2006-2017	NOTR 2017-2018	OPTN 2006-2017	OPTN 2006-2017 with US recalibration
Schold	C-statistic (95% CI)	0.562 (0.532-0.591)	0.555 (0.495-0.615)	0.577 (0.567-0.586)	0.577 (0.567-0.586)
	Calibration slope	0.917	0.737	1.113	1.113
	Calibration-in-the-large	16.4% vs 15.2%	16.4% vs 10.4%	17.0% vs 10.2%	10.7% vs 10.2%
KDRI full model	C-statistic (95% CI)	0.572 (0.542-0.601)	0.560 (0.495-0.625)	0.592 (0.582-0.601)	0.592 (0.582-0.601)
	Calibration slope	1.049	0.866	1.321	1.321
	Calibration-in-the-large	16.4% vs 15.2%	16.7% vs 10.4%	17.1% vs 10.2%	10.7% vs 10.2%
KDRI donor-only	C-statistic (95% CI)	0.571 (0.541-0.600)	0.559 (0.495-0.623)	0.590 (0.581-0.600)	0.590 (0.581-0.600)
	Calibration slope	1.195	0.898	1.507	1.507
	Calibration-in-the-large	16.4% vs 15.2%	16.8% vs 10.4%	15.9% vs 10.2%	10.7% vs 10.2%
Kasiske (5y) original formula	C-statistic (95% CI)	0.593 (0.574-0.612)	-	0.604 (0.598-0.610)	-
	Calibration slope	0.895	-	0.895	-
	Calibration-in-the-large	41.6% vs. 36.9%	-	41.2% vs 31.7%	-
Kasiske (1y)	C-statistic (95% CI)	0.584 (0.556-0.612)	0.547 (0.484-0.610)	0.609 (0.599-0.618)	0.609 (0.599-0.618)
	Calibration slope	0.816	0.389	0.968	0.968
	Calibration-in-the-large	16.3% vs 15.2%	16.8% vs 10.4%	16.1% vs 10.2%	10.7% vs 10.2%
Watson	C-statistic (95% CI)	0.544 (0.515-0.574)	0.538 (0.473-0.603)	0.552 (0.542-0.562)	0.552 (0.542-0.562)
	Calibration slope	0.634	0.623	0.796	0.796
	Calibration-in-the-large	16.3% vs 15.2%	16.0% vs 10.4%	15.5% vs 10.2%	10.7% vs 10.2%
Molnar (5y) original formula	C-statistic (95% CI)	0.568 (0.549-0.588)	-	0.567 (0.561-0.574)	-
	Calibration slope	0.745	-	0.588	-
	Calibration-in-the-large	14.4% vs. 36.9%	-	12.7% vs 31.7%	-

Table S6: Continued.

		NOTR 2006-2017	NOTR 2017-2018	OPTN 2006-2017	OPTN 2006-2017 with US recalibration
Molnar (1y)	C-statistic (95% CI)	0.566 (0.537-0.596)	0.575 (0.515-0.636)	0.578 (0.569-0.588)	0.578 (0.569-0.588)
	Calibration slope	0.7639	0.654	0.685	0.685
	Calibration-in-the-large	16.3% vs 15.2%	17.5% vs 10.4%	14.4% vs 10.2%	10.6% vs 10.2%
Vinson	C-statistic (95% CI)	0.598 (0.569-0.626)	0.573 (0.510-0.636)	0.611 (0.601-0.620)	0.611 (0.601-0.620)
	Calibration slope	0.927	0.583	0.907	0.907
	Calibration-in-the-large	16.3% vs 15.2%	16.5% vs 10.4%	16.6% vs 10.2%	10.7% vs 10.2%

Calibration-in-the-large is given as predicted versus observed. CI: confidence interval. The Kasiske and Molnar 5 year original formulas were not validated in the NOTR 2017-2018 as there was less than 5 years follow up.

Table S7: Example recipient-donor combinations for kidneys from deceased donors aged 50 years or older. These examples were compiled by a panel of nephrologists. For examples 1 and 2 the nephrologists indicated they would accept the kidney. For example 3 they would doubtfully accept and in example 4 they would decline the kidney offer.

	Predictors	**Example 1:** ideal scenario	**Example 2:** average scenario	**Example 3:** substandard scenario	**Example 4:** poor scenario
Donor characteristics	Age	50	65	72	75
	Gender	Male	Male	Male	Male
	BMI	22	25	27	20
	Cause of death	Trauma	Trauma	CVA	CVA
	Donation after circulatory death	No	No	Yes	Yes
	CPR performed	No	No	Yes	Yes
	Inotropes use	No	Yes	Yes	Yes
	Last serum creatinine (mmol/L)	70	80	85	150
	Proteinuria	No	No	No	Yes
	Hypertension	No	Yes	Yes	Yes
	Hypotension	No	Yes	No	No
	Diabetes mellitus	No	No	No	Yes
	Smoker	No	No	No	Yes
	Left or right kidney	Left	Left	Left	Left
	Warm ischemic time (minutes)	0	0	15	30
	Cold ischemic time (hours)	6	8	10	12
Recipient characteristics	Age	50	60	70	70
	Gender	Female	Male	Female	Male
	BMI	25	28	28	30
	Diabetes mellitus	No	No	Yes	Yes
	Cardiovascular disease	No	No	No	Yes
	Primary kidney disease	GN	GN	Diabetes	Hypertension
	Dialysis duration (months)	0	12	18	12
	Number of tprevious kidney Tx	0	0	1	1
	Total number of HLA mismatches	0	2	3	4
	Peak PRA	0%	0%	10%	10%
	AO full model, predicted risk	10.7%	30.6%	63.1%	82.6%
	AO data-driven model, predicted risk	11.1%	34.1%	58.4%	78.5%
	AO expert model, predicted risk	8.6%	30.6%	69.8%	81.4%

Abbreviations: BMI: body mass index; CPR: cardiopulmonary resuscitation; Tx: transplantation; PRA: panel reactive antibody; HLA: human leukocyte antigen; AO: adverse outcome; GN: glomerulonephritis.

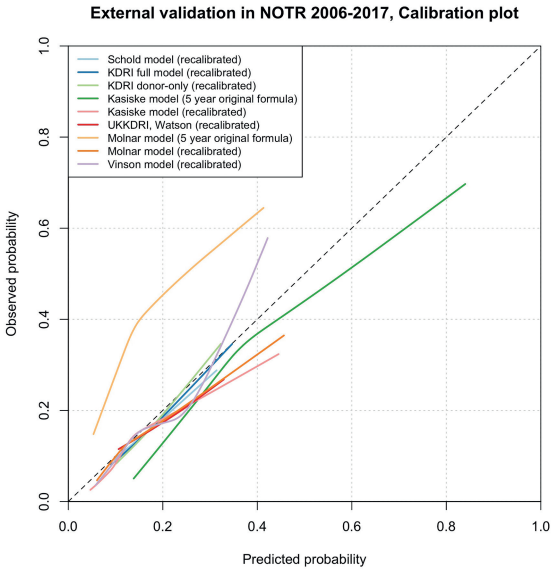

Figure S2: Calibration plot of externally validated models in the NOTR 2006-2017 (Dutch data). The models were recalibrated to the baseline risk of graft failure/recipient death in 1 year in this cohort. The full model was only given for the Kasiske and Molnar model, the calibration is shown for these full models as well as the recalibrated version.

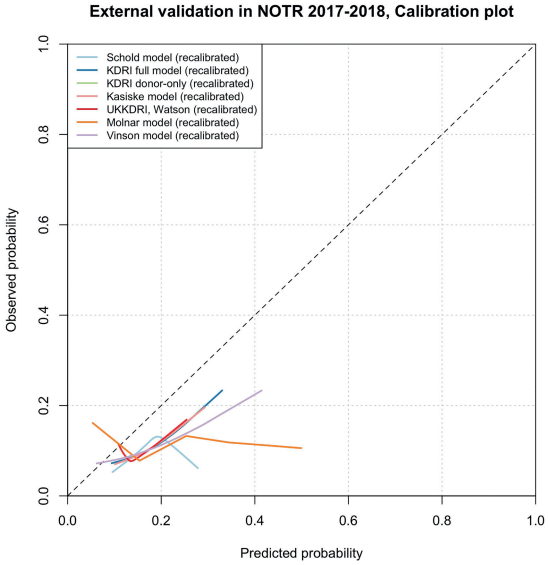

Figure S3: Calibration plot of externally validated models in the NOTR 2017-2018 (Dutch data). The models were recalibrated to the baseline risk of graft failure/recipient death in 1 year in the NOTR 2006-2017 cohort.

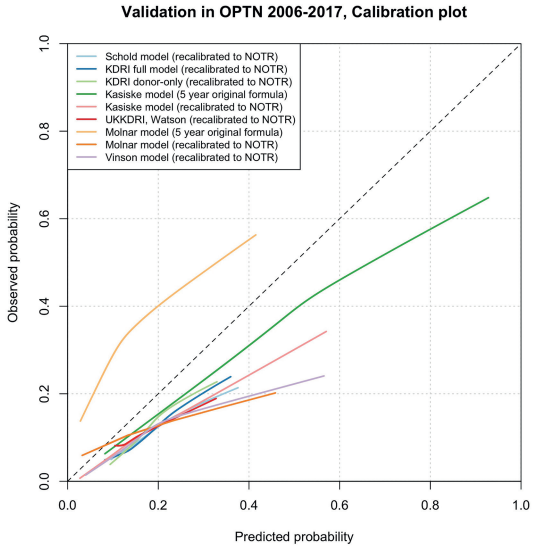

Figure S4: Calibration plot of externally validated models in the OPTN (US data). The models were recalibrated to the baseline risk of graft failure/recipient death in 1 year in the NOTR 2006-2017 cohort.

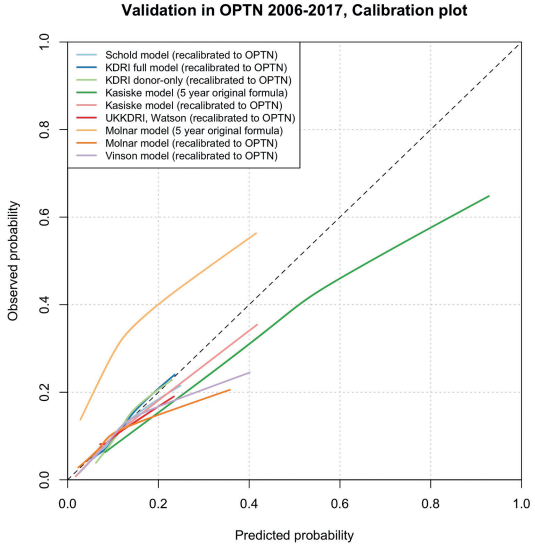

Figure S5: Calibration plot of externally validated models in the OPTN (US data). The models were recalibrated to the baseline risk of graft failure/recipient death in 1 year in the same OPTN cohort.

Table S8: AO model external validation model performance, without recalibration to the United States outcome incidence.

		External validation OPTN 2006-2017
AO full model	C-statistic (95% CI)	0.630 (0.622-0.637)
	Calibration slope	0.739
	Calibration intercept	-0.927
	Calibration-in-the-large[a]	36.4% vs 21.1%
AO data-driven model	C-statistic	0.624 (0.617-0.631)
	Calibration slope	0.796
	Calibration intercept	-0.785
	Calibration-in-the-large[a]	33.4% vs 21.1%
AO expert model	C-statistic	0.619 (0.612-0.627)
	Calibration slope	0.761
	Calibration intercept	-0.768
	Calibration-in-the-large[a]	32.4% vs 21.1%

[a]Calibration-in-the-large is given as predicted versus observed.

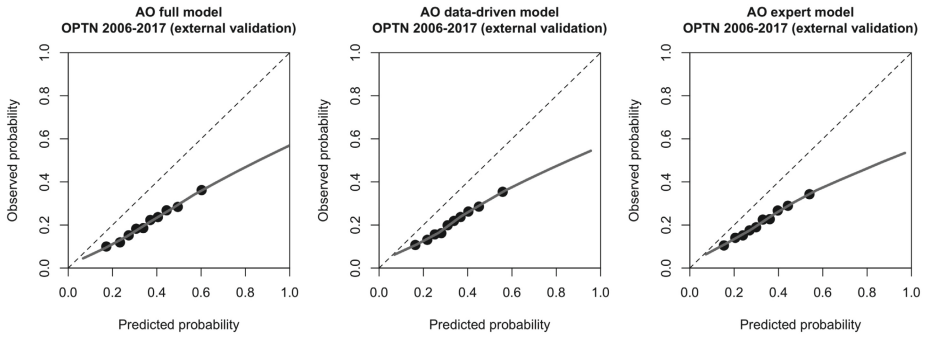

Figure S6: AO model external validation model calibration, without recalibration to the United States outcome incidence

2. Prediction model formulas:

Validated model: Schold et al.[2]

Regression coefficients provided in Table 1 of original publication.
PI = 0.136*(donor CMV+ & recipient CMV-)+0.165*(donor black)+0.378*(donor age 0-6)+ 0.303*(donorage 7-11) + 0.138*(donorage 29-39) + 0.268*(donorage 40-49) + 0.422*(donorage 50-59) + 0.676*(donorage 60-69) + 0.770*(donorage 70+) +0.089*(COD: CVA) + 0.069*(HLA-A: 2 MM)+0.111*(HLA-B:2 MM) + 0.085*(HLA-DR:1 MM) + 0.163*(HLA-DR:2MM) +0.108*(CIT 10-18h) +0.152*(CIT: 19-29h) +0.261*(CIT 30+h) + 0.138*(donor hypertension) +0.156*(donor DM)

1-year Probability (recalibrated to NOTR) = 1 – 0.8396 ^ exp(PI – 0.9723)
1-year Probability (recalibrated to OPTN) = 1 – 0.8954 ^ exp(PI – 1.013)

Proxies used in the NOTR: race was not available, everyone was assumed to be white.
Proxies used in the OPTN: -

Validated model: KDRI, Rao et al.[3]

Regression coefficients provided in Table 2 subscript of original publication.

KDRI Full model (14 predictors)

PI= -0.0194*(donorage<18y)*(donorage-18) + 0.0128*(donorage-40) + 0.0107*(donorage>50y) *(donorage-50) + 0.179*(donor black)+0.126*(donor hypertension) +0.130*(DM donor) +0.220*(s.creat-1mg/dL)-0.209*(s.creat>1.5)*(s.creat-1.5 mg/dL)+0.0881*(COD=CVA)-0.0464*((donor height -170)/10) – 0.0199*(donor weight < 80 kg)*((donorweight-80)/5) +0.133* (DCD) + 0.240*(donor HCV positive) – 0.0766*(HLA-B MM=0)-0.0610*(HLA-B MM=1) – 0.130* (HLA-DR MM=0) +0.0765* (HLA-DR MM=2) +0.00548 *(CIT-20h) -0.364*(enbloc)-0.148*(double tx)

1-year Probability (recalibrated to NOTR) = 1 – 0.8397 ^ exp(PI – 0.4159)
1-year Probability (recalibrated to OPTN) = 1 – 0.8954 ^ exp(PI – 0.4635)

Proxies used in the NOTR: race was not available, everyone was assumed to be white.
Proxies used in the OPTN: -

KDRI donor-only model (10 predictors)

PI= -0.0194*(donorage<18y)*(donorage-18) + 0.0128*(donorage-40) + 0.0107*(donorage>50y) *(donorage-50) + 0.179*(donor black)+0.126*(donor hypertension) +0.130*(DM donor)

+0.220*(s.creat-1mg/dL)-0.209*(s.creat>1.5)*(s.creat-1.5 mg/dL)+0.0881*(COD=CVA)-0.0464*((donor height -170)/10) – 0.0199*(donor weight < 80 kg)*((donorweight-80)/5) +0.133* (DCD) + 0.240*(donor HCV positive)

1-year Probability (recalibrated to NOTR) = 1 – 0.8385 ^ exp(PI – 0.5120)
1-year Probability (recalibrated to OPTN) = 1 – 0.8946 ^ exp(PI – 0.4800)

Proxies used in the NOTR: race was not available, everyone was assumed to be white.
Proxies used in the OPTN: -

Validated model: Kasiske et al.[4]

Prediction model 1: Full prediction formula provided in Table 1 of original publication.
PI= (donor age – 38)*0.0087 + (donor age -38)^2 *0.0003 + (recipient black)*0.1897 + (recipient Asian)*-0.3129 + (recipient white) * 0 + (recipient other/unkown)*-0.2033 + (preemptive tx & first tx)*-0.2847 + (<1y dialysis & first tx)*-0.0787 + (3-5y dialysis & first tx)*0.0221 + (≥5y dialysis & first tx)*0.1546 + (<9y RRT & subsequent tx)*0.3768 +(9-14y RRT & subsequent tx)*0.3350 + (≥14 RRT & subsequent tx)*0.2598 + (recipient age-50)*0.0088 + (recipient age-50)^2 *0.0006 + ((recipient age-50)*(donor age -38))*-0.0001 + (diabetic nephropathy) * 0 + (hypertension nephropathy)*-0.1718 + glomerulonephritis*-0.2523 + (cystic kidney disease)*-0.5137 + (other kidney disease)*-0.2071 + (donor HCV) *0.3929 + (donor hypertension) *0.1984 + (primary insurance medicare)*0 + (primary insurance private) *-0.1859 + (primary insurance other)*0.1003 + (trauma COD)*-0.1603 + (HLA 1-3 MM) *0.1448 + (HLA 4-6MM)*0.2347 + (HLA MM unknown)*0.3035

5 -year Probability = 1 – 0.74666^ exp(PI)

1-year Probability (recalibrated to NOTR) = 1 – 0.8455 ^ exp(PI – 0.5817)
1-year Probability (recalibrated to OPTN) = 1 – 0.9005 ^ exp(PI – 0.5587)

Proxies used in the NOTR: race was not available, everyone was assumed to be white. As the insurance system is different in the Netherlands and everyone is insured by law, everyone was filled in to have 'medicare'.
Proxies used in the OPTN: -

Validated model: UKKDRI, Watson et al.[5]

Regression coefficients provided in the results section of original publication. Donor days in hospital were truncated at 100 days, as longer times highly skewed computed probabilities.

PI = -0.245*(donorage<40)+0.396*(donorage≥60)+0.265*(history of hypertension)+0 .0253*((donorweight-75)/10)+0.00461*(donor days in hospital)+0.0465*(adrenaline administered to donor)

1-year Probability (recalibrated to NOTR) = 1 – 0.8405 ^ exp(PI – 0.3623)
1-year Probability (recalibrated to OPTN) = 1 – 0.8955 ^ exp(PI – 0.3049)

Proxies used in the NOTR: use of inotropic medication used as proxy for adrenaline administration.
Proxies used in the OPTN: use of inotropic medication used as proxy for adrenaline administration.

Validated model: Molnar et al.[6]

Cox regression model for predicting combined outcome (mortality or graft failure) with all variables (main model). Regression coefficients given in Table 4 of original publication. PI= 0.0910*(recipient age 18-34)-0.0718*(recipientage 35-49)+0.1723* (recipientage ≥ 65) + 0.1474*(rec. Hispanic) - 0.2339*(rec.black) + 0*(rec.white) - 0.4099*(rec. other/ unknown) + 0*medicare -0.1426*Medicaid – 0.3927*other – 0.3767*unknown + 0*(diabetic nephropathy) +0.1412*(hypertensive kidney disease) +0.1325*(glomerular nephritis) – 0.1713 * (cystic kidney disease)+0.2939 *(other PKD) – 0.2398*(1-3y on dialysis) – 0.3432*(3-5y on dialysis) -0.1312*(>5y on dialysis)+0.2767*(rec DM)+0.2397*(rec CAD)-0.2421*(rec s. albumin g/dL)-0.0413*(rec Hb g/dL) + 0.0054*(donor age) + 0.4210*(donor DM) – 0.303*(donor DM unkown)+0.1907*(donor ECD)+0.2969*(1/2/3 HLA MM) + 0.2853*(4/5/6 HLA MM)

5-year Probability = 1 – 0.752292^ exp(PI)
1-year Probability (recalibrated to NOTR) = 1 – 0.8438 ^ exp(PI + 0.6501)
1-year Probability (recalibrated to OPTN) = 1 – 0.9004 ^ exp(PI + 0.8054)

Proxies used in the NOTR: race was not available, everyone was assumed to be white. As the insurance system is different in the Netherlands and everyone is insured by law, everyone was filled in to have 'medicare'. Recipient albumin and recipient haemoglobin were not available, the mean albumin value (4.0) and mean haemoglobin value (12.2) in the article of Molnar were imputed for everyone.

Proxies used in the OPTN: Recipient coronary artery disease was not available; recipient peripheral vascular disease was used as proxy. Recipient albumin and recipient haemoglobin were not available, the mean albumin value (4.0) and mean haemoglobin value (12.2) in the article of Molnar were imputed for everyone.

Validated model: Vinson et al.[7]

Model 3 (independent D&R variable and DR pairing variables) validated.
HR's provided in Table 3. HRs were transformed to regression coefficients by taking the natural logarithm of the HRs. WDWR: white donor white recipient, BDBR: black donor black recipient, ODOR: other donor other recipient.

Prognostic Index = (donor age^2/1000)*0.1476 + (diabetic nephropathy)*0.1587 + (polycystic kidney disease)*-0.2705 + (glomerulonephritis) * 0 + (other kidney disease)*0.1240 + WDWR*0 + BDBR*0.2406 + ODOR*-0.3052 + WDBR*0.0980 + WDOR*-0.2383 + BDWR*0.1415 +BDOR*-0.0502 +ODWR*-0.0284 + ODBR*0.0564 + (recipient age 35-65) *-0.3092 + (recipient age>65)*0.0030 + (dialysis vintage >1y)*0.2639 + (donor HCV)*0.4675 + (recipient PVD)*0.2814 + (donor height – recipient height (meter))^2 *-0.3202 + (donor DM)* 0.2343 + (peak PRA^2/1000)*0.0286 + (recipient CAD)*0.1898 + (donor hypertension) *0.1406 + (recipient DM)*0.1856 + (HLA 1 MM)*0.1007 + (HLA 2 MM)*0.1302 + (HLA 3 MM)*0.1275 + (HLA 4 MM)*0.1467 + (HLA 5 MM)*0.1914 + (HLA 6 MM)*0.2469 + (CIT(h))*0.0060 +donorweight -recipientweight< -30)*-0.0010 +(donorweight - recipientweight = -30 – (-10))*0.0010 +(donorweight-recipientweight=10-30)*0.0469 + (donorweight-recipientweight>30)*0.1527

1-year Probability (recalibrated to NOTR) = 1 – 0.8465 ^ exp(PI – 1.0651)
1-year Probability (recalibrated to OPTN) = 1 – 0.9007 ^ exp(PI – 1.0855)

Proxies used in the NOTR: race was not available, everyone was assumed to be white.
Proxies used in the OPTN: Recipient coronary artery disease was not available; recipient peripheral vascular disease was used as proxy.

Developed adverse outcomes prediction models' formulas

Restricted cubic splines with 4 knots were used to model continuous predictors and splines were retained if the non-linear term was statistically significant in an analysis of variance. In the NOTR, recipient cardiovascular disease was defined as the occurrence of a myocardial infarction, percutaneous coronary intervention, bypass surgery, cerebrovascular accident, aortic bifurcation graft, percutaneous angioplasty, peripheral bypass surgery or amputation for vascular reasons. In the OPTN, recipient peripheral vascular disease was used as a proxy for cardiovascular disease. Furthermore, donor hypotension was not recorded and imputed to 'no' for everyone.

Adverse outcome full model Europe (Netherlands):
$\mathbf{PI}_{\text{full model}}$ = -9.2256335 + 0.14226206* (donor age) + 0.032697001* (donor BMI) + 0*(donor cause of death = trauma) + 0.39901351*(donor cause of death = CVA) 0.37023348*(donor cause of death = Anoxia) + 0.25623746*(other cause of death donor) - 0.042428701* (cold ischemic time, h) +0.00060902324*pmax((cold ischemic time, h) -7.3741667,0)^3-

0.0017966623*pmax((cold ischemic time, h) -13,0)^3+0.0013968265*pmax((cold ischemic time, h) -17.25,0)^3-0.00020918745*pmax((cold ischemic time, h) -25,0)^3-0.026253521* (CPR performed on donor) +0.023980171* (non-heartbeating donor)*(cold ischemic time, h) +0.10690675* (donor with diabetes) -0.46290564* (non-heartbeating donor) -0.078442953* (male donor) +0.094361445* (donor with hypertension) -0.095333163* (donor with hypotension previous to death) - 0.16863536* (donor admitted inotropics) -0.0031793722* (donor last serum creatinine, µmolL) +8.1093576e-06*pmax((donor last serum creatinine, µmolL) -36,0)^3-2.4275464e-05*pmax((donor last serum creatinine, µmolL) -58,0)^3+ 1.8011509e-05*pmax((donor last serum creatinine, µmolL) -74.257428,0)^3-1.8454024e-06*pmax((donor last serum creatinine, µmolL) -120,0)^3+ 0.062181839* (left kidney donated) -0.039771841*(donor had proteinuria, yes/no) +0.016255085* (donor with smoking history) +0.028227912* (first warm ischaemic time, minutes) +.07830389* (recipient age) +1.7360882e-05*pmax((recipient age) -33,0)^3-3.0810687e-05*pmax((recipient age) -54,0)^3-1.3629029e-05*pmax((recipient age) -65,0)^3+2.7078834e-05*pmax((recipient age) -73,0)^3+0.0015987313* (recipient BMI) +0.10023038* (recipient with cardiovascular disease) +0.068470786* (recipient with diabetes) +0.00010238671* (recipient number of days on dialysis) - 0.23420186* (male recipient) +0.062103814* (recipient number of previous kidney transplants received) + 0*(recipient primary kidney disease=diabetes) 0.046446946*(recipient primary kidney disease= hypertension)-0.055092072*(recipient primary kidney disease=glomerular nephritis)-0.14192488*(recipient primary kidney disease=cystic kidney disease)+0.11284731*(recipient primary kidney disease=other) - 0.0016482676*(recipient age)*(donor age)+0.18429126*(Donor height – Recipient height = -20 – (-10) cm) + 0.13241334*(Donor height – Recipient height = -10 – 10 cm) +0.23776639*(Donor height – Recipient height = 10 – 20 cm) +0.16348181* (Donor height – Recipient height^{3}20cm) - 0.33816064* (Donor weight – Recipient weight = -30 – (-10) kg)-0.61045866* (Donor weight – Recipient weight = -10 – 10 kg) -0.83696762*(Donor weight – Recipient weight = 10 – 30 kg)-0.88536114*(Donor weight – Recipient weight ^{3} 30) +0.042584943* (total number of HLA mismatches) +0.0035453836* (Peak PRA percentage)

Probability = $1/(1+\exp(-\text{PI}_{\text{full model}}))$

Adverse outcome full model North-America:
Probability = $1/(1+\exp(-(\text{PI}_{\text{full model}} - 0.7596))$

Adverse outcome data-driven model Europe (Netherlands):
$\mathbf{PI_{data\text{-}driven\ model}}$ = -9.4654947+0.150404* (donor age) +0.033207517* (donor BMI) + 0*(donor cause of death = trauma) + 0.38524276*(donor cause of death = CVA)+0.32300252*(donor cause of death = Anoxia)+0.22096086*(other cause of death donor) +0.034602112*(non-heartbeating donor)*(cold ischemic time, h) -0.65498356* (non-heartbeating donor) -0.18651035* (donor admitted inotropics) -0.0044266414* (donor last serum creatinine, µmolL)+8.3184702e-06*pmax((donor last serum creatinine, µmolL)-36,0)^3-2.4610941e-

05*pmax((donor last serum creatinine, µmolL)-58,0)^3+ 1.8082212e-05*pmax((donor last serum creatinine, µmolL)-74.257428,0)^3-1.7897404e-06*pmax((donor last serum creatinine, µmolL)-120,0)^3+0.027036962* (first warm ischaemic time, minutes) +0.083370358* (recipient age) +1.7282631e-05*pmax((recipient age)-33,0)^3-1.7011519e-05*pmax((recipient age)-54,0)^3-4.6010797e-05*pmax((recipient age)-65,0)^3+ 4.5739686e-05*pmax((recipient age)-73,0)^3 +0.11715439* (recipient with cardiovascular disease)+ 0.00011067474* (recipient number of days on dialysis)-0.27923847* (male recipient) -0.0017830732*(recipient age)*(donor age) -0.36903991* (Donor weight – Recipient weight = -30 – (-10) kg) - 0.63722528*(Donor weight – Recipient weight = -10 – 10 kg) - 0.84232781*(Donor weight – Recipient weight = 10 – 30 kg)-0.86368954*(Donor weight – Recipient weight^{3}30) +0.035743551* (total number of HLA mismatches)

Probability = 1/(1+exp(- $PI_{data\text{-}driven\ model}$))

Adverse outcome data-driven model North-America:
Probability = 1/(1+exp(-($PI_{data\text{-}driven\ model}$ – 0.6273))

Expert adverse outcome model Europe (Netherlands):
$PI_{expert\ model}$ = -9.3698629+0.14972394* (donor age) + 0*(donor cause of death = trauma) + 0.44853721*(donor cause of death = CVA)+0.37546206*(donor cause of death = Anoxia) +0.29525571*(donor cause of death=other) -0.042659043* (cold ischemic time, h) + 0.00066728464 *pmax((cold ischemic time, h)-7.3741667,0)^3-0.0019891102*pmax((cold ischemic time, h)-13,0)^3 +0.0015623064 *pmax((cold ischemic time, h)-17.25,0)^3 -0.00024048082*pmax((cold ischemic time, h)-25,0)^3+0.020926381*(non-heartbeating donor)*(cold ischemic time, h) +0.16702693* (donor with diabetes) -0.35021544*(non-heartbeating donor)+0.11818282* (donor with hypertension) -0.11182244* (donor with hypotension previous to death) -0.0062187199* (donor last serum creatinine, µmolL)+7.9753811e-06*pmax((donor last serum creatinine, µmolL)-36,0)^3-2.2735249e-05*pmax((donor last serum creatinine, µmolL)-58,0)^3+1.6169913e-05*pmax((donor last serum creatinine, µmolL)-74.257428,0)^3 - 1.4100453e-06*pmax((donor last serum creatinine, µmolL)-120,0)^3 +0.028169337* (first warm ischemic time, minutes)+0.083271651* (recipient age)+1.7906896e-05*pmax((recipient age)-33,0)^3-3.4822504e-05*pmax((recipient age)-54,0)^3-6.831031e-06* pmax((recipient age)-65,0)^3+ 2.374664e-05*pmax((recipient age)-73,0)^3+0.13772337* (recipient with diabetes) + 0.00012085464* (recipient number of days on dialysis)+0.080699178* (recipient number of previous kidney transplants received) - 0.0017254551* (recipient age)*(donor age) + 0.034202039* (total number of HLA mismatches)

Probability = 1/(1+exp(- $PI_{expert\ model}$))

Adverse outcome expert model North-America:
Probability = 1/(1+exp(-($PI_{expert\ model}$ – 0.5799))

References

1. Kabore R, Haller MC, Harambat J, et al. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2017;32(suppl_2):ii68-ii76. doi: 10.1093/ndt/gfw405 [published Online First: 2017/02/17]
2. Schold JD, Kaplan B, Baliga RS, et al. The broad spectrum of quality in deceased donor kidneys. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2005;5(4 Pt 1):757-65. doi: 10.1111/j.1600-6143.2005.00770.x [published Online First: 2005/03/12]
3. Rao PS, Schaubel DE, Guidinger MK, et al. A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. *Transplantation* 2009;88(2):231-6. doi: 10.1097/TP.0b013e3181ac620b [published Online First: 2009/07/23]
4. Kasiske BL, Israni AK, Snyder JJ, et al. A simple tool to predict outcomes after kidney transplant. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2010;56(5):947-60. doi: 10.1053/j.ajkd.2010.06.020 [published Online First: 2010/08/31]
5. Watson CJE, Johnson RJ, Birch R, et al. A simplified donor risk index for predicting outcome after deceased donor kidney transplantation. *Transplantation* 2012;93(3):314-18. doi: 10.1097/TP.0b013e31823f14d4
6. Molnar MZ, Nguyen DV, Chen Y, et al. Predictive Score for Posttransplantation Outcomes. 2017;101(6):1353-64. doi: 10.1097/tp.0000000000001326
7. Vinson AJ, Kiberd BA, Davis RB, et al. Nonimmunologic Donor-Recipient Pairing, HLA Matching, and Graft Loss in Deceased Donor Kidney Transplantation. *Transplantation direct* 2019;5(1):e414. doi: 10.1097/txd.0000000000000856 [published Online First: 2019/01/19]