Development and external validation study combining existing models and recent data into an up-to-date prediction model for evaluating kidneys from older deceased donors for transplantation

Ramspek, C.L.; Moumni, M. el; Wali, E.; Heemskerk, M.B.A.; Pol, R.A.; Crop, M.J.; ... ; Moers, C.

Development and external validation study combining existing models and recent data into an up-to-date prediction model for evaluating kidneys from older deceased donors for transplantation

OPEN

Chava L. Ramspek[1], Mostafa El Moumni[2], Eelaha Wali[1], Martin B.A. Heemskerk[3], Robert A. Pol[4], Meindert J. Crop[5], Nichon E. Jansen[3], Andries Hoitsma[3], Friedo W. Dekker[1], M. van Diepen[1] and Cyril Moers[4]

[1]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands; [2]Department of Trauma Surgery, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; [3]Research Division, Dutch Transplant Foundation, Leiden, the Netherlands; [4]Department of Surgery–Organ Donation and Transplantation, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands; and [5]Department of Internal Medicine–Nephrology, University Medical Center Groningen, Groningen, the Netherlands

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. Here, we externally validate existing prediction models in a European population of older deceased donors, and subsequently developed and externally validated an adverse outcome prediction tool. Recipients of kidney grafts from deceased donors 50 years of age and older were included from the Netherlands Organ Transplant Registry (NOTR) and United States organ transplant registry from 2006-2018. The predicted adverse outcome was a composite of graft failure, death or chronic kidney disease stage 4 plus within one year after transplantation, modelled using logistic regression. Discrimination and calibration were assessed in internal, temporal and external validation. Seven existing models were validated with the same cohorts. The NOTR development cohort contained 2510 patients and 823 events. The temporal validation within NOTR had 837 patients and the external validation used 31987 patients in the United States organ transplant registry. Discrimination of our full adverse outcome model was moderate in external validation (C-statistic 0.63), though somewhat better than discrimination of the seven existing prediction models (average C-statistic 0.57). The model's calibration was highly accurate. Thus, since existing adverse outcome kidney graft survival models performed poorly in a population of older deceased donors, novel 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.

Correspondence: *Chava L. Ramspek, Department of Clinical Epidemiology, Leiden University Medical Center, Albinusdreef 2, Leiden 2300 RC, the Netherlands. E-mail: c.l.ramspek@lumc.nl*

Received 2 July 2020; revised 11 October 2020; accepted 5 November 2020; published online 16 December 2020

Kidney International (2021) **99,** 1459–1469; https://doi.org/10.1016/j.kint.2020.11.016

KEYWORDS: adverse outcome; graft failure; kidney allocation; kidney transplantation; mortality; prediction model

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 suboptimal 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 pretransplant assessment of organ quality and selection of the best recipient-to-donor match to minimize unjust discard and maximize graft and patient survival have 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 proved to be challenging, several models have included predictors measured during transplant surgery or shortly after transplantation, such as the iBox risk score.[11] Although 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).[12] The Kidney Donor Profile Index (KDPI), derived from this KDRI, has been implemented in the new US kidney allocation system in effect since 2014.[13] 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 labeling may cause unjust and almost automatic discard of kidneys with a high KDPI.[14–17]

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 have been constructed on transplant data from the United States.[18–20] As patient populations, kidney transplant procedures, and policies differ considerably between Europe and the United States, 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. Subsequently, our aim was to improve on 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.

RESULTS

Existing prediction models

Following the systematic screening, 6 studies, presenting 7 prediction models, were considered appropriate for validation (flowchart in Supplementary 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 Prediction model Risk Of Bias ASsessment Tool (PROBAST; Supplementary Table S1).[23] Only 2 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 (Table 2). Most existing models only included donor characteristics.

Baseline characteristics

In total, 3333 transplant recipients of kidneys from deceased donors aged ≥50 years were included from the Netherlands Organ Transplant Registry (NOTR). These patients were split into a development cohort (2510 patients) and a temporal validation cohort (837 patients). From the US transplant registry (Organ Procurement and Transplantation Network [OPTN]), 31,987 recipients were included as the external validation cohort. At baseline, the OPTN data set had slightly younger donors with more diabetes and hypertension and substantially fewer donations after circulatory death (Table 3). More extensive baseline tables, including percentage of missing data and stratified by outcome, are given in Supplementary Tables S2–S4. In the NOTR development cohort, a total of 10.2% (n = 257) experienced graft failure, 6.9% (n = 172) experienced death, and 17.8% (n = 446) experienced chronic kidney disease (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) experienced death, and 17.4% (n = 146) experienced CKD stage ≥4; 4% (n = 35) were lost to follow-up. In the OPTN validation cohort 6.2% (n = 1992) experienced graft failure, 5.3% (n = 1711) experienced death, and 12.8% (n = 4094) experienced 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.

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 donors aged ≥50 years (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 (Table 4). The models' discrimination was slightly better in the OPTN data of kidneys from donors aged ≥50 years (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.*[19] Models were conservatively updated to assess the calibration, which was generally reasonable (Supplementary Table S5 and Supplementary Figures S2–S5). The best calibration was seen for the Schold model and the KDRI model.

AO models

For the newly developed AO models, the predicted outcome was a combined end point, including at least one of the following within 1 year after transplantation: graft failure, recipient death, or CKD stage ≥4. In the AO full model, all candidate predictors, predefined 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 are shown in Supplementary Table S6. In general, there was a lot of variation in ranking between individual nephrologists, although all agreed that donor age and donor serum creatinine were the most important predictors.

Table 1 | Characteristics of externally validated models

Prediction model	Time horizon, yr	Outcome	Development cohort	Population	Mean donor age, yr	Reported C-statistic	Overall risk of bias	Model information
Donor risk score by Schold *et al.* (2005)[21]	—	Death/GF	US population 1996–2002 (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),[12] 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, HRs given
Kasiske *et al.* pretransplant model (2010)[18]	5	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)[22]	9	Death/GF	European population 2000–2007 (UK Transplant Registry)	Adult kidney transplant recipients Adult deceased donors	49[a]	0.62 (IV)	High	Cox model, HRs given
Molnar *et al.* (2017)[20]	5	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)[19]	—	Death/GF	US population 2000–2014 (OPTN: SRTR data)	Single kidney only, adult recipients Deceased donors	39	0.63 (IV)	High	Cox model, HRs given

—, Unknown/not reported; GF, graft failure; HR, hazard ratio; IV, internal validation; KDRI, Kidney Donor Risk Index; OPTN, Organ Procurement and Transplantation Network; SRTR, Scientific Registry of Transplant Recipients (this data set, OPTN data that are supplemented with data from various secondary sources); USRDS, US Renal Data System (OPTN data supplemented with data from Centers for Medicare & Medicaid Services).
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.
[a]Median age.
[b]OPTN data linked to dialysis facility data.

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, respectively, and the AO expert model had slightly lower C-statistics of 0.609 and 0.619, respectively (Table 5). Calibration was generally good, although the models tended to overpredict in higher-risk patients (Table 5 and Figure 1). Without recalibration, the AO models generally overpredicted risks in the OPTN data set (Supplementary Table S7 and Supplementary Figure S6). In sensitivity analyses, we also built a model that predicts the more conventional graft survival outcome. The performance of this model was poorer than that of the AO models, although slightly better than that of most existing models (data not shown).

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, can be calculated using the formulas provided in the Supplementary Material or R-script. Both a European formula and a North American formula are provided. Risk predictions for 4 hypothetical patients are shown in Supplementary Table S8. 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 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 <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.

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 pretransplant donor and recipient characteristics. In addition, 7 existing prediction models of graft survival

Table 2 | Final predictors in developed and validated models

Predictors	AO full model	AO data-driven model	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	✓		✓							✓
Inotrope 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							✓			
Medical insurance							✓		✓	
No. of previous kidney Tx	✓		✓				✓			
Peripheral vascular disease										✓
Primary kidney disease	✓						✓		✓	✓
Serum albumin									✓	
Sex	✓	✓								
Donor-recipient										
Donor age * recipient age	✓	✓	✓				✓			
Donor-recipient CMV match				✓						
Donor-recipient ethnicity difference										✓
Donor-recipient height difference	✓									✓
Donor-recipient weight difference	✓	✓								✓
HLA mismatches	✓	✓	✓	✓	✓		✓		✓	✓
Peak PRA	✓									✓

AO, adverse outcome; BMI, body mass index; CIT, cold ischemic time; CMV, cytomegalovirus; CPR, cardiopulmonary resuscitation; DCD, donor after cardiac death; ECD, expanded criteria donor; HCV, hepatitis C virus; HLA, human leukocyte antigen; KDRI, Kidney Donor Risk Index; PRA, panel-reactive antibody; Tx: transplantation.

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 with 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

Table 3 | Baseline characteristics stratified by cohort

Characteristics	Development cohort (NOTR 2006–2017) (n = 2510)	Temporal validation cohort (NOTR 2017–2018) (n = 837)	External validation cohort (OPTN 2006–2017) (n = 31,987)
Donor characteristics			
Age, yr	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, min	17 (14–21)	15 (13–18)	18 (11–27)
Cold ischemic time, h	15.8 (5.8)	13.3 (5.7)	18.2 (9.2)
Recipient characteristics			
Age, yr	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, mo	39 (25–57)	25 (15–42)	40 (13–66)
≥1 Previous kidney transplant, %	12.9	15.2	8.9
Donor-recipient characteristics			
Total No. of HLA mismatches	3 (2–4)	3 (2–4)	5 (4–5)
Peak PRA	0 (0–0)	0 (0–0)	0 (0–13)

BMI, body mass index; DCD, donation after circulatory death; HLA, human leukocyte antigen; NOTR, Netherlands Organ Transplant Registry; OPTN, Organ Procurement and Transplantation Network; PRA, panel-reactive antibody; WIT, warm ischemia time.
Data are given as mean (SD) and median (interquartile range). Laboratory values are shown in SI units and can be converted to conventional units as follows: serum creatinine in mg/dL, multiply by 0.011.

externally validated numerous times and consistently showed a moderate discrimination with a C-statistic of around 0.62.[10,12,24] 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 United States. By doing so, the same KDPI score may translate to different absolute risks of graft failure in different years (Salkowski N, Gustafson S, Wey A, Snyder J. KDPI obscures trends in absolute donor risk [abstract]. *Am J Transplant.* 2017;17(suppl 3):A98. Available at: https://atcmeetingabstracts.com/abstract/kdpi-obscures-trends-in-absolute-donor-risk/. Accessed February 15, 2021). If, over time, more high-risk kidneys are discarded on the basis of their KDRI score, the donor pool will continue to decrease each year. So far, no prediction model has been developed that 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 several 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 health care professionals' judgment. Declined kidneys were not included in the current analyses, although these models would be used on such kidneys in

Table 4 | External validation results: the Harrel C-statistics for 1-year risk of the combined end point: graft failure and recipient death

Model	NOTR 2006–2017	NOTR 2017–2018	OPTN 2006–2017
Schold	0.562 (0.532–0.591)	0.555 (0.495–0.615)	0.577 (0.567–0.586)
$KDRI_{full\ model}$ (Rao)	0.572 (0.542–0.601)	0.560 (0.495–0.625)	0.592 (0.582–0.601)
$KDRI_{donor\text{-}only\ model}$ (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)

KDRI, Kidney Donor Risk Index; NOTR, Netherlands Organ Transplant Registry; OPTN, Organ Procurement and Transplantation Network.
Data are given as C-statistic (95% confidence interval).

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 AO 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. In addition, 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.[25] Furthermore, the 7 existing models were not specifically developed for an older donor population. By validating these in such a subpopulation, 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. Nonetheless, the AO models showed consistent performance in the recent 2017 to 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

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

Model performance measure	Development (NOTR 2006–2017)	Internal validation	Temporal validation (NOTR 2017–2018)	External validation (OPTN 2006–2017)[a]
AO full model				
C-statistic (95% CI)	0.680 (0.657–0.703)	0.646[b]	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[b]	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[b]	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

AO indicates adverse outcome; CI, confidence interval; NOTR, Netherlands Organ Transplant Registry; OPTN, Organ Procurement and Transplantation Network.
[a]Correction factor was added to the model, to recalibrate to the US outcome incidence. The results without recalibration are shown in the Supplementary Material.
[b]No CIs computed as it concerns a bootstrap shrinkage corrected C-statistic.
[c]Calibration in the large is given as predicted vs. observed.

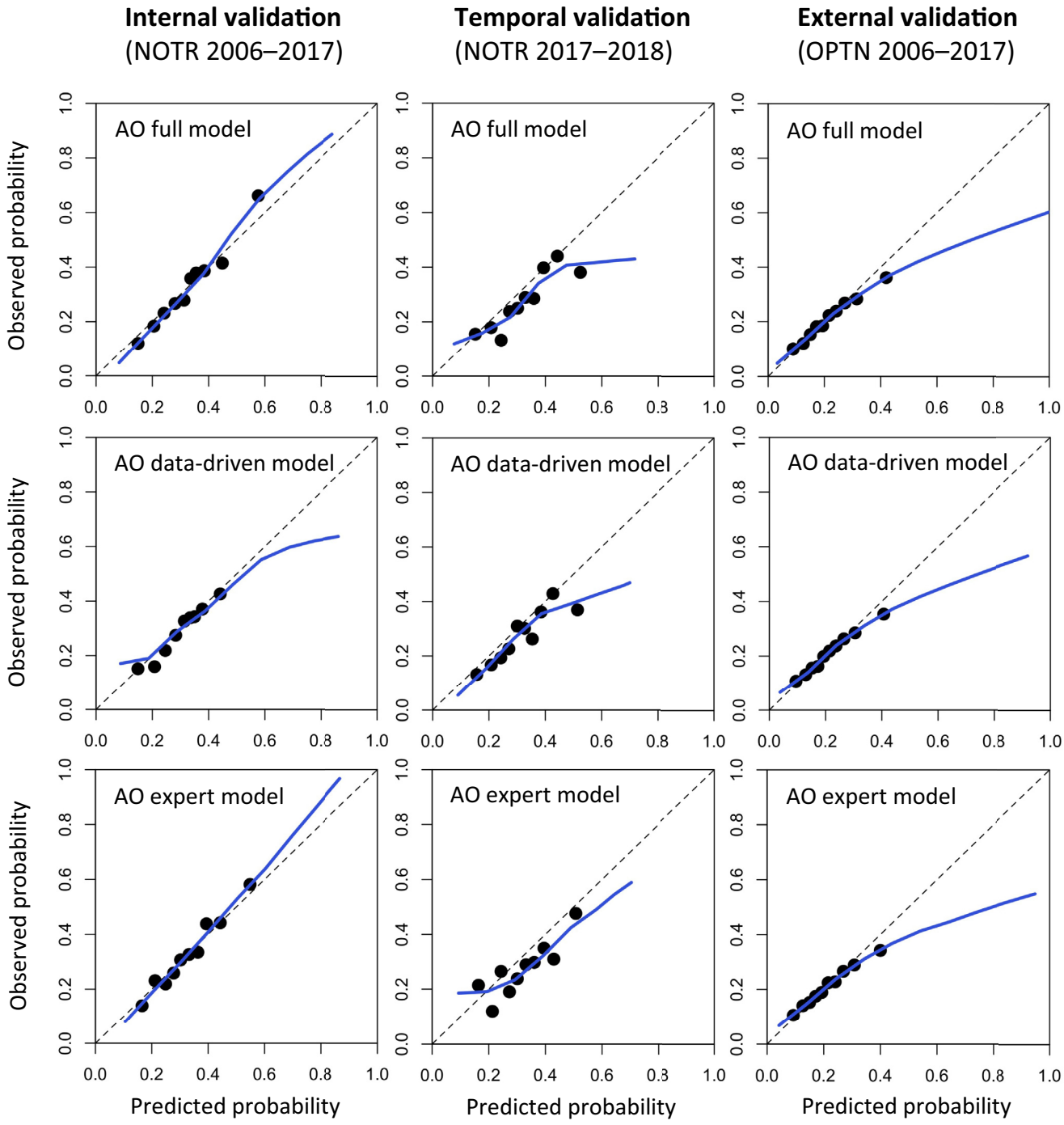

Figure 1 | Calibration plots of adverse outcome (AO) models in internal, temporal, and external validation. The external validation plots are recalibrated to the US outcome incidence. Predicted risk on the x-axis and observed risk on the y-axis per decile of predicted probability, augmented by a smoothed (lowess) regression line. The 45° dotted line indicates perfect agreement between predicted and observed risks. NOTR, Netherlands Organ Transplant Registry; OPTN, Organ Procurement and Transplantation Network.

consensus among physicians whether to accept or decline were included. Last, by including a multitude of donor and recipient characteristics, interaction terms, nonlinear 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

Table 6 | Sensitivity, specificity, PPV, and NPV for various hypothetical risk thresholds based on the developed AO full prediction model

Thresholds	Sensitivity, %	Specificity, %	PPV, %	NPV, %	No. risk < threshold	No. of false negatives	No. risk ≥ threshold	No. of false positives
$P \geq 70\%$	0.4	99.8	50.0	72.6	835	229	2	1
$P \geq 65\%$	0.9	99.3	33.3	72.6	831	228	6	4
$P \geq 60\%$	2.2	99.0	45.5	72.8	826	225	11	6
$P \geq 55\%$	4.3	98.0	45.5	73.0	815	220	22	12
$P \geq 50\%$	8.7	94.9	39.2	73.3	786	210	51	31
$P \geq 45\%$	19.1	89.1	40.0	74.5	728	186	110	66
$P \geq 40\%$	35.7	81.2	41.8	76.9	641	148	196	114

AO, adverse outcome; NPV, negative predictive value; P, predicted risk; PPV, positive predictive value.
Calculated on the temporal Netherlands Organ Transplant Registry validation cohort. The "No. risk < threshold" would be the number of accepted kidneys. The false negatives are the number of these transplanted recipients who would experience graft failure, death, or chronic kidney disease stage ≥4 within 1 year of transplantation. The "No. risk ≥ threshold" would be the number of rejected kidneys. The false positives are the number of rejected donor-recipient pairs who did not experience the outcome within 1 year.

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 the AO data-driven model. Although the full model has a slightly higher discrimination, the data-driven model is more convenient as it contains fewer 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 that 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, because of 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 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 aged ≥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 AO prediction models with maximum achievable performance for renal grafts recovered from older deceased donors.

METHODS

This study was conducted in accordance with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement (checklist in the Supplementary Material).[26,27]

Existing models

To identify existing prediction models that were suitable for external validation, a systematic search was performed, of which the details are reported in the Supplementary 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 after transplantation.

Dutch transplantation cohort

The NOTR prospectively collects data from all 8 transplant centers 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 aged ≥50 years were included. Recipients aged <18 years were excluded as were recipients of multiple-organ transplants. For the development and temporal validation of our AO models, the NOTR data set was split on the basis of 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.

US transplantation cohort

Patients who received a solitary deceased donor kidney transplant in the United States between 1 January 2006 and 1 January 2017 were included as validation cohort. Data from a United Network for Organ Sharing Standard Transplant Analysis and Research File from the OPTN, as of 1 March 2018, were used. Deceased donors aged <50 years were excluded, as were recipients aged <18 years and recipients wait-listed 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.[10,28–31] This list was presented to an expert panel of 10 nephrologists working at 4 different transplant centers 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 2 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 on the basis of 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, donation after circulatory death, and cold ischemic time were added on the basis of clinical expertise and literature.[19,32,33]

Predicted outcome

For the newly developed AO models, the predicted outcome was a combined end point, including at least one of the following within 1 year after transplantation: graft failure, recipient death, or CKD stage $\geq$4. This composite outcome was defined by an expert panel of nephrologists, transplant surgeons, and epidemiologists. The estimated glomerular filtration rate was calculated using the Chronic Kidney Disease Epidemiology Collaboration formula, and CKD stage $\geq$4 was defined as an unrecovered estimated glomerular filtration rate $<$30 ml/min per 1.73 m^2.[34] 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 SDs, or medians with interquartile ranges. Missing data were assumed to be largely missing at random, and a 10-fold multiple imputation, including all predictors and outcome, was performed.[35,36]

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 until graft failure/recipient death within 1 year. Discrimination was calculated by the Harrel 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.[37,38] 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.[39] Therefore, we recalibrated all existing models in a conservative manner by updating the baseline hazard of the outcome (updated results and model formulas are given in the Supplementary Material). In addition, the 2 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.[37] Nonlinear continuous predictors were modeled using restricted cubic splines.[40] For the data-driven models, we used a backward elimination procedure with $P < 0.157$ as stopping criterion.[41] For the AO expert model, the top-ranked predictors were entered in a logistic regression model. The 3 developed AO models were first internally validated. This internal validation was done by a 250-fold bootstrapping analysis, as recommended by the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guidelines. On the basis of the bootstrapped results, the models were adjusted for overfitting by multiplying each coefficient by a shrinkage factor (the bootstrapped slope).[42] These optimism-corrected models were subsequently validated in the Dutch temporal and US external validation cohorts. 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. In addition, certain practices may change over time, which differentiates the temporal validation cohort from the development cohort. As the incidence of AO differs between the Netherlands and the United States, the models were conservatively recalibrated for the US outcome incidence, by adding a correction factor to the model formula.[43] This improves calibration but does not affect discrimination. Discrimination was assessed in the development, internal validation, temporal validation, and external validation cohorts by calculating the C-statistic. Calibration was assessed by plotting the predicted risks against the observed risks in calibration plots. In addition, the calibration in the large, which is the average predicted risk in the entire population compared with 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 $<$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 Supplementary Material. Model performance measures and coefficients were pooled over the 10 imputation data sets, according to the rules of Rubin.[44] All analyses were performed in R version 3.6.1.

DISCLOSURE

All authors declared no conflict of interest.

ACKNOWLEDGEMENTS

The Dutch Transplant Foundation is gratefully acknowledged for providing the data, particularly Cynthia Konijn-Janssen. Part of the data reported herein has been supplied by the United Network for Organ Sharing as the contractor for the Organ Procurement and Transplantation Network (OPTN). The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the OPTN or the US government. This work was supported in part by Health Resources and Services Administration contract 234-2005-370011C. The content is the responsibility of the authors alone and does not necessarily reflect the views or policies of the Department of Health and Human

Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. This study was financially supported by a grant from the Dutch Transplantation Foundation. The work on this study by MvD was further supported by a grant from the Dutch Kidney Foundation (16OKG12).

AUTHOR CONTRIBUTIONS

CM, MeM, FWD, MvD, and CLR conceived of and designed the study. MH, NJ, and AH collected data. MeM, CR, and EW analyzed data. All authors interpreted data. CR wrote the manuscript. All authors critically revised the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary File (R-Script)

An R-script with the developed prediction models can be found as supplementary information. In addition, the file supplementary material contains the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis checklist, systematic search and PubMed search strategy, prediction model formulas, and use of proxies as well as the following tables and figures:

Supplementary File (PDF)

Table S1. Risk of bias.
Table S2. Baseline characteristic development cohort.
Table S3. Baseline characteristic temporal validation cohort.
Table S4. Baseline characteristic external validation cohort.
Table S5. External validation results for existing models.
Table S6. Expert opinion predictor ranking.
Table S7. AO model performance in OPTN without recalibration.
Table S8. Example recipient-donor combinations.
Figure S1. Flowchart.
Figure S2. Calibration plot NOTR 2006 to 2017 for existing models.
Figure S3. Calibration plot NOTR 2017 to 2018 for existing models.
Figure S4. Calibration plot OPTN for existing models.
Figure S5. Calibration plot OPTN for existing models, recalibrated.
Figure S6. AO model calibration in OPTN without recalibration.

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;341:1725–1730.
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;10:390–397.
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;74:1281–1286.
4. Tullius SG, Rabb H. Improving the supply and quality of deceased-donor organs for transplantation. *N Engl J Med*. 2018;378:1920–1929.
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;29:403–415.
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;30:14–28.
7. Cooper M, Formica R, Friedewald J, et al. Report of National Kidney Foundation Consensus Conference to Decrease Kidney Discards. *Clin Transplant*. 2019;33:e13419.
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;94:187–198.
9. Mathur AK, Ashby VB, Sands RL, Wolfe RA. Geographic variation in end-stage renal disease incidence and access to deceased donor kidney transplantation. *Am J Transplant*. 2010;10(pt 2):1069–1080.
10. Kabore R, Haller MC, Harambat J, et al. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrol Dial Transplant*. 2017;32(suppl 2):ii68–ii76.
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. *BMJ*. 2019;366:l4923.
12. Rao PS, Schaubel DE, Guidinger MK, et al. A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. *Transplantation*. 2009;88:231–236.
13. Stewart DE, Kucheryavaya AY, Klassen DK, et al. Changes in deceased donor kidney transplantation one year after KAS implementation. *Am J Transplant*. 2016;16:1834–1847.
14. Massie AB, Luo X, Lonze BE, et al. Early changes in kidney distribution under the new allocation system. *J Am Soc Nephrol*. 2016;27:2495–2501.
15. Ruggenenti P, Remuzzi G. Invited letter in response to: "Is the kidney donor profile index (KDPI) universal or UNOS-specific? *Am J Transplant*. 2018;18:1033–1034.
16. Stallone G, Grandaliano G. To discard or not to discard: transplantation and the art of scoring. *Clin Kidney J*. 2019;12:564–568.
17. Bae S, Massie AB, Luo X, et al. Changes in discard rate after the introduction of the Kidney Donor Profile Index (KDPI). *Am J Transplant*. 2016;16:2202–2207.
18. Kasiske BL, Israni AK, Snyder JJ, et al. A simple tool to predict outcomes after kidney transplant. *Am J Kidney Dis*. 2010;56:947–960.
19. Vinson AJ, Kiberd BA, Davis RB, Tennankore KK. Nonimmunologic donor-recipient pairing, HLA matching, and graft loss in deceased donor kidney transplantation. *Transplant Direct*. 2019;5:e414.
20. Molnar MZ, Nguyen DV, Chen Y, et al. Predictive score for posttransplantation outcomes. *Transplantation*. 2017;101:1353–1364.
21. Schold JD, Kaplan B, Baliga RS, Meier-Kriesche HU. The broad spectrum of quality in deceased donor kidneys. *Am J Transplant*. 2005;5(pt 1):757–765.
22. Watson CJE, Johnson RJ, Birch R, et al. A simplified donor risk index for predicting outcome after deceased donor kidney transplantation. *Transplantation*. 2012;93:314–318.
23. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170:51–58.
24. 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;102:162–170.
25. Friedewald JJ, Samana CJ, Kasiske BL, et al. The kidney allocation system. *Surg Clin North Am*. 2013;93:1395–1406.
26. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
27. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1–73.
28. Jeldres C, Cardinal H, Duclos A, et al. Prediction of delayed graft function after renal transplantation. *Can Urol Assoc J*. 2009;3:377–382.
29. 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;10:135–141.
30. 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. *Am J Transplant*. 2010;10:2279–2286.
31. 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;86:1130–1139.
32. 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;381:727–734.
33. Miller AJ, Kiberd BA, Alwayn IP, et al. Donor-recipient weight and sex mismatch and the risk of graft loss in renal transplantation 2017;. 2017;12:669–676.
34. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150:604–612.
35. de Goeij MC, van Diepen M, Jager KJ, et al. Multiple imputation: dealing with missing data. *Nephrol Dial Transplant*. 2013;28:2415–2420.
36. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16:219–242.

37. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:b604.
38. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem*. 2008;54:17–23.
39. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med*. 2014;33:517–535.
40. Harrell F. *REGRESSION MODELING STRATEGIES: With Applications to Linear Models, Logistic and Ordinal ... Regression, and Survival Analysis*. SPRINGER; 2016.
41. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med*. 2000;19:1059–1079.
42. 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*. 2012;98:683–690.
43. Janssen KJ, Vergouwe Y, Kalkman CJ, et al. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth*. 2009;56:194–201.
44. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley; 1987.