32. Rao PS, Ojo A. The alphabet soup of kidney transplantation: SCD, DCD, ECD Impact of static cold storage vs hypothermic machine preservation on ischemic kidney graft: inflammatory cytokines and adhesion molecules as markers of ischemia/reperfusion tissue damage. Our preliminary results fundamentals for the practicing nephrologist. *Clin J Am Soc Nephrol* 2009; 4: 1827–1831
33. Rodrigue JR, Paek MJ, Egbuna O *et al.* Making house calls increases living donor inquiries and evaluations for blacks on the kidney transplant waiting list. *Transplantation* 2014; 98: 979–986
34. 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
35. Snyder JJ, Salkowski N, Kim SJ *et al.* Developing statistical models to assess transplant outcomes using national registries: the process in the United States. *Transplantation* 2016; 100: 288–294.
36. Watson CJ, Wells AC, Roberts RJ *et al.* Cold machine perfusion versus static cold storage of kidneys donated after cardiac death: a UK multicenter randomized controlled trial. *Am J Transplant* 2010; 10: 1991–1999
37. Belzer FO, Ashby BS, Huang JS *et al.* Etiology of rising perfusion pressure in isolated organ perfusion. *Ann Surg* 1968; 168: 382–391
38. Poston RS, Gu J, Prastein D *et al.* Optimizing donor heart outcome after prolonged storage with endothelial function analysis and continuous perfusion. *Ann Thorac Surg* 2004; 78: 1362–1370, discussion 1362–1370
39. Xu H, Lee CY, Clemens MG *et al.* Pronlonged hypothermic machine perfusion preserves hepatocellular function but potentiates endothelial cell dysfunction in rat livers. *Transplantation* 2004; 77: 1676–1682
40. Kayler LK, Srinivas TR, Schold JD. Influence of CIT-induced DGF on kidney transplant outcomes. *Am J Transplant* 2011; 11: 2657–2664
41. Kayler LK, Magliocca J, Zendejas I *et al.* Impact of cold ischemia time on graft survival among ECD transplant recipients: a paired kidney analysis. *Am J Transplant* 2011; 11: 2647–2656
42. Stahl CC, Wima K, Hanseman DJ *et al.* Organ quality metrics are a poor predictor of costs and resource utilization in deceased donor kidney transplantation. *Surgery* 2015; 158: 1635–1641
43. Jochmans I, O'Callaghan JM, Pirenne J *et al.* Hypothermic machine perfusion of kidneys retrieved from standard and high-risk donors. *Transpl Int* 2015; 28: 665–676
44. Groen H, Moers C, Smits JM *et al.* Cost-effectiveness of hypothermic machine preservation versus static cold storage in renal transplantation. *Am J Transplant* 2012; 12: 1824–1830

Received: 14.9.2017; Editorial decision: 27.12.2017

Nephrol Dial Transplant (2018) 33: 1259–1268
doi: 10.1093/ndt/gfy019
Advance Access publication 16 February 2018

Prediction models for delayed graft function: external validation on The Dutch Prospective Renal Transplantation Registry

Jesper Kers[1,*], Hessel Peters-Sengers[2,*], Martin B.A. Heemskerk[3], Stefan P. Berger[4], Michiel G.H. Betjes[5], Arjan D. van Zuilen[6], Luuk B. Hilbrands[7], Johan W. de Fijter[8], Azam S. Nurmohamed[9], Maarten H. Christiaans[10], Jaap J. Homan van der Heide[2], Thomas P.A. Debray[11] and Fréderike J. Bemelman[2]

[1]Department of Pathology, Academic Medical Center (AMC), Amsterdam, The Netherlands, [2]Department of Internal Medicine, Renal Transplant Unit, Academic Medical Center (AMC), Amsterdam, The Netherlands, [3]Dutch Transplant Foundation, Leiden, The Netherlands, [4]Department of Nephrology, University Medical Center Groningen (UMCG), Groningen, The Netherlands, [5]Department of Nephrology, Erasmus University Medical Center (Erasmus MC), Rotterdam, The Netherlands, [6]Department of Nephrology, University Medical Center Utrecht (UMCU), Utrecht, The Netherlands, [7]Department of Nephrology, Radboud University Nijmegen Medical Center (RUNMC), Nijmegen, The Netherlands, [8]Department of Nephrology, Leiden University Medical Center (LUMC), Leiden, The Netherlands, [9]Department of Nephrology, Free University Medical Center (VUMC), Amsterdam, The Netherlands, [10]Department of Nephrology, Maastricht University Medical Center (MUMC+), Maastricht, The Netherlands and [11]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht (UMCU), Utrecht, The Netherlands

Correspondence and offprint requests to: Jesper Kers; E-mail: j.kers@amc.uva.nl
*These authors contributed equally to the study.

ABSTRACT

Background. Delayed graft function (DGF) is a common complication after kidney transplantation in the era of accepting an equal number of brain- and circulatory-death donor kidneys in the Netherlands. To identify those cases with an increased risk of developing DGF, various multivariable algorithms have been proposed. The objective was to validate the reproducibility of four predictive algorithms by Irish *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) (USA), Jeldres *et al.* (Prediction of delayed graft function after renal transplantation. Can Urol Assoc J 2009;3:377-382)

(Canada), Chapal *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) (France) and Zaza *et al.* (Predictive model for delayed graft function based on easily available pre-renal transplant variables. Intern Emerg Med 2015;10:135-141) (Italy) according to a novel framework for external validation.

Methods. We conducted a prospective observational study with data from the Dutch Organ Transplantation Registry (NOTR). Renal transplant recipients from all eight Dutch academic medical centers between 2002 and 2012 who received a deceased allograft were included (N = 3333). The four prediction algorithms were reconstructed from donor, recipient and transplantation data. Their predictive value for DGF was validated by c-statistics, calibration statistics and net benefit analysis. Case-mix (un)relatedness was investigated with a membership model and mean and standard deviation of the linear predictor.

Results. The prevalence of DGF was 37%. Despite a significantly different case-mix, the US algorithm by Irish was best reproducible, with a c-index of 0.761 (range 0.756 – 0.762), and well-calibrated over the complete range of predicted probabilities of having DGF. The US model had a net benefit of 0.242 at a threshold probability of 0.25, compared with 0.089 net benefit for the same threshold in the original study, equivalent to correctly identifying DGF in 24 cases per 100 patients (true positive results) without an increase in the number of false-positive results.

Conclusions. The US model by Irish et al. was generalizable and best transportable to Dutch recipients with a deceased donor kidney. The algorithm detects an increased risk of DGF after allocation and enables us to improve individual patient management.

Keywords: delayed graft function, external validation, kidney transplantation, net benefit analysis, prediction

INTRODUCTION

Renal transplantation is the preferred type of renal replacement therapy with respect to patient outcome as compared with long-term dialysis [1, 2]. Although improvement in short-term management of renal transplant recipients has been established over the past decades, patient and graft survival has hardly progressed due to chronic rejection and comorbidity associated with immunosuppressive medication. The incidence of infections and malignancies in renal transplant patients is manifold increased [3–5]. Dutch dialysis patients wait for more than 3 years for a deceased donor kidney and >10% of patients died while on the waiting list in 2014 [6]. This persisting shortage of donor organs has led to an upward trend in accepting older deceased donors fulfilling expanded criteria donor (ECD), including donations after circulatory death (DCD) [7]. These ECD kidneys have been shown to be a valuable source to meet the growing demand, at the expense of having an increased risk of delayed graft function (DGF) [8].

DGF is defined as the need for concomitant dialysis within the first week after transplantation. The causes for DGF are multiple and include prolonged ischaemia times, donor and recipient age, type of donation and calcineurin inhibitor pharmacodynamics (nephrotoxicity) that result in renal allograft failure [9–11]. This in turn provides an optimal inflammatory milieu to prime the alloimmune response and therefore, predicting DGF is of high importance for future graft management [12, 13]. Besides an increased risk of individual patient morbidity, also the societal impact of an increase in DGF due to extended criteria donations and DCD is enormous. The prolonged hospitalization attributed to the use of these types of donor organs results in an ∼50% increase in costs (around $70 000/person) as compared with standard criteria donors [14]. It is therefore of importance to predict those cases with an increased risk of developing DGF so as to improve individual patient management and allocate hospital resources to those patients in need of extra surveillance.

Multiple demographic and clinical donor and recipient parameters have been associated with the development of DGF. This has led to the construction of various multivariable algorithms that were shown to predict DGF. External validation of these algorithms in different patient settings (independent validation) is considered the gold standard to measure generalizability. With the use of data from the prospectively collected, nation-wide renal transplantation registry of The Netherlands [The Netherlands Organ Transplant Registry (NOTR)], we performed external validation of four algorithms according to a novel framework as proposed by Debray *et al.* [15].

MATERIALS AND METHODS

Selection of algorithms for the prediction of DGF

Until 23 September 2015, we identified 1147 articles through Medline with the search term: (Prognosis/Broad[filter]) AND ('Delayed Graft Function'[Mesh] OR 'delayed graft function'[tiab]). From these 1147 articles, four unique or updated algorithms were identified for validation [16–19].

Study population

This validation cohort comprised of adult renal transplant recipients that were transplanted in any of the eight university hospitals covering all renal transplantations in the Netherlands. Donor kidneys were acquired through allocation by the Eurotransplant allocation programme, Leiden, the Netherlands [20]. From January 2002 until January 2012, data from 3333 recipients of a deceased renal transplant were collected in the NOTR, a nation-wide prospectively collected registry by the Dutch Transplant Foundation. Patients with primary non-function were excluded from this cohort. The clinical and research activities being reported are consistent with the Principles of the Declaration of Istanbul as outlined in the 'Declaration of Istanbul on Organ Trafficking and Transplant Tourism'.

Study outcome

DGF was the primary outcome. The four identified articles that proposed an algorithm for the prediction of DGF used the same definition of the endpoint as proposed by Mallon *et al.* [21]: the use of dialysis within the first week after transplantation. Per definition, this includes a heterogeneous group of

diagnoses, including acute tubular necrosis, acute rejection (T cell-mediated and antibody-mediated rejection), acute calcineurin toxicity, early recurrent renal diseases and primary non-function due to technical complications. None of the articles differentiated between underlying disease entities that resulted in DGF. The studies by Irish *et al.* (>24 h graft survival) [18] and Chapal *et al.* (>7 days of graft survival) [19] specifically indicated that patients with primary non-function were excluded from analysis, albeit with different definitions, whereas the studies by Jeldres *et al.* [16] and Zaza *et al.* [17] did not. In the current study, cases with primary non-function, defined as definite transplant failure due to continuous need for additional renal replacement therapy (return to dialysis or retransplantation), were excluded.

Available data

The parameters necessary to calculate the four algorithms are depicted in Supplementary data, Table S1. In the NOTR database, recipient race [Organ Procurement and Transplantation Network/United Network for Organ Sharing (OPTN/UNOS) 6/350 points in the nomogram of Irish *et al.*] and recipient pre-transplantation blood transfusions (OPTN/UNOS 6/350 points) were not available. In general, very few black recipients specifically from African-American ancestry live in The Netherlands and therefore this parameter was set to 'no' in all recipients. Pretransplantation transfusion was, in concordance with the UNOS nomogram, set to 'no/unknown', which corresponds to the reference category. Supplementary data, Table S2 lists the percentage of missing data in the NOTR cohort. For 2803 of the 3333 (84.1%) patients, we had complete information available. From the articles, we extracted the exclusion criteria. Jeldres *et al.* did not mention specific exclusion criteria in their methods [16]. Irish *et al.* excluded donors younger than 16 years, recipients of multi-organ transplantations, pre-emptive renal transplantations, recipients of machine perfused grafts and living donors [18]. Zaza *et al.* excluded patients with a pre-transplantation panel reactive antibody (PRA) >20%, recipients with a body mass index (BMI) >30 kg/m^2 and pre-emptive renal transplantations [17]. The authors did not state whether they included living donations in their model as well. Lastly, Chapal *et al.* excluded patients younger than 18 years, recipients of multi-organ transplantations, pre-emptive renal transplantations, recipients of a pulsatile machine perfused graft, recipients of a living and deceased after brain death donor, donors older than 54 years, recipients who were immunized prior to transplantation and recipients on peritoneal dialysis prior to transplantation [19].

Statistical analyses

All analyses were performed within the R computing environment v3.2.1 GUI 1.66 Mavericks build (6956) in R studio v0.99.467 for Macintosh OS X 10.11.1 (www.r-project.org) with use of various packages as listed below. Differences in study characteristics between the NOTR cohort before and after application of the exclusion criteria were calculated by chi-square tests or non-parametric Mann–Whitney tests and Bonferroni-corrected P-values were reported. A full description of the methods to perform all statistical analyses are described in Supplemental R syntax and Supplementary Materials and Methods: multiple imputation chained equations [22], imputation result pooling [23], construction of calibration plots, net benefit analysis [24], Kaplan–Meier analysis, reconstruction of the OPTN/UNOS individual patient data and membership model analysis [15].

Institutional Review Board Approval

The Institutional Review Board did not review the study, because research on already available data from patients' medical records is not covered under the Dutch Medical Research Involving Human Subjects Act (WMO). The Academic Medical Center Research Code for Scientific Integrity can be found at the following link: https://www.amc.nl/web/file?uuiduid0739524-e2b3-4c29-9e65-aa97d7903433&owner=4d928ce4-e557-4531-8337-ee2442f7f9d8&contentid=17768.

RESULTS

Baseline characteristics and the indirect effects of patient exclusion

Table 1 shows the baseline characteristics of the 3333 patients in the NOTR database. To expose the hidden effects of patient exclusion, we applied the same exclusion criteria to the NOTR database (Table 1). Since Jeldres *et al.* did not describe any exclusion criteria, we did not observe a change in case-mix of the NOTR database. Based on the exclusion criteria as described in the study by Irish *et al.*, 167 of 3333 patients were excluded (5%). This indirectly led to a significant increase in the percentage of recipients that were treated with haemodialysis before transplantation (63% versus 68%, P = 0.01). The prevalence of DGF after exclusion in these patients remained unchanged (37%). Application of the exclusion criteria as described in the study by Zaza *et al.* resulted in exclusion of 1165 recipients (32%). Indirectly, this led to a higher number of patients that acquired a first renal transplant (87% versus 94%, P < 0.0001) and more male recipients (60% versus 65%, P < 0.0001). The prevalence of DGF remained unchanged (37%). Application of the exclusion criteria as described by Chapal *et al.* resulted in exclusion of 1971 patients, corresponding to 59% of the NOTR data. These exclusion criteria indirectly changed the composition of the cohort by means of a reduced prevalence of DGF (37% versus 22%, P < 0.0001) and fewer donors with antemortal anoxia (10% versus 5%, P = 0.0002).

Discrimination of the four multivariable models

We calculated the c-indices with use of the parameters as provided by the original articles [16–19]. In Table 2, we can observe that the multivariable models by Jeldres *et al.*, Zaza *et al.* and Chapal *et al.* had a lower c-statistic as compared with the original articles: Jeldres *et al.* median (range) c-statistic 0.565 (0.561–0.571), Zaza *et al.* c-statistic 0.561 (0.560–0.563) and Chapal *et al.* c-statistic 0.581 (0.574–0.594). After application of the exclusion criteria, the c-statistic worsened for Zaza *et al.*, 0.546 (0.545–0.548) and improved for Chapal *et al.*, 0.617 (0.596–0.641). A model that only included cold ischaemia time

Table 1. Characteristics of the NOTR cohort and the effects of patient exclusion on parameter distribution

Parameters	NOTR	Jeldres *et al.* [16]	Irish *et al.* [18]	Zaza *et al.* [17]	Chapal *et al.* [19]
Transplantation					
N (% of total)	3333 (100)	3333 (100)	3166 (95)	2268 (68)	1362 (41)
DGF, *n* (%)	1183 (37)	1183 (37)	1148 (37)	801 (37)	292 (22)**
HLA class I and II, no. of 0 mismatches (%)	411 (13)	411 (13)	396 (13)	269 (13)	210 (16)
HLA-A, no. of 0 mismatches (%)	1177 (37)	1177 (37)	1133 (38)	773 (36)	503 (38)
HLA-B, no. of 0 mismatches (%)	789 (25)	789 (25)	765 (25)	518 (24)	343 (26)
HLA-DR, no. of 0 mismatches (%)	1235 (39)	1235 (39)	1198 (40)	799 (37)	535 (41)
Cold ischaemia time, h, mean (95% CI)	18 (9–28)	18 (9–28)	18 (10–28)	18 (9–28)	18 (9–29)
Anastomosis time, min, mean (95% CI)	35 (19–55)	35 (19–55)	35 (19–55)	35 (19–55)	35 (19–56)
Induction therapy, *n* (%)	945 (29)	945 (29)	911 (29)	599 (27)	389 (29)
Induction therapy by ATG, *n* (%)	116 (4)	116 (4)	111 (4)	64 (3)	41 (3)
Donor					
Age, years, mean (95% CI)	47 (17–68)	47 (17–68)	48 (18–68)	48 (18–68)	48 (18–69)
Gender, males, *n* (%)	1703 (51)	1703 (51)	1623 (51)	1142 (50)	635 (47)
BMI, kg/m^2, mean (95% CI)	25 (19–33)	25 (19–33)	25 (19–33)	25 (19–33)	25 (20–33)
Donor type, no. of DBD (%)	2040 (61)	2040 (61)	1917 (60)	1330 (59)	1362 (100)
Cerebrovascular accident, *n* (%)	1830 (59)	1830 (59)	1736 (59)	1268 (59)	842 (62)
Antemortal anoxia, *n* (%)	296 (10)	296 (10)	284 (10)	215 (10)	68 (5)***
Antemortal hypotensive period, *n* (%)	905 (27)	905 (27)	852 (27)	690 (30)	394 (29)
Antemortal inotropic medication use, *n* (%)	1001 (30)	1001 (30)	955 (30)	708 (31)	464 (34)
Last serum creatinine, µmol/L, mean (95% CI)	75 (39–125)	75 (39–125)	75 (39–126)	75 (39–125)	76 (39–128)
Recipient					
First transplantation, *n* (%)	2886 (87)	2886 (87)	2730 (87)	2100 (94)**	1123 (83)
Age, years, mean (95% CI)	52 (28–71)	52 (28–71)	52 (28–71)	52 (28–71)	52 (28–71)
Gender, males, *n* (%)	1988 (60)	1988 (60)	1891 (60)	1478 (65)**	791 (58)
BMI, kg/m^2, mean (95% CI)	25 (19–33)	25 (19–33)	25 (19–33)	24 (19–29)	25 (19–33)
Haemodialisys prior to transplantation, *n* (%)	2081 (63)	2081 (63)	2073 (68)*	1429 (66)	1362 (100)
Prior dialysis time, years, mean (95% CI)	4 (1–9)	4 (1–9)	4 (1–9)	4 (1–9)	5 (1–10)
Pre-emptive transplantation, *n* (%)	155 (5)	155 (5)	0 (0)	0 (0)	0 (0)
Primary renal disease with possibility of recurrence in transplant, *n* (%)	1741 (52)	1741 (52)	1675 (53)	1344 (52)	690 (51)
Peak PRA, mean (95% CI)	13 (0–86)	13 (0–86)	13 (0–86)	9 (0–57)	18 (0–93)
Pre-transplant PRA, mean (95% CI)	4 (0–35)	4 (0–35)	4 (0–35)	1 (0–4)	6 (0–50)

*P = 0.01 versus complete NOTR cohort (after Bonferroni correction),
**P < 0.0001, ***P = 0.0002.
DBD, donation after brain death.

Table 2. Discriminative value of the four algorithms for DGF

Model	*N* included	Patients included (%)	c-statistic (ROC)	Range
CIT only	3333	100	0.567	0.543–0.593
Jeldres *et al.* [16]				
Development data set (original)	532	100	0.74	
Complete NOTR data	3333	100	0.565	0.561–0.571
NOTR data, exclusion criteria applied	3333	100	0.565	0.561–0.571
Irish *et al.* [18]				
Development data set (original)	24 337	100	0.70	
Complete NOTR data	3333	100	0.761	0.756–0.762
NOTR data, exclusion criteria applied	3166	95	0.758	0.756–0.760
Zaza *et al.* [17]				
Development data set (original)	2755	100	0.63	
Complete NOTR data	3333	100	0.561	0.560–0.563
NOTR data, exclusion criteria applied	2268	68	0.546	0.545–0.548
Chapal *et al.* [19]				
Development data set (original)	1238	100	0.73	
Complete NOTR data	3333	100	0.581	0.574–0.594
NOTR data, exclusion criteria applied	1362	41	0.617	0.596–0.641

CIT, cold ischemia time; ROC, receiver operating characteristics.

had a c-statistic of 0.567 (0.543–0.593), indicating very little to no added value of these algorithms in our hands. The more extensive model by Irish *et al.* had a c-statistic 0.761 (0.756–0.762), which was considerably higher than the c-statistic in the original study, 0.70. Application of the exclusion criteria only lowered the c-statistic marginally 0.758 (0.756–0.760).

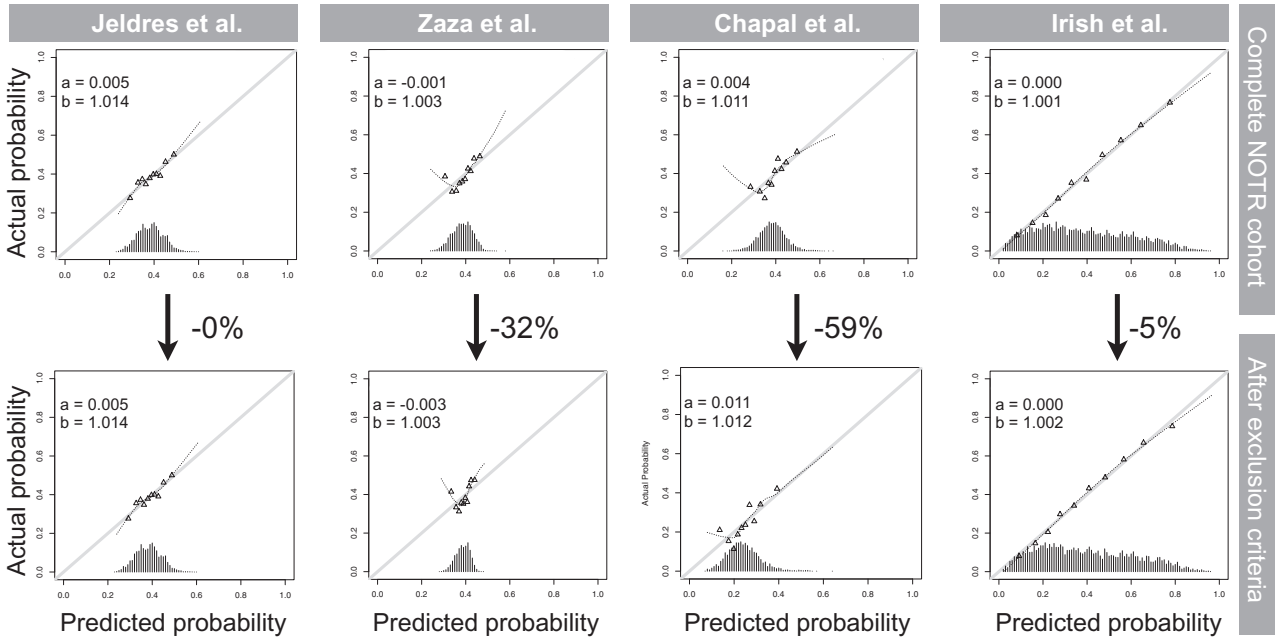

FIGURE 1: Calibration plots of the multivariable prediction models for DGF when applied to the Dutch NOTR validation cohort. The upper row shows the calibration plots of the four algorithms on the full NOTR cohort ($N = 3333$). The panels on the lower row represent the calibration plots after application of the exclusion criteria. The percentages in between the upper and the lower row indicate what percentage of patients had to be excluded. Perfect calibration is represented by the thick grey line through the origin, i.e. an intercept (a) of 0 and a slope (b) of 1. Ten quantile groups predicted probabilities were created and illustrated their corresponding outcome proportion with a triangle.

Calibration of the four multivariable models

With calibration plots, we can visualize whether the predicted and observed probabilities are appropriately scaled over the entire range of predicted probabilities. The algorithm of Irish *et al.* had the widest dynamic range of predicted probabilities (Figure 1). The intercept of the calibration curves was 0.000 and the slope was 1.001 before and 1.002 after application of the exclusion criteria, which is virtually overlapping with an ideal intercept and slope of 0 and 1, respectively, indicative of a well-calibrated model. The model by Jeldres *et al.* had a narrow range of predicted probabilities and model calibration was not as good as the model by Irish *et al.* (intercepts 0.005, calibration slopes 1.014). The models of Zaza *et al.* and Chapal *et al.* had reasonable calibration intercepts and slopes (Figure 1). Application of the exclusion criteria led to a further narrowing of the dynamic range for both models, and both were still not well calibrated. We observed a shift in the distribution of the predicted probabilities to lower risks in the model of Chapal *et al.* when we applied the exclusion criteria, which can be explained by the exclusion of 59% of patients with a relatively high risk of DGF.

Decision curve analysis for the prediction of DGF

Since the model by Irish *et al.* was best at discriminating patients with or without DGF with a good calibration, we decided to validate other predictive features of the model. Irish *et al.* performed decision curve analysis to investigate the clinical utility of the prediction model with the net benefit for different threshold probabilities [24]. If there was to be a preventive measure to mitigate the effects of DGF, the likelihood of accepting this intervention would depend on the balance between clinical benefit and adverse side effects. The threshold probability is defined as the level of risk for DGF where the expected benefit of intervention equals the expected benefit of withholding the intervention. This probability threshold is subjective by nature, depending on the likelihood of each outcome and the judgement of these outcomes by physicians or patients. The net benefit of the prediction model is calculated as the rate of true positives minus the rate of false positives weighed by the odds of DGF for the probability threshold. We compared the net benefit of the prediction model with two opposites: treat all patients with the intervention or treat none with the intervention. Figure 2 shows that the model by Irish *et al.* had a net benefit of 0.242 at a threshold probability of 0.25, which is considerably higher than the 0.089 net benefit for the same threshold in the original study (Figure 2A). A net benefit of 0.242 is equivalent to correctly identifying DGF in 24 cases per 100 patients (true-positive results) without an increase in the number of false-positive results, as compared with assuming absence of DGF in all patients. However, applying this threshold, the net benefit of the model compared with the treat-all strategy for DGF is rather small ($0.24 - 0.20 = 0.04$ net benefit). At a threshold probability of 37% of DGF, our current rate of DGF, the net benefit of the DGF model is 0.12 greater than the treat-all strategy for DGF, equivalent to identifying 21 fewer false-positive results per 100 patients (Figure 2B).

Predicted probability of DGF and graft survival

We validated whether the predicted probabilities as calculated for the model by Irish *et al.* also associated with graft

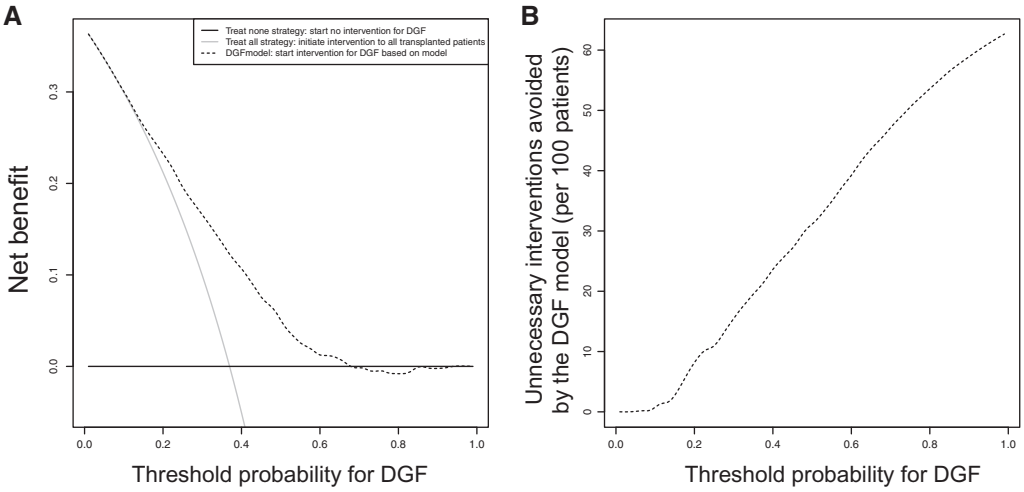

FIGURE 2: Net benefit decision curve for predicted probabilities of having DGF based on the Dutch validation cohort. **(A)** The model by Irish *et al.* [18] on the US OPTN/UNOS cohort showed that a threshold probability of 0.25 will results in a net benefit of 0.089. The net benefit of the algorithm by Irish *et al.* applied to the Dutch NOTR cohort rendered a slightly higher net benefit of 0.242 at the same threshold probability. We observed a higher net benefit for an interventional approach based on the Irish model as compared with assuming all patients have DGF and will undergo an intervention accordingly (e.g. a preventive treatment or the choice to perform an invasive transplant biopsy) between a DGF threshold probability of ∼0.10 to 0.70 **(B)** In line with this higher net benefit in the Dutch NOTR data, we calculated that in 12.7 per 100 patients, unnecessary treatment could be avoided when we would apply this algorithm to stratify patients compared with 8.8 per 100 patients in the original article by Irish *et al.* Net benefit: NB = (TP − w FP)/*N*, where TP is the number of true-positive decisions, FP the number of false-positive decisions, *N* is the total number of patients and w is a weight equal to the odds of the cut-off.

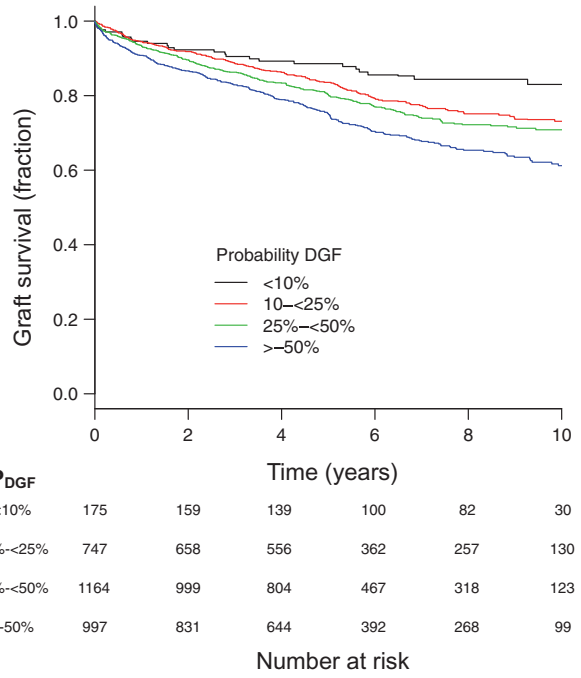

FIGURE 3: Graft survival according to the categories of predicted probabilities of DGF.

survival. For direct comparison, we chose to use the same cut-off values for the predicted probabilities to group the patients as described in the original study (<10% risk of DGF, 10–25% risk, >25–50% risk and >50% risk). Figure 3 shows the Kaplan–Meier curve. Compared with <25% risk of DGF, patients with a 25–50% risk of DGF had an increased risk of graft failure [hazard ratio (HR) = 1.25, 95% confidence interval (CI) = 1.03–1.53, P = 0.02] as well as patients with a >50% risk of DGF (HR = 1.68, 95% CI = 1.39–2.03, P < 0.0001). In our external validation study, a 2-fold increase in the predicted odds for having DGF associated with an increased risk of 23% (HR = 1.23) for graft failure compared with 30% in the original study. A deleterious effect of DGF on death-censored graft failure was only observed in old recipients who had received a graft from on old donor (both aged ≥65 years, Supplementary data, Figure S1).

Case-mix relatedness between the Dutch NOTR cohort and the OPTN/UNOS cohort

Knowledge on the relatedness between the two cohorts allows us to interpret the model performance in terms of clinical transportability. We reconstructed the individual patient data of the OPTN/UNOS cohort with a copula (see R codes in Supplementary Text S1) to compare it directly to the individual patient data of the Dutch NOTR cohort in a membership model as proposed in the framework by Debray *et al.* [15]. Supplementary data, Table S3 shows that our reconstructed OPTN/UNOS database indeed had a similar parameter distribution to the original OPTN/UNOS data. Figure 4 shows a higher mean, but lower standard deviation of the predicted risks for DGF in our validation cohort compared with the original OPTN/UNOS cohort. A membership c-statistic of 0.5 indicates complete relatedness between cohorts. The c-statistic in this study was 0.87 (Figure 4A and B), which is indicative of a difference in case-mix between the OPTN/UNOS and the NOTR cohort.

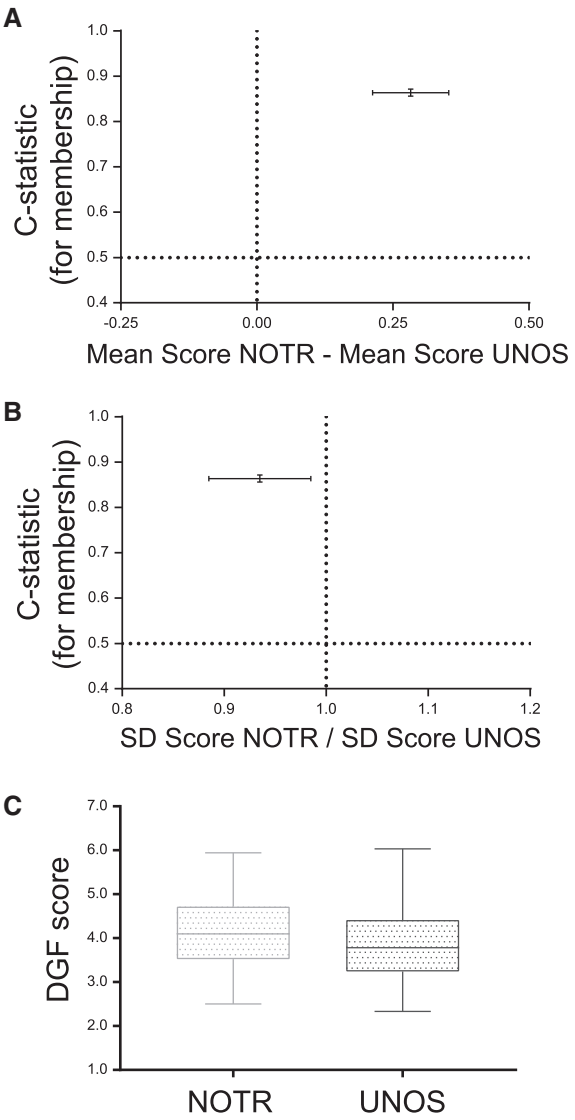

FIGURE 4: Case-mix differences between the Dutch NOTR cohort and the US OPTN/UNOS cohort. In graphs **(A)** and **(B)**, the *y*-axis reflects the extent to which the Dutch cohort is different but related to the US (OPTN/UNOS) cohort (as indicated by the c-statistic of the membership model). In graph (A), the *x*-axis reflects the difference between the means of the DGF score of the Irish *et al.* [18] model. In graph (B), the *x*-axis reflects the potential for good performance indicated by the relative difference in standard deviation of the DGF score. Panel (C) shows that the boxplot of the DGF score is higher in the Dutch cohort compared with the US cohort.

DISCUSSION

In the current study, we observed that the model by Irish *et al.* [18] performed well in Dutch patients, even though the case-mix between the USA (OPTN/UNOS) and the Netherlands (NOTR) differed substantially. Importantly, we observed a good calibration of the US model with a broad dynamic range of predicted probabilities, which indicates that also at the extremes (patients at a very low or very high risk of DGF), we can apply the US algorithm on the Dutch transplant population. The algorithm by Irish *et al.* performed equally well for the prediction of graft failure. We believe that this extensive model covers a large part of the variation among renal transplant recipients [13] and all data that are needed to construct patient-specific absolute risks are readily available at the institutes in the Netherlands (all donor-derived data are available through standardized Eurotransplant files upon transplantation) [20].

Discrepancies between the Dutch and the USA have been reported previously, also depending on life expectancy, access to dialysis, length of waiting time on the transplantation list and health insurance regulations [25]. Another explanation for the case-mix difference is that the OPTN/UNOS cohort included transplants procured between January 2003 and December 2006 while the Dutch NOTR cohort included transplants in a later era (until January 2012). Corroborating our results, a recent external validation study by Decruyenaere *et al.*, which included 497 renal transplant recipients from Belgium, described a similar predictive value by the algorithm of Irish *et al.* (c-index = 0.78) [26]. In contrast to our study, they observed that the model was not well calibrated due to an overestimation of the risk of DGF in low- to intermediate-risk patients. An explanation could be found in a case-mix difference between the Belgian and US cohorts, but this was not investigated as such. Recently, a second cohort from Belgium also validated the US model [27]. The authors performed a retrospective analysis on patients from the Antwerp University Hospital ($N = 253$) and they calculated a c-index of 0.69, which is comparable to the original US data. This shows that even within one country, there is some difference in performance of the algorithm, indicating that local protocols can influence accuracy. The model was well calibrated, although the dynamic range of prediction was small, most probably reflecting the remarkably low prevalence of DGF in this cohort (15.3%). Also in a Spanish cohort, the algorithm by Irish *et al.* performed well (c-index = 0.71) and here the model was well calibrated [28]. A somewhat lower c-index was calculated by Gourishankar *et al.* (c-index = 0.69), who studied 730 kidney transplant recipients at the University of Alberta in Canada [29]. These authors did not report statistics on model calibration and we therefore cannot compare them to those of the current study. Altogether, these studies show that the US model has comparable accuracy when validated in cohorts from various countries (c-indices ranging from 0.68 to 0.78), although differences in calibration occur. Institutions that externally validated the US algorithm on their own patients might benefit from strategies to improve model fit, which include intercept and slope updating or recalculation of the model's estimates [30, 31].

By applying all exclusion criteria that were used in the original study by Chapal *et al.*, we had to exclude 59% of our data set, which resulted, as a secondary effect, in a significant reduction of the incidence of DGF from 37% to 16%. This suggests that the algorithm that was developed in this study might be more suitable for patients with an *a priori* low risk of DGF (and other posttransplant pathology). The model did not perform as well in our cohort as compared with the original study (c-index = 0.581 in the validation cohort versus a c-index = 0.73 in the original article). The model by Chapal *et al.* was not well calibrated and we found a small dynamic range of the predicted probabilities, which limits its use in the clinical setting as well. Although we observed an increase in model performance after

application of the same exclusion criteria (area-under-the-curve = 0.617), this did not match the model performance in the original publication and calibration did not improve either. This difference in model performance could be due to kidney allocation (France is not included in the Eurotransplant programme), but most importantly, as mentioned by the authors themselves as well [19], inclusion of induction therapy with anti-thymocyte globulins (ATG) might bias their model substantially, since the indication to treat with ATG differs not only per country, but also among different centres within a country. Corroborating on this discussion, it has been shown that also in randomized studies that investigated the influence of induction therapy on the development of DGF, outcomes differed between studies [13]. The recent study by Decruyenaere *et al.* also failed to reproduce the model's performance [26]. There the authors calculated a c-index = 0.59 in 497 Belgian renal transplant recipients, which is comparable to the c-index we calculated. They found that application of the French algorithm to the Belgian patients severely overestimated the risk of DGF. Similar misclassification and miscalibration was found in another cohort from Belgium [27]. The authors calculated a c-index of 0.51 and again the model by Chapal *et al.* overestimated the risk of DGF substantially.

Now, how could we use the US prediction model for DGF in daily practice for the individual patient? The prediction model is based on donor, transplantation and recipient factors and is therefore specifically designed in patients who have already been transplanted. We want to stress that the prediction model is not designed to accept or decline a transplant offer prior to transplantation to have the best match between donor and recipient, since this requires a risk estimation for the comparison between acquiring a transplant with DGF versus remaining on dialysis. A study by Tonelli *et al.* has shown that renal transplantation in general has a better long-term cardiovascular and quality of life outcome as compared with dialysis, even when controlling for risk factors of DGF [1]. To see the potential clinical value of the US algorithm for the prediction of DGF, we should (i) consider DGF as a composite surrogate endpoint defined as the need for supportive renal replacement therapy (dialysis) within the first week after transplantation, whatever the underlying cause and (ii) consider that prevention of DGF by an intervention delays death-censored graft failure. We would like to stress that, as of yet, no drugs specifically treat all underlying causes of DGF at once, but attempts have been made to reduce DGF, for instance, in the CALLISTO trial by delayed introduction of calcineurin inhibitors (by early introduction of the mTOR inhibitor everolimus) [32]. There are currently various randomized trials with biologicals under investigation that aim at preventing DGF. The medicaments include the complement inhibitors eculizumab (NCT01403389, NCT01919346 and NCT02145182) and a C1 esterase inhibitor (NCT02134314) as well as the TLR2-antagonist OPN-305 (NCT01794663). Although the CALLISTO trial did some risk stratification to select patients at the highest risk of DGF prior to randomization, a validated, accurate and more comprehensive risk stratification tool like the US algorithm would probably be more worthwhile for individual patient risk management.

An increasingly important metric to help physicians with their decision to treat renal transplant patients with the help of prediction models is decision analysis [24]. Prediction models may lead to the early identification of patients at risk for an outcome, in this case DGF, but this is at the cost of identifying patients who in fact will not have the outcome (false positives). In decision analysis, the trade-off between benefit and harm (the net benefit) for that model is analysed over a range of threshold probabilities. In a decision curve, the net benefit for applying the intervention/treatment according to the prediction model is compared with the net benefit assuming all patients will have DGF and will undergo an intervention accordingly (e.g. perform an early biopsy or preventive treatment in all patients) versus the net benefit, under the assumption that none of the patients will have DGF and therefore no patients will be treated. For each DGF threshold probability, the decision curve shows the strategy with the highest net benefit. The DGF threshold is defined as the number of patients that a physician would have to treat to prevent DGF in one patient; at a 25% threshold probability, no more than four patients should be treated to prevent DGF in one patient. When we applied decision curve analysis for the Irish model, we observed a higher net benefit on the NOTR data (0.242) compared with the OPTN/UNOS data in the original article (0.089) for the (arbitrary) DGF threshold probability of 25% [18]. When we compare the strategy of treating patients according to the predictions of the US model to the strategy of treating all patients, we observe a positive net benefit for the prediction model-guided strategy from a DGF threshold probability of ∼10% up to 70%. This indicates that over this wide range of risk thresholds for DGF, it is always better to treat an individual patient with prior stratification based on the US model compared with treating all patients, because unnecessary treatment of patients that are not at risk for DGF can be avoided. Such risk stratification could help the treating physician to make a more evidence-based decision on the expected impact of a treatment based on the individual patient risk profile. In the end, a patient management plan that includes the harms (side effects, costs) and benefits (efficacy) of particular drugs will lead to the most informed treatment choice. Whether such a model-guided treatment strategy to prevent DGF will also lead to a delay in death-censored graft failure is not known, since this depends on the causal effect of DGF on graft failure and the dynamics of alloimmunity, infections and other post-transplant diseases on follow-up. In fact, we observed that DGF only has a deleterious effect on death-censored graft survival when old patients had received a transplant from an old donor (Supplementary data, Figure S1), which is in agreement with the literature and limits the use of DGF as a surrogate outcome [33, 34].

Our study has limitations. We used data from a registry database and we are therefore not able to check the validity of all parameters that we included. For 15.9% of the patients, we had missing data for one or more parameters. We performed multiple imputations to account for these missing data. Also, allocation algorithms differ from country to country. Although this is not a limitation of our study *per se*, but rather in general,

it makes a direct comparison among cohorts difficult. As a solution, to investigate the case-mix between cohorts, we chose to reconstruct the individual patient data with a copula, which allowed us to calculate the non-relatedness between the OPTN/UNOS and NOTR database in a membership model as proposed by Debray *et al.* [15]. We are limited by the data that are collected by the registry, e.g. we did not have data available on recipient race or pre-transplantation blood transfusions (together accounting for 12/350 points in the normogram by Irish *et al.*). These parameters could alter the individual risk by ∼5%.

In conclusion, in a prospective cohort of 3333 renal transplant patients from the eight university hospitals that perform renal transplantations in the Netherlands, we could validate the OPTN/UNOS algorithm by Irish *et al.* to predict DGF.

SUPPLEMENTARY DATA

Supplementary data are available at ndt online.

FUNDING

T.P.A.D. is financially supported by a grant from The Netherlands Organization of Health Research and Development (grant no 91617050) and consulting fees from Biogen MA Inc. F.J.B. received financial grant support from Astellas Pharma and Chiesi, received consulting fees from Chiesi and lecture fees from Baxter. A.D.v.Z. received grant and travel support from Astellas Pharma and consulting fees from Novartis.

CONFLICT OF INTEREST STATEMENT

The results presented in the paper have not been published previously.

REFERENCES

1. Tonelli M, Wiebe N, Knoll G *et al.* Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. *Am J Transplant* 2011; 11: 2093–2109
2. 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
3. Lamb KE, Lodhi S, Meier-Kriesche H-U. Long-term renal allograft survival in the United States: a critical reappraisal. *Am J Transplant* 2011; 11: 450–462
4. Meier-Kriesche HU, Schold JD, Kaplan B. Long-term renal allograft survival: have we made significant progress or is it time to rethink our analytic and therapeutic strategies? *Am J Transplant* 2004; 4: 1289–1295
5. Meier-Kriesche H-UU, Schold JD, Srinivas TR *et al.* Lack of improvement in renal allograft survival despite a marked decrease in acute rejection rates over the most recent era. *Am J Transplant* 2004; 4: 378–383
6. Nederlandse Transplantatie Stichting. Nederlandse Transplantatie Stichting: jaarverslag 2014 [Internet]. Leiden; 2015 [cited 23 February 2016]. http://www.transplantatiestichting.nl/sites/default/files/product/downloads/nts_jaarverslag_2014.pdf
7. Metzger RA, Delmonico FL, Feng S *et al.* Expanded criteria donors for kidney transplantation. *Am J Transplant* 2003; 3: 114–125
8. Snoeijs MGJ, Winkens B, Heemskerk MBA *et al.* Kidney transplantation from donors after cardiac death: a 25-year experience. *Transplantation* 2010; 90: 1106–1112
9. Shoskes DA, Cecka JM. Deleterious effects of delayed graft function in cadaveric renal transplant recipients independent of acute rejection. *Transplantation* 1998; 66: 1697–1701
10. Wu WK, Famure O, Li Y *et al.* Delayed graft function and the risk of acute rejection in the modern era of kidney transplantation. *Kidney Int* 2015; 88: 851–858
11. Boom H, Mallat MJ, de Fijter JW *et al.* Delayed graft function influences renal function, but not survival. *Kidney Int* 2000; 58: 859–866.
12. Leemans JC, Kors L, Anders H-J *et al.* Pattern recognition receptors and the inflammasome in kidney disease. *Nat Rev Nephrol* 2014; 10: 398–414
13. Siedlecki A, Irish W, Brennan DC. Delayed graft function in the kidney transplant. *Am J Transplant* 2011; 11: 2279–2296
14. Saidi RF, Elias N, Kawai T *et al.* Outcome of kidney transplantation using expanded criteria donors and donation after cardiac death kidneys: realities and costs. *Am J Transplant* 2007; 7: 2769–2774
15. Debray TPA, Vergouwe Y, Koffijberg H *et al.* A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015; 68: 279–289
16. Jeldres C, Cardinal H, Duclos A *et al.* Prediction of delayed graft function after renal transplantation. *Can Urol Assoc J* 2009; 3: 377–382
17. 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
18. 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
19. 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
20. Persijn GG, Smits JM, Frei U. Eurotransplant kidney allocation. *Lancet* 2000; 355: 71
21. Mallon DH, Summers DM, Bradley JA *et al.* Defining delayed graft function after renal transplantation: simplest is best. *Transplantation* 2013; 96: 885–889
22. Buuren SV, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011; 45: 1–67
23. Rubin DB. Introduction and Summary of Repeated-Imputation Inferences. In: Rubin DB (ed). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons, Inc., 1987, 1–26.
24. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; 352: i6
25. Gondos A, Döhler B, Brenner H *et al.* Kidney graft survival in Europe and the United States. *Transplant J* 2012; 0: 1
26. Decruyenaere A, Decruyenaere P, Peeters P *et al.* Validation in a single-center cohort of existing predictive models for delayed graft function after kidney transplantation. *Ann Transplant* 2015; 20: 544–552
27. Michalak M, Wouters K, Fransen E *et al.* Prediction of delayed graft function using different scoring algorithms: a single-center experience. *World J Transplant* 2017; 7: 260–268
28. Rodrigo E, Miñambres E, Ruiz JC *et al.* Prediction of delayed graft function by means of a novel web-based calculator: a single-center experience. *Am J Transplant* 2012; 12: 240–244
29. Gourishankar S, Grebe SO, Mueller TF. Prediction of kidney graft failure using clinical scoring tools. *Clin Transplant* 2013; 27: 517–522
30. Vergouwe Y, Nieboer D, Oostenbrink R *et al.* A closed testing procedure to select an appropriate method to update prediction models. *Stat Med* 2014; 36: 4529–4539
31. Nieboer D, Vergouwe Y, Ankerst DP *et al.* Improving prediction models with new markers: a comparison of updating strategies. *BMC Med Res Methodol* 2016; 16: 128
32. Dantal J, Berthoux F, Moal M-C *et al.* Efficacy and safety of de novo or early everolimus with low cyclosporine in deceased-donor kidney transplant recipients at specified risk of delayed graft function: 12-month results of a randomized, multicenter trial. *Transpl Int* 2010; 23: 1084–1093

33. Gill J, Dong J, Johnston O *et al.* Delayed graft function increases the risk of early death post-transplant in elderly kidney transplant recipients [abstract]. *Am J Transpl* 2013; 13 (Suppl 5). http://atcmeetingabstracts.com/abstract/delayed-graft-function-increases-the-risk-of-early-death-post-transplant-in-elderly-kidney-transplant-recipients/

34. Sola R, Alarcon A, Jimenez C *et al.* The influence of delayed graft function. *Nephrol Dial Transplant* 2004; 19: iii32–iii37

Received: 19.10.2017; Editorial decision: 8.1.2018

Nephrol Dial Transplant (2018) 33: 1268–1277
doi: 10.1093/ndt/gfx240
Advance Access publication 7 September 2017

ORIGINAL ARTICLE

Meta-analysis of cognitive functioning in patients following kidney transplantation

Paras Joshee[1], Amanda G. Wood[2,3], Eleri R. Wood[4] and Elizabeth A. Grunfeld[5,6]

[1]School of Psychology, University of Birmingham, Birmingham, UK, [2]School Life and Health Sciences & Aston Brain Centre, Aston University, Birmingham, UK, [3]Murdoch Childrens Research Institute, Clinical Sciences, Flemington Road, Parkville, VIC, Australia, [4]Kings College Hospital NHS Trust, London, UK, [5]Centre for Technology Enabled Health Research, Coventry University, Coventry, UK and [6]School of Psychological Sciences, Birkbeck, University of London, UK

Correspondence and offprint requests to: Amanda Wood; E-mail: a.wood4@aston.ac.uk

ABSTRACT

Background. There is mixed evidence regarding the nature of cognitive function in patients who have undergone renal transplantation. The aim of this meta-analysis was to examine which cognitive domains are impacted following kidney transplantation and how performance compares with non-transplanted patients or healthy controls/normative data.

Method. A systematic search was conducted using keywords within three databases (Embase, MEDLINE and PsychINFO), yielding 458 unique studies, 10 of which met the inclusion criteria. Neuropsychological tests were grouped into nine cognitive domains and three separate analyses were undertaken within each domain: (i) within subjects pre- versus post-transplant, (ii) transplanted versus non-transplanted patients and (iii) transplanted versus healthy matched controls and standardized normative data.

Results. Transplanted patients showed moderate to large improvements in the domains of general cognitive status ($g = 0.526$), information and motor speed ($g = 0.558$), spatial reasoning ($g = 0.376$), verbal memory ($g = 0.759$) and visual memory ($g = 0.690$) when compared with their pre-operative scores. Test scores in the same five domains were significantly better in post-transplanted patients when compared with dialysis-dependant or conservatively managed chronic kidney disease patients. However, post-transplanted patients' performance was significantly low compared with that of healthy controls (and standardized normative data) in the domains of executive functioning ($g = -0.283$), verbal fluency ($g = -0.657$) and language ($g = -0.573$).

Conclusions. Two key issues arise from this review. First, domain-specific cognitive improvement occurs in patients after successful transplantation. Nevertheless, transplanted patients still performed significantly below healthy controls in some domains. Second, there are important shortcomings in existing studies; the length of follow-up is typically short and only limited neuropsychological test batteries are employed. These factors are important in order to support the recovery of cognitive function among patients following renal transplant.

Keywords: chronic kidney disease, cognition, dialysis, kidney transplant, systematic review

INTRODUCTION

Chronic kidney disease (CKD) can result from multiple factors, including hypertension, diabetes and genetic disorders [1–3]. End-stage kidney disease (ESKD) is the fifth and most severe category of CKD and is defined as the inability of the kidneys to metabolize and remove waste substances such as creatinine [4]. Patients suffering from ESKD depend on renal replacement therapy (RRT) for survival. RRT options include kidney transplantation, peritoneal dialysis (PD) and haemodialysis (HD) [5] and transplantation is currently the recommended gold standard [6]. The majority of transplanted patients spend a period of