NEPHROLOGY

Check for updates

Nephrology **24** (2019) 855–862

Original Article

Development and future deployment of a 5 years allograft survival model for kidney transplantation

DEREK A DUBAY,[1] ZEMIN SU,[2] THOMAS A MORINELLI,[1] PRABHAKAR BALIGA,[1] VINAYAK ROHAN,[1] JOHN BIAN,[2] DAVID NORTHRUP,[3] NICOLE PILCH,[4] VINAYA RAO,[1] TITTE R SRINIVAS,[5] PATRICK D MAULDIN[2] and DAVID J TABER[1,6]

[1]Department of Surgery, [2]Division of General Internal Medicine and Geriatrics, [3]Office of the Chief Information Officer, Medical University of South Carolina, [4]Department of Clinical Pharmacy and Outcomes Sciences, College of Pharmacy, Medical University of South Carolina, [6]Department of Pharmacy, Ralph H Johnson VAMC, Charleston, South Carolina, and [5]Intermountain Transplant Services, Salt Lake City, Utah, USA

KEY WORDS:
data analysis, decision support technique, kidney, graft survival, transplant.

Correspondence
Dr Derek A DuBay, MD, Associate Professor of Surgery, Division Director, Transplant Surgery, Department of Surgery, 96 Jonathan Lucas, MSC 613, Office #409, Charleston, SC 29464, USA. Email: dubay@musc.edu

Accepted for publication 4 September 2018.
Accepted manuscript online 10 September 2018.

doi: 10.1111/nep.13488

ABSTRACT:

Aim: **Identifying kidney transplant patients at highest risk for graft loss prior to loss may allow for effective interventions to improve 5 years survival.**

Methods: **We performed a 10 years retrospective cohort study of adult kidney transplant recipients (n = 1747). We acquired data from electronic health records, United Network of Organ Sharing, social determinants of health, natural language processing data extraction, and real-time capture of dynamically evolving clinical data obtained within 1 year of transplant; from which we developed a 5 years graft survival model.**

Results: **Total of 1439 met eligibility; 265 (18.4%) of them experienced graft loss by 5 years. Graft loss patients were characterized by: older age, being African–American, diabetic, unemployed, smokers, having marginal donor kidneys and cardiovascular comorbidities. Predictive dynamic variables included: low mean blood pressure, higher pulse pressures, higher heart rate, anaemia, lower estimated glomerular filtration rate peak, increased tacrolimus variability, rejection and readmissions. This Big Data analysis generated a 5 years graft loss model with an 82% predictive capacity, *versus* 66% using baseline United Network of Organ Sharing data alone.**

Conclusion: **Our analysis yielded a 5 years graft loss model demonstrating superior predictive capacity compared with United Network of Organ Sharing data alone, allowing post-transplant individualized risk-assessed care prior to transitioning back to community care.**

SUMMARY AT A GLANCE

The study focusses on the development of a predictive model for 5 years overall allograft survival using big data approach, using registry data as well as granular and dynamic data from electronic health care records. This model has over an 80% predictive capacity in determining 5 years allograft survival and may be a useful adjunct to standard post-transplant care. Validation of this predictive model in other cohorts will be essential.

The current regulatory environment places a high priority on short-term graft survival, with corresponding improvements in the 1 year renal graft survival.[1,2] Current induction and maintenance immunosuppressive regimens have resulted in 1 year graft survival reaching 90–97%, dependent upon deceased or living donor organ source.[3–5] In the United States, 1 year graft survival is comparable to other western countries, whereas long-term graft survival (3–10 years) is significantly worse.[4] The aetiology of this outcome discrepancy is not clearly understood, but may be due to less rigorous long-term post-transplant care provided by transplant centres in concert with increased reliance upon primary care physicians.[6,7] To improve long-term outcomes within the United States, there is a need to accurately identify patients at higher risk of graft loss and focus interventions on this vulnerable population.[3,4] Reliable predictors of long-term graft loss are the first step in developing and testing effective interventions aimed at improving long-term graft survival.

We previously developed predictive models for 1- and 3 years graft survival using a Big Data approach.[8] This model used dynamically evolving clinically relevant patient-level information along with administrative data; taking advantage of the vast array of information available in the electronic health record (EHR) by using real-time electronic capture and abstraction of unstructured data fields through natural language processing (NLP) to improve predictability of early patient and graft survival. While this Big Data approach to analysis improved predictability, many 1- and 3 years graft loss and patient mortality were random events; thus, the clinical application of the model was limited. In the current study, we incorporated a Big Data approach into an analytical model using data obtained up to 1 year post-transplant to develop robust predictive models for 5 years graft loss. Our hypothesis is that graft loss 1–5 years post-renal transplant is a predictable event involving mutable factors.

METHODS

Patient population

This is a retrospective cohort study of adult (≥18 years of age) solitary kidney transplant recipients at the Medical University of South Carolina (MUSC), Charleston, SC during the time period from 1 January 2007 to 30 June 30 2017. We chose this period of time as a consequence of the accuracy and the availability of EHR and administrative records for data capture. Patients were excluded if they were less than 18 years old, non-renal transplant, experienced graft loss or death in the first year or had less than 1 year follow-up time. Patients with graft failure during the first year post-transplant were excluded from this study because the aim of this study was to predict late graft failure and because of early graft loss is rare and usually due to very early postoperative complications related to the donor, surgery or cardiovascular (CV) events that occurred in the peri-operative time period. Patients with less than 1 year follow-up were excluded because of the necessary construction of the dynamic variables. The MUSC Institutional Review Board for Human Research approved this study (#00064075).

Data sources

Structured data were acquired from EHR using Practice Partner (McKesson, Seattle, WA) prior to May 2012 and Epic® (Epic Corp, Madison, WI) from July 2011 onwards. Elements from the (United Network of Organ Sharing (UNOS)) database containing Organ Procurement and Transplantation Network data were acquired since 1986. The key social determinants of health were obtained from the transplant database Velos® (Velos, Inc., Fremont, CA) before September 2014 and Epic® afterwards. NLP was used to extract text Banff data from biopsies and vital signs found in records that predated electronic capture.

Primary outcome measures

The primary outcome for this study was 5 years graft loss. Graft loss data were retrieved from internal records and UNOS files. A graft loss event was defined as a return to chronic dialysis, re-transplantation, or death. For patients who received more than one transplant, multiple graft losses per patient were considered unique observations. However, death, which was considered a graft loss, was linked to the most recent transplant event. A 1 year exposure period was used to derive 5 years graft loss model (i.e. the model was run only with cumulative data up to 365 days post-transplant).

Covariates

United Network of Organ Sharing data elements was used for key donor and recipient demographics and transplant-related variables in accordance with the published methodology used by the Scientific Registry of Transplant Recipients. These data, in concert with the MUSC EHR, were used to extract key social determinants of health. EHR data were also used to identify patient comorbidities, vital signs, CV events, laboratory data, transplant length of stay and post-transplant acute care utilization data (including both inpatient and emergency department visits).

Using enhanced International Classification of Diseases (ICD-9-CM and ICD-10-CM) codes, patient comorbidities were derived from a modified Elixhauser coding algorithm using select Charlson comorbidities.[9] Post-transplant CV events, such as arrhythmias and myocardial infarction were included collectively in determination of CV risk. Immunological risk was captured by rejection rates (Banff scores; defined as ≥1A), s (CMV) infection, BK virus (BK) infection (viral load ≥500 copies/mL), and tacrolimus trough concentrations. Social determinants of health were summarized as demographics, education level and income variables.

Statistical analysis

Event data were captured beginning at transplant date and extending to 365 days post-transplant, which included data up to 24 h prior to a death and/or Graft Loss (GL) event. Means, standard deviations, maximums and the slope of the regression lines were used to represent dynamic variables, capturing effects of change, direction of change and magnitude of change throughout the 1 year post-transplant exposure period. This treatment was applied to estimated glomerular filtration rates (eGFR), pulse rates, blood pressures, glucose, tacrolimus and haemoglobin (HGB) levels. eGFR and HGB within the first week post-transplant were excluded because of inherent post-transplant fluctuations in

values prior to patient stabilization. A multivariable Cox regression model was developed using baseline and follow-up data obtained up to 365 days post-transplant exposure period to develop a 5 years graft loss model. Statistical significance was determined at the two-sided 5% level. The Harrell's concordance[10,11] and time-dependent receiver operating characteristic (ROC) and area under the curve (AUC)[12] were used to summarize the predictive accuracy of the fitted model. IBM Watson Content Analytics Suite IBM SPSS Modeler (Version 17), IBM SPSS Statistics – Essentials for R (IBM Corporation, Armonk, New York, USA) and SAS 9.4 were used for the statistical analysis.

RESULTS

Study population and baseline characteristics

Of the 1747 kidney transplants performed between 1 January 2007 and 30 June 2017, 1439 unique transplant events met eligibility criteria to be included in the analysis. There were 265 graft loss events (18.4%) during the 5 years post-transplant follow-up time period (see Fig. 1 for the Consort diagram depicting how the cohort was developed). Demographic and clinical characteristics of patients included in the study, stratified by graft loss, are summarized in Table 1. Adjusted risk factors for graft loss are displayed in Table 2. Significant baseline risk factors for graft loss included: male gender, unemployment, smokers, comorbidities, receiving a marginal donor kidney and living greater than 200 miles from the transplant centre.

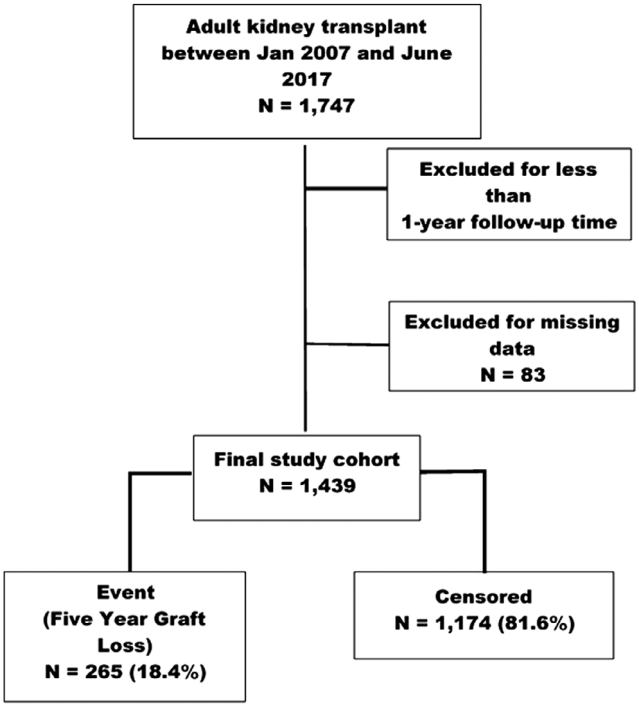

Fig. 1 Consort diagram for study enrolment.

Clinical variables during the first year follow-up as predictors of graft loss

Several clinical variables obtained from the EHR during the first year post-transplant were significant risk factors for 5 years graft loss (Table 2). A low mean blood pressure, higher pulse pressures, a higher mean heart rate, anaemia, decreasing HGB slope, lower eGFR peak, reduced eGFR slope, increased tacrolimus variability and number of post-transplant readmissions were significant and independent predictors of 5 years graft loss. A positive unit rate of HGB change was associated with a 28% decreased risk of graft loss. Acute rejection within 1 year post-transplant was significantly more common in those that developed graft loss within 5 years of transplant (11% *vs* 3%, $P < 0.001$), and associated with more than twice the risk of graft loss. The 5 years graft loss group also demonstrated higher rates of rejection for each of the acute rejection classifications (Table S1, Supporting information).

Model predictability performance

The 5 years Cox survival model was tested for its discriminatory ability using the C-statistic and construction of time-dependent ROC curves. The findings of this assessment indicate that the model has strong predictive ability and is capable of identifying those at high risk of graft loss in a sensitive manner. The overall Harrell's concordance from this model was 0.759 (SE = 0.018) and the integrated AUC (iAUC) over time was 0.819. Thus, by using baseline and 1 year post-transplant clinical data, the model was capable of accurately identifying 8 out of every 10 patients that developed graft loss within 5 years of transplant. The utilization of clinical data electronically captured and abstracted through the EHR using NLP significantly improved the predictive performance of the model, as compared with a model using only baseline UNOS data (Fig. 2). Finally, internal validation was conducted for the final model using bootstrapping (1000 replicates) and an unrestricted random sampling method. The mean of bootstrapped iAUC was 0.823 (95% CI = 0.791–0.852), indicating this model was stable with minimal bias.

DISCUSSION

The significant findings of this study demonstrate that combining donor and recipient baseline variables (i.e. UNOS model variables) with granular EHR data generated during the first year post-transplant produced a robust model capable of predicting more than 80% of patients that will have graft loss within 5 years of transplant. This Big Data approach used traditional transplant registry donor and recipient variables along with detailed sociodemographic and clinical elements obtained from EHR data, including

Table 1 Patient characteristic comparison

Variable	5 years graft loss (*N* = 265)	Censored (*N* = 1174)	*P* value
UNOS			
Age at Transplant (years; mean ± SD)	54 ± 14	51 ± 14	0.005
Categorical age at transplant			0.005
Age < = 40	20%	23%	0.202
Age = 41–60	41%	47%	0.049
Age > 60	40%	29%	0.001
Female	38%	40%	0.428
African–American	57%	54%	0.521
Deceased donor type	89%	83%	0.018
African–American donor	33%	26%	0.020
KDPI (mean ± SD)	52 ± 31	39 ± 29	<0.001
Categorical KDPI			<0.001
KDPI = 1% –20%	20%	34%	<0.001
KDPI = 21% –84%	61%	60%	0.633
KDPI = 85%–100%	19%	7%	<0.001
Blood type B	17%	16%	0.708
Waiting time (years)	1.8 ± 1.7	1.7 ± 1.6	0.337
Delayed graft function	18%	15%	0.117
Diabetes	41%	34%	0.033
BMI (mean ± SD)	28.8 ± 5.7	29.0 ± 5.4	0.497
Categorical BMI			0.907
BMI < 20	5%	4%	0.689
BMI = 20–35	80%	81%	0.732
BMI > = 35	15%	15%	0.882
Private insurance	22%	28%	0.072
Distance to MUSC_km (mean ± SD)	249 ± 172	238 ± 196	0.358
Categorical distance to MUSC			0.295
Distance <80 km	23%	26%	0.399
Distance = 80–161 km	9%	11%	0.484
Distance = 161–322 km	38%	40%	0.661
Distance > = 322 km	29%	23%	0.063
Previous kidney transplant	11%	8%	0.135
Velos ± EHR			
Finish high school	85%	87%	0.243
Employed	25%	33%	0.013
Received disability	50%	47%	0.338
Married	59%	62%	0.480
Smoker	12%	6%	0.001
EHR			
Congestive heart failure	17%	12%	0.036
Peripheral vascular disorders	14%	11%	0.231
Cerebrovascular disease	5%	6%	0.644
Cardiac arrhythmias	30%	22%	0.005
Valvular disease	9%	9%	0.767
Hypertension	96%	96%	0.931
Alcohol abuse	5%	3%	0.201
Drug abuse	4%	3%	0.328
Depression	13%	12%	0.891
Myocardial infarction	12%	5%	<0.001
Transplant LOS_days (mean ± SD)	4.0 ± 3.9	3.4 ± 2.3	0.024
Transplant LOS > 3 days	26%	20%	0.057
Acute MI[‡]	3.4%	0.9%	0.002
Cardiac or vascular event	41%	20%	<0.001
Any CMV > 500	14%	12%	0.338
SBP mean (mm Hg mean ± SD)	141 ± 13	142 ± 12	0.357
Categorical SBP Mean			0.172
SBP mean < 110	3%	1%	0.061
SBP mean = 110–159	92%	93%	0.455
SBP mean > = 160	6%	6%	0.946
Pulse pressure SD	14.8 ± 3.7	13.8 ± 3.8	<0.001
Pulse mean	81 ± 9	80 ± 9	0.095
Glucose mean (g/dL)	140 ± 33	134 ± 32	0.002
HGB mean (g/dL)	10.7 ± 1.6	11.3 ± 1.4	<0.001
HGB slope per 30 days (g/dL per 30 days)[†]	−0.21 ± 3.72	0.33 ± 0.60	0.021
Maximum eGFR (mL/min per 1.73 m^2)	64 ± 41	69 ± 23	0.043
eGFR SD (mL/min per 1.73 m^2)	11.0 ± 8.4	10.1 ± 5.3	0.073
eGFR slope per week: maximum value (mL/min per 1.73 m^2/week)	−5.5 ± 22.1	−1.4 ± 3.7	0.003
Tacrolimus mean (ng/mL)	8.6 ± 1.6	8.8 ± 1.3	0.043
Tacrolimus SD (ng/mL)	4.0 ± 1.4	3.7 ± 1.1	<0.001
INP readm count	1.6 ± 1.7	0.9 ± 1.3	<0.001
ED readm count	0.17 ± 0.74	0.15 ± 0.51	0.666
EHR ± NLP			
BK plus rejection			<0.001
Any BK > 500 and rejection	4%	3%	0.413
Any BK > 500 only	11%	14%	0.124
Rejection only	11%	3%	<0.001
No BK > 500 and norrejection	74%	79%	0.057

[†]HGB slope determined between 7 days post-transplant until the end of the exposure. [‡]Exposure period for 5 years model is 365 days. BK, BK virus; BMI, body mass index; CMV, cytomegalovirus; ED, emergency department; eGFR, estimated glomerular filtration rate; EHR, electronic health record; HGB, haemoglobin; IP, inpatient; KPDI, Kidney Donor Profile Index; LOS, length of stay; MI, myocardial infarction; MUSC, Medical University of South Carolina; NLP, natural language processing; SBP, systolic blood pressure; SD, standard deviation; UNOS, United Network of Organ Sharing.

comorbidities, complications, vital signs, change in lab values, health care utilization, and NLP for Banff scores.

In lieu of a 'one size fits all' approach to follow-up patient management, post-transplant care has the potential to become truly personalized for patients through estimating risk of graft loss with the use of Big Data to develop and autonomously deploy sensitive risk models. It may be most appropriate to closely follow patients at highest risk of graft loss and apply centre-level resources in a more intense manner to these vulnerable populations.[7] As an example of how to use this approach, let us compare two individual patients in our cohort with similar baseline variables that presented during the 1 year follow-up with distinctly different HGBs, eGFR slopes and tacrolimus variability with the end result of one patient suffering graft loss 3 years post-transplant. These two non-smoking female African–American patients received kidneys with KPDI of less than 20%, lived greater than 200 miles from the transplant centre, received

Table 2 Cox proportional regression hazard ratio and 95% confidence intervals for 5 years graft survival model

Variable	Hazard ratio	HR 95% CI		*P* value
Female	0.665	0.511	0.866	**0.003**
KDPI (ref = '1–20')				
21–84	1.541	1.119	2.122	**0.008**
85–100	2.285	1.522	3.428	**<0.001**
Distance to MUSC_km (ref = '>322')				
161–322	0.706	0.519	0.959	**0.026**
80–161	0.569	0.356	0.908	**0.018**
<80	0.470	0.324	0.680	**<0.001**
Employed	0.757	0.550	1.042	0.088
Receive disability	0.708	0.538	0.932	**0.014**
Smoker	1.954	1.323	2.884	**0.001**
Myocardial infarction	1.713	1.151	2.548	**0.008**
Cardiac arrhythmias	1.450	1.080	1.947	**0.014**
Drug abuse	1.737	0.929	3.250	0.084
Cardiac or vascular event	2.069	1.550	2.761	**<0.001**
BK > 500 and rejection (ref = 'No BK No REJ)				
BK > 500 and rejection	0.999	0.538	1.856	0.998
BK > 500 only	1.050	0.701	1.573	0.812
Rejection only	2.189	1.453	3.298	**<0.001**
SBP mean (ref = '110–159')				
<110	4.274	1.702	10.731	**0.002**
> = 160	0.714	0.425	1.197	0.201
Pulse pressure SD	1.065	1.029	1.103	**<0.001**
HGB mean	0.858	0.774	0.951	**0.004**
HGB slope per month	0.719	0.630	0.822	**<0.001**
Heart rate mean	1.023	1.008	1.038	**0.002**
Max eGFR	0.994	0.988	0.999	**0.017**
eGFR slope per week	0.978	0.972	0.983	**<0.001**
Tacrolimus mean	0.913	0.828	1.007	0.069
Tacrolimus SD	1.176	1.065	1.299	**0.001**
Inpatient readm count	1.101	1.011	1.199	**0.027**

BK, BK virus; eGFR, estimated glomerular filtration rate; HGB,haemoglobin; MUSC, Medical University of South Carolina; SBP, systolic blood pressure; SD, standard deviation.

disability income, had neither CV comorbidities at transplant nor demonstrated BK viral infection or acute rejection, and had similar mean SBP values (110–159 mmHg). Immediate post-transplant mean HGB levels and maximum eGFR values were also similar. However, in the patient suffering from graft loss at 3 years, her 30-day HGB slope demonstrated a change of −0.25 g/dL per month vs. a 1.66 g/dL per month increase for the non-graft loss patient; an eGFR slope of −1.92 mL/min/1.73 m^2 per week *versus* a flat 0.00 mL/min per1.73 m^2 and a mean tacrolimus value of 4.7 ng/mL *versus* 8.6 ng/mL, respectively. In this example, these two patients have identical baseline risk, yet their post-transplant clinical trajectories were vastly different. Theoretically, our model would provide information allowing clinicians to focus on the high-risk patient.

It has yet to be determined whether utilization of this model to identify at-risk individuals can lead to better population outcomes overall. However, there are a number of significant mutable baseline and follow-up clinical variables that can be the focus of deliberate intervention. Patients that resided farther distances from the transplant centre were at higher risk of graft loss. This may be a reflection of care coordination and logistical challenges in providing optimal follow-up at a substantial distance from the transplant center.[7] Efforts to improve remote monitoring and follow-up may mitigate this risk. Active tobacco users at the time of transplant were at twice the risk of graft loss. Interventions to expand smoking cessation endeavours may help to mitigate this risk factor.[13,14] The predominant pre- and post-transplant risk factors were in the domain of CV events. CV health and risk factor control is a crucial part of optimizing long-term outcomes in all patient populations, but this is particularly the case within kidney transplantation. To significantly improve long-term outcomes, better management of CV disease and risk factors –particularly hypertension –is needed.[15,16]

A number of additional, potentially modifiable, post-transplant risk factors were identified through this model. Tacrolimus variability during the 1 year post-transplant was a significant risk factor for graft loss. Tacrolimus variability has previously been associated with graft loss in a number of studies, including within our own centre; this may be a reflection of increased non-adherence.[17–19] Acute rejection has traditionally been a strong risk factor for graft loss,

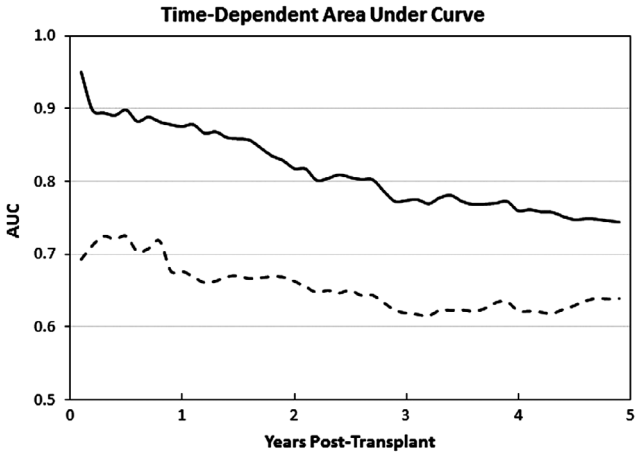

Fig. 2 Comparison of time-dependent area under curve (AUC) for 5 years graft survival model. Time-dependent receiver operating characteristic (ROC) and AUC used to determine predictive accuracy of the Big Data (UNOS + EHR + NLP) and the UNOS only models. UNOS iAUC = 0.660 (95% confidence intervals (CI) 0.601–0.706); Big Data iAUC = 0.819 (95% CI = 0.771–0.857). Time (t) expressed in years. EHR, electronic health record; iAUC, integrated AUC NLP, Natural Language Processing; UNOS, United Network of Organ Sharing. (——) UNOS + EHR + NLP, iAUC = 0.819; (-----) UNOS baseline only, iAUC = 0.660.

which is consistent within our model, although it must be noted these events are quite rare. Acute rejection is not necessarily preventable in all cases, but transplant centres may need to closely follow these patients for several years to prevent downstream negative clinical sequelae associated with these events.[20,21]

Decreasing HGB slope was one of the strongest predictors of graft loss. Anaemia is present in as many as 72% of patients 1 month post-transplant, decreasing to 40% of the patients by 3 months and 20% at 1 year.[22] The early stages of anaemia have been attributed to several factors, including: a consequence of end-stage renal disease pre-transplant nadir, surgical blood loss, inadequate nutrition, fluid overload, ischaemic reperfusion injury, poor graft quality, delayed graft function and known adverse effects from immunosuppressive.[22–24] Late (>1 year post-transplant) anaemia is thought to be attributable to impaired graft function, iron deficiency, female gender and adverse effects of pharmaceuticals.[22] In two retrospective studies, HGB levels at 12-month post-transplant were correlated with long-term graft outcome. Anaemia at 12 months, when stratified according to kidney function (i.e. eGFR), predicted graft loss in live donor recipients.[25] Additionally, higher levels of HGB at 1 year were predictive of improved long-term graft survival.[26] These studies also suggest that utilization of erythropoietin stimulating agents (ESA to counter late stage anaemia may benefit graft survival. Such was the case in the CAPRIT trial, in which epoetin-β was used to correct post-transplant.[27] In this study, use of ESA for 2 years normalized HGB levels and produced a 95% graft survival *versus* 80% for the partially normalized group. However, there have been inconsistencies in the utilization and reported outcomes of ESA in long-term post-transplant anaemia,[22,28] possibly as a consequence of the concerns regarding adverse CV events associated with the chronic use of these agents.[29] Thus, further study into this potentially modifiable risk factor is clearly warranted.

There have been previous attempts at developing accurate predictive models for graft survival, designed to identify the combination of factors that result in poorer outcomes. Methodologies include data mining, computational forecasting, clinical markers, 'Big Data', and others.[30–32] As a consequence of risk factor heterogeneity affecting graft survival, predictive models have met with mixed success. Krikov et al. developed a tree-based learning model using the United States Renal Data System data from 1990 to1999 to predict kidney graft loss using donor, recipient and transplant variables.[30] More recently, a predictive model of 5 years graft survival was based solely upon pre-transplant data using 10 predictors.[30] Their results (www.transplantscore.com) gave a predictive value C-statistic = 0.70 for patient mortality, 0.63 for graft failure and 0.63 for combined mortality/graft failure. A Korean group developed a combination model using machine learning with survival statistics, immunological factors and specific donor and recipient variables to develop a long-term graft survival model.[32] The ROC AUC values for their model ranged from 0.975 for 1 year, to 0.701 at 5 years, to 0.707 for 10 years post-transplant.[32] There are other reports using analytical forecasting models in conjunction with clinical outcomes to predict early allograft rejection,[30] delayed graft function,[31] 3 years deceased donor graft survival,[33] CV-related deaths post-transplant,[34,35] and health care utilization post-transplant.[36] As there are numerous examples of risk-prediction models in the transplant literature, future studies are now urgently required to determine if such models can be used as clinical-decision aids and lead to improved outcomes.[20–26] Previously, we implemented such a model for 1- and 3 years graft survival.[8] In that study, clinically relevant, daily patient and administrative data were used to improve predictability of early patient and graft survival. It is our intent in future studies to use a similar automated approach to demonstrate value of the current model in predicting 5 years graft survival.

There are several limitations to our current study. The analysis that we developed used a retrospective cohort of patients from a single academic transplant centre located in the Southeast. As such, our study population may not represent the larger cohort of transplant patients across the United States. Distinct differences include a high proportion of African–Americans (68% compared with 38% in the United States), predominantly rural population, and relatively low socioeconomic status. Also, there were few patients with glomuleronephritis aetiologies of end-stage renal disease. Furthermore, it is unclear whether or not the Big Data variables included in this algorithm are partially or

fully generalizable across kidney transplant recipients in the United States or are mutable through intervention. Although our cohort was small ($n = 1445$) as compared with the US registry databases, it did provide for development of a strong predictive 5 years graft loss model. Lastly, as a matter of policy, patients were not transplanted with pre-existing donor-specific antibodies. Therefore, data on donor-specific antibodies was not included in the modelling. This model would benefit from validation by an analysis of a larger patient cohort from additional academic transplant centres to ensure widespread applicability.

Using Big Data predictive analytics, a 5 years graft loss model was developed from data autonomously captured during the first post-transplant year. The model resulted in strong predictability and discernment and substantially improves upon models using only baseline patient and donor information. Through automation and utilization of this model, there is the strong potential for post-transplant care to be individualized for patients, with the goal of mitigating important risk factors in these patients to improve long-term graft survival.

ACKNOWLEDGEMENTS

Support for these studies was provided by National Institutes of Health; K23DK091514 (DD) and R03DK106432 (DD). Authors thank to Ms Rachel Mehard for editorial assistance.

CONFLICTS OF INTEREST

We have no conflict of interest to report.

REFERENCES

1. Lodhi SA, Meier-Kriesche HU. Kidney allograft survival: The long and short of it. *Nephrol. Dial. Transplant.* 2011; **26**: 15–7.
2. Wekerle T, Segev D, Lechler R, Oberbauer R. Strategies for long-term preservation of kidney graft function. *Lancet* 2017; **389**: 2152–62.
3. Galichon P, Xu-Dubois YC, Finianos S, Hertig A, Rondeau E. Clinical and histological predictors of long-term kidney graft survival. *Nephrol. Dial. Transplant.* 2013; **28**: 1362–70.
4. Wang JH, Skeans MA, Israni AK. Current status of kidney transplant outcomes: Dying to survive. *Adv. Chronic Kidney Dis.* 2016; **23**: 281–6.
5. El-Zoghby ZM, Stegall MD, Lager DJ *et al*. Identifying specific causes of kidney allograft loss. *Am. J. Transplant.* 2009; **9**: 527–35.
6. Gupta G, Unruh ML, Nolin TD, Hasley PB. Primary care of the renal transplant patient. *J. Gen. Intern. Med.* 2011; **25**: 731–40.
7. Schachtner T, Reinke P. Provision of highly specialized aftercare by the transplant center strongly improves patient and allograft survival in long-term follow-up after kidney transplantation. *Transplantation* 2017; **101**: S27.
8. Srinivas TR, Taber DJ, Su Z *et al*. Big Data, predictive analytics, and quality improvement in kidney transplantation: A proof of concept. *Am. J. Transplant.* 2017; **17**: 671–81.
9. Quan H, Sundararajan V, Halfon P *et al*. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care* 2005; **43**: 1130–9.
10. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* 1996; **15**: 361–87.
11. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* 2011; **30**: 1105–17.
12. Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules fort-year survivors with censored regression models. *J. Am. Stat. Assoc.* 2007; **102**: 527–37.
13. Yaribakht S, Malartic C, Grange G, Morel O. Operative risk related to tobacco in gynecology. *Gynecol. Obstet. Fertil.* 2014; **42**: 343–7.
14. Khullar D, Maa J. The impact of smoking on surgical outcomes. *J. Am. Coll. Surg.* 2012; **215**: 418–26.
15. Taber DJ, Hunt KJ, Fominaya CE *et al*. Impact of cardiovascular risk factors on graft outcome disparities in black kidney transplant recipients. *Hypertension* 2016; **68**: 715–25.
16. Bostom AD, Brown RS, Chavers BM *et al*. Prevention of post-transplant cardiovascular disease—Report and recommendations of an ad hoc group. *Am. J. Transplant.* 2002; **2**: 491–500.
17. Borra LCP, Roodnat JI, Kal JA, Mathot RAA, Weimar W, van Gelder T. High within-patient variability in the clearance of tacrolimus is a risk factor for poor long-term outcome after kidney transplantation. *Nephrol. Dial. Transplant.* 2010; **25**: 2757–63.
18. Sapir-Pichhadze R, Wang Y, Famure O, Li Y, Kim SJ. Time-dependent variability in tacrolimus trough blood levels is a risk factor for late kidney transplant failure. *Kidney Int.* 2014; **85**: 1404–11.
19. van Gelder T. Within-patient variability in immunosuppressive drug exposure as a predictor for poor outcome after transplantation. *Kidney Int.* 2014; **85**: 1267–8.
20. Ferguson R. Acute rejection episodes – best predictor of long-term primary cadaveric renal transplant survival. *Clin Transplant.* 1994; **8**: 328–31.
21. Meier-Kriesche HU, Schold JD, Srinivas TR, Kaplan B. 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**: 378–83.
22. Bamgbola OF. Spectrum of anemia after kidney transplantation: Pathophysiology and therapeutic implications. *Clin. Transplant.* 2016; **30**: 1185–94.
23. Parajuli S, Clark DF, Djamali A. Is kidney transplantation a better state of CKD? Impact on diagnosis and management. *Adv. Chronic Kidney Dis.* 2016; **23**: 287–94.
24. Joist H, Brennan DC, Coyne DW. Anemia in the kidney-transplant patient. *Adv. Chronic Kidney Dis.* 2006; **13**: 4–10.
25. Huang Z, Song T, Fu L *et al*. Post-renal transplantation anemia at 12 months: Prevalence, risk factors, and impact on clinical outcomes. *Int. Urol. Nephrol.* 2015; **47**: 1577–85.
26. Lofaro D, Greco R, Papalia T, Bonofiglio R. Increasing levels of hemoglobin improve renal transplantation outcomes. *Transplant. Proc.* 2011; **43**: 1036–8.
27. Choukroun G, Kamar N, Dussol B *et al*. Correction of postkidney transplant anemia reduces progression of allograft nephropathy. *J. Am. Soc. Nephrol.* 2012; **23**: 360–8.
28. Molnar MZ, Mucsi I, Macdougall IC *et al*. Prevalence and management of anaemia in renal transplant recipients: Data from ten European centres. *Nephron Clin. Pract.* 2011; **117**: c127–34.
29. Molnar MZ, Tabak AG, Alam A *et al*. Serum erythropoietin level and mortality in kidney transplant recipients. *Clin. J. Am. Soc. Nephrol.* 2011; **6**: 2879–86.
30. Furness PN, Kazi J, Levesley J, Taub N, Nicholson M. A neural network approach to the diagnosis of early acute allograft rejection. *Transplant. Proc.* 1999; **31**: 3151.
31. Decruyenaere A, Decruyenaere P, Peeters P, Vermassen F, Dhaene T, Couckuyt I. Prediction of delayed graft function after kidney

transplantation: Comparison between logistic regression and machine learning methods. *BMC Med. Inform. Decis. Mak.* 2015; **15**: 83.
32. Yoo KD, Noh J, Lee H *et al*. A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: A multicenter cohort study. *Sci. Rep.* 2017; **7**: 8904.
33. Goldfarb-Rumyantzev AS, Scandling JD, Pappas L, Smout RJ, Horn S. Prediction of 3-yr cadaveric graft survival based on pre-transplant variables in a large national dataset. *Clin. Transplant.* 2003; **17**: 485–97.
34. Mansell H, Stewart SA, Shoker A. Validity of cardiovascular risk prediction models in kidney transplant recipients. *ScientificWorldJournal* 2014; **2014**: 750579.
35. Pita-Fernández S, Pértega-Díaz S, Valdés-Cañedo F *et al*. Incidence of cardiovascular events after kidney transplantation and cardiovascular risk scores: Study protocol. *BMC Cardiovasc. Disord.* 2011; **11**: 2.
36. Keong FM, Afshar YA, Pastan SO, Chowdhury R, Binongo JN, Patzer RE. Decreasing estimated glomerular filtration rate is associated with increased risk of hospitalization after kidney transplantation. *Kidney Int. Rep.* 2016; **1**: 269–78.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website:

Table S1 Rate and severity of acute rejection rates.