ASTS
Leadership Development Program
October 18–21, 2020

The Premier Executive Management Course Designed Exclusively for the Field of Transplantation

In a field of increasing regulation and complexity, you need an edge to succeed.

The ASTS Leadership Development Program will introduce you to expert coaches to improve your transplant center's performance and accelerate your career success.

Learn. Lead. Succeed.

Learn more at ASTS.org/LDP.

ASTS
American Society of Transplant Surgeons®

Kellogg School of Management
Northwestern

Northwestern University Kellogg School of Management James L. Allen Center, Evanston, IL

Received: 28 September 2018 | Revised: 31 December 2018 | Accepted: 6 February 2019

DOI: 10.1111/ajt.15312

AJT

ORIGINAL ARTICLE

Modeling graft loss in patients with donor-specific antibody at baseline using the Birmingham-Mayo (BirMay) predictor: Implications for clinical trials

Andrew Bentall[1,2] | Byron H. Smith[2] | Manuel Moreno Gonzales[2,3] | Keisha Bonner[2] | Walter D. Park[2] | Lynn D. Cornell[4] | Patrick G. Dean[2] | Carrie A. Schinstock[2] | Richard Borrows[1] | Carmen Lefaucheur[5] | Alexandre Loupy[6,7] | Mark D. Stegall[2]

[1]Department of Renal Medicine, Queen Elizabeth Hospital Birmingham, Birmingham, UK

[2]William J. von Liebig Transplant Center, Mayo Clinic, Rochester, Minnesota

[3]Servicio de Cirugía General, Americana, Lima, Perú

[4]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN

[5]Department of Nephrology and Kidney Transplantation, Saint-Louis Hospital, Assistance Publique - Hôpitaux de Paris, Paris, France

[6]Department of Nephrology-Transplantation Necker Hospital, Assistance Publique-Hôpitaux de Paris, University Paris Descartes, Paris, France

[7]Paris Cardiovascular Research Centre – Biostatistics Unit University Paris Descartes, UMR-S970, Paris, France

Correspondence
Andrew Bentall
Email: Bentall.andrew@mayo.edu

Predicting which renal allografts will fail and the likely cause of failure is important in clinical trial design to either enrich patient populations to be or as surrogate efficacy endpoints for trials aimed at improving long-term graft survival. This study tests our previous Birmingham-Mayo model (termed the BirMay Predictor) developed in a low-risk kidney transplant population in order to predict the outcome of patients with donor specific alloantibody (DSA) at the time of transplantation and identify new factors to improve graft loss prediction in DSA+ patients. We wanted define ways to enrich the population for future therapeutic intervention trials. The discovery set included 147 patients from Mayo Cohort and the validation set included 111 patients from the Paris Cohort—all of whom had DSA at the time of transplantation. The BirMay predictor performed well predicting 5-year outcome well in DSA+ patients (Mayo C statistic = 0.784 and Paris C statistic = 0.860). Developing a new model did not improve on this performance. A high negative predictive value of greater than 90% in both cohorts excluded allografts not destined to fail within 5 years. We conclude that graft-survival models including histology predict graft loss well, both in DSA+ cohorts as well as DSA- patients.

KEYWORDS

alloantibody, clinical research/practice, kidney (allograft) function/dysfunction, kidney transplantation/nephrology, pathology/histopathology, protocol biopsy, risk assessment/risk stratification

1 | INTRODUCTION

Predicting which renal allografts will fail and the likely cause of failure is important in clinical trial design to either enrich patient populations to be treated (eg, studies design to treat antibody-mediated rejection [ABMR]) or as surrogate efficacy endpoints for trials aimed at improving long-term graft survival. Several studies have demonstrated that kidney transplant recipients who have DSA at the time of transplantation have inferior outcomes to those without DSA.[1-3] Clearly, new therapy is needed to overcome the immunologic hurdle of preformed

Abbreviations: ABMR, antibody-mediated rejection; ACR, albumin creatinine ratio; cABMR, chronic antibody-mediated rejection; CDC, complement dependent cytotoxic crossmatch; DCGF, death-censored graft failure; DSA, donor specific alloantibody; eACR, estimated albumin:creatinine ratio; eGFR, estimated glomerular filtration rate; FXM, T and B flow cytometric crossmatch; MFI, median fluorescence intensity; NDS, Anon-donor specific alloantibody; NRI, net classification index; ROC, receiver operating characteristic; SAB, Single Antigen Bead.

DSA.[4] However, determining the appropriate design of clinical trials in this area has been vexing. Similar to the issue in low-risk populations is that only a relatively small subset of patients experience graft loss and the time to graft loss may be many years. Despite many reports having suggested that factors such as high levels of DSA or the presence of microvascular inflammation on 1-year protocol biopsies are associated with graft loss, quantification of these risk factors has been difficult and particularly treatment effect limited.

We have previously used mathematical Modeling to predict the likelihood of graft loss in erstwhile low-risk/DSA- patients, with both clinical parameters (Birmingham Model) and the inclusion of histology termed the BirMay Model[5,6] (Table 1). The former model initially used primarily clinical parameters (estimated glomerular filtration rate [eGFR], urinary albumin:creatinine ratio [ACR], acute rejection in first year, age, race, serum albumin) present at 1 year to predict the outcome of renal allografts at 5 years. The Birmingham Model performed much better compared to using eGFR or ACR alone in predicting graft loss. In order to assess the effect of histology on the model, we identified the addition of Banff ci score and g score from 1-year protocol biopsy data as improving the ability to predict death-censored graft failure (DCGF). Recently a prognostic model was published describing outcomes following treatment for ABMR developed in Paris, but not for an at-large population with preformed DSA at the time of transplant.[7]

The goal of the current study was twofold. First, we aimed to test the robustness of the existing BirMay Predictor in predicting all-cause graft loss in a cohort of higher-risk patients with DSA at the time of transplantation. Secondly, we then sought to develop a new model for DSA+ patients, using a discovery set at Mayo Clinic and a validation set from the Paris group in order to examine predictors that might be specific for graft loss due to chronic antibody mediated rejection (cABMR) and thus either be more useful as a surrogate endpoint for disease-specific graft loss or as inclusion criteria to enrich populations for graft loss due to cABMR.

2 | METHODS

2.1 | Patient population

In this institutional review board–approved study, 147 patients at Mayo Clinic, Rochester were included in the discovery cohort received a transplant between January 2000 to December 2010 who had positive crossmatch with either a complement dependent cytotoxic (CDC) or a flow cytometric (FXM) positive crossmatch at initiation of treatment. There was heterogeneity in the types of treatments during this period prior to transplantation—the results of which have previously been published.[2,8] For the validation set, 111 patients from the Paris Transplant Group (both Necker Hospital and Saint-Louis Hospital, Paris) transplanted between March 2003 and May 2011 who were in the high immunologic risk program as defined by a high peak or day zero DSA levels (mean fluorescence intensity [MFI] > 3000) with a CDC-negative crossmatch. The majority of this cohort predated the prognostic scores for treatment of ABMR (after 2008 inclusion). Both of these cohorts were included due to available histology at 1-year post transplantation. Comparison between the groups was made on the DSA with the highest MFI in each cohort.

2.2 | Immunosuppression

Patients at Mayo Clinic, Rochester were treated according to an evolving protocol which was aimed at reduction in DSA levels prior to transplantation, using anti-thymocyte globulin (ATG) at induction and triple immunosuppression with calcineurin inhibition, mycophenolic acid, and prednisone as ongoing therapies. In addition to ATG at induction and then a calcineurin inhibitor, mycophenolate mofetil and prednisolone, as previously described,[9] the Paris population received a posttransplant desensitization protocol starting at day 0 with high-dose Intravenous Immunoglobulin plasma exchanges and later rituximab was used as the program evolved.

2.3 | Variables and study endpoints

The fixed endpoints of DCGF and overall graft failure (including death or graft failure) were used and follow-up was a minimum of 5 years for surviving allografts, thus actual follow-up data was used for analysis. Histological data at 1-year protocol biopsies and anti-HLA antibodies at the same time point were collected. Biopsies were interpreted according to clinical guidelines in each center and done according to protocol at 1-year posttransplant.

Table 1 shows both existing published models, with all the variables and the weight given to each variable, as well as the interaction of variables within these models. Where albumin:creatinine ratio was not available, it was estimated using equations previously published.[10,11] In brief, the urinary albumin was estimated as a proportion of the total 24 hour urinary protein as described by Halimi and colleagues with albumin constituting 24% of proteinuria if less than 250mg, 35% if less than 750mg, 43% if less than 1000mg and 56% if greater than 1000mg. The estimated urinary creatinine (mg) using the formula of 879.89 + 12.51 * (weight [kg]) - 6.19*(age) +(34.51 if black) - (379 if female). The estimated ACR (eACR) was the calculated by the estimated albumin divided by the estimated creatinine in the urine.

In the Mayo Clinic cohort, clinical noting and subsequent biopsies were reviewed to define the causes of allograft failure, specifically chronic antibody-mediated rejection (cABMR) where preexisting DSA is a significant risk factor. Whilst determining the exact cause of graft loss is always problematic, grafts without chronic glomerulopathy in their biopsies prior to graft loss were designated as graft loss due to other causes; these included recurrent disease, oxalate deposition, and systemic infections. Other causes of allograft failure, including recurrent disease may include patients with existing chronic glomerulopathy, but had reported other causes of allograft failure per clinical record. Recurrent disease was clearly defined by pathologists and often shown on multiple biopsies after the 1-year biopsy, per our protocol biopsy practice. The Paris cohort had two nephrologists review clinical record for designation of cause of graft failure.

TABLE 1 Weighted variables in previous models and new model

	Birmingham model			Birmingham-Mayo model (BirMay)			DSA+ new model		
	Transformation	Death-censored failure	Overall transplant failure	Transformation	Death-censored failure	Overall transplant failure	Transformation	Death-censored failure	Overall transplant failure
Urine ACR (in mg/mmol)	Log scale	—	0.666	Log_{10} value	—	0.3712			
Urine ACR (in mg/mmol)	Log scale—0.46	1.107	—	Log_{10} value—0.46	0.5053	—	Log_{10} value—0.72	1.398	1.239
Serum albumin (in g/L)	(Value—40)/5	—	−0.217	(Value—40)/5	—	0.4000			
eGFR (in mL/min/1.73 m^2	(Value—47)/10	−0.297	−0.206	(Value—47)/10	−0.3969	−0.2367	(Value—51)/10	−0.345	−0.301
eGFR (in mL/min/1.73 $m^2)^2$	$([Value—47]/10)^2$	0.0711	0.0669	$([Value—47]/10)^2$	0.0677	0.06613			
Any Rejection Episode		1.038	0.550	—	0.2806	0.4988			
Black ethnicity		1.324	1.095	—	0.7582	—			
Asian ethnicity	—	1.039	0.653						
Recipient age (in y)	(Value—46)/10	0.138	0.00149	(Value—46)/10	−0.2366	0.1202			
Recipient age (in y) squared	$([Value—46]/10)^2$	0.204	0.187	$([Value—46]/10)^2$	—	0.0858			
Urine ACR with rejection interaction	$Log(ACR)^2$ 0.46 if rejection = yes	−0.543	—	Log_{10}(ACR value)—0.46 if rejection = yes	0.5579	—			
g					0.917	0.5985			
ci					0.5074	—			

ACR, albumin:creatinine ratio; ci, chronic interstitial fibrosis Banff score; DSA, donor specific antibody; eGFR, estimated glomerular filtration rate; g, glomerulitis in Banff score.

2.4 | Statistical analysis

All analyses were carried out using R v3.2 (Vienna, Austria). Risk scores of 5-year death-censored and all-cause kidney failure based on the Birmingham and Birmingham-Mayo models for high-risk patients were calculated for comparison.[5,6] New models were constructed using Cox proportional hazards regression. All clinically-relevant predictors were first tested at the univariate level (Table S1). These included donor and recipient demographic information, baseline serologic factors, 1-year biopsy Banff scores and rejection status, and 1-year HLA antibody variables. Any variable with a $P < .1$ was carried forward for a multivariate model. Variable selection for multivariate models was performed via forward stepwise, using a criterion of $P < .05$ to remain in the model.

Model performance was evaluated using concordance (Harrell's C-statistic) on the full dataset as well as through a 10-fold cross-validation. In each fold of the cross-validation, a multivariate Cox model is fit, with forward stepwise variable selection, using all possible predictors on 90% of the data. The concordance is then estimated by predicting the risk of the remaining 10% of the data. This is repeated 10 times to loop through each of the possible 10% hold-out samples.

Model calibration was evaluated using the Hosmer-Lemeshow statistic and test using risk categorizations of 0% to 10%, 10% to 30%, and 30% to 100%. External cross-validation was carried out on a cohort from the Paris group. New models were compared versus old models based on net reclassification index.

Risk scores in Cox models are defined by the covariate values as well as the baseline hazard at specific times. In order to recalibrate models in the Paris cohort, Cox models were fit using the initial models' linear predictor and zero iterations of the optimization algorithm (setting the "iter" option to 0 in the coxph() function in R). This allows for an updated estimate of the baseline hazard function without modification to the linear predictor. Note that concordance does not change when a model is re-calibrated, but the Hosmer-Lemeshow statistic will.

3 | RESULTS

3.1 | Demographics and clinical outcomes of the patient cohorts

The demographics and clinical outcomes of the Mayo and Paris Cohorts are compared in Table 2. The Mayo DSA+ Cohort had a higher preemptive transplantation rate (40, 27.2% vs. 1, 0.9%, $P < .01$), higher rates live donor transplants (143, 97.3% vs. 8, 7.2%, $P < .01$), higher re-transplantation rates (55, 37.4% vs. 57, 51.4%, $P < .01$) and more females (102, 69.4% vs. 60, 54.1%, $P = .02$) in comparison to the Paris cohort (Table 2). At 1 year after transplant, the Mayo DSA+ cohort had median follow-up time from transplantation of 82.4 months (59.1-108.0) while the Paris Cohort follow up was 49.8 months (34.2-64.6). Death between 1 and 5 years was 2.0% (3/147) in the Mayo Cohort compared to 4.5% (5/111, $P = .448$) in the Paris cohort. In the same period, DCGF was 20.4% (30/147) in the Mayo Cohort compared to 17.1% (19/111) in the Paris Cohort. The comparative causes of 5-year graft loss in both cohorts are shown in Figure 1, but in particular 66.7% (Mayo) and 73.6% (Paris) of graft losses were attributed to chronic ABMR.

There were also significant histological differences at 1 year between the validation set and the test cohort. In particular there was greater incidence of glomerulitis in the Paris (64.0% vs. 41.5%, $P < .001$), more arteriolar hyalinosis (53.2% vs. 26.4%, $P < .001$) and vascular fibrous intimal thickening (68.2% vs. 49.0%, $P = .002$) in the Paris cohort. There were similar rates of both interstitial inflammation (31.5% vs. 25.2%, $P = .226$), tubulitis (31.5% vs. 28.6%, $P = .681$), peritubular capillaritis (62.2% vs. 68.0%, $P = .390$) and interstitial fibrosis and tubular atrophy (64.9% vs. 66.3%, $P = .792$) between groups and a higher incidence of chronic glomerulopathy (27.2% vs. 12.6%, $P = .005$) in the

TABLE 2 Demographic table of new model cohorts

	Rochester (N = 147)	Paris (N = 111)	*P* value
Age			.746
Mean (SD)	47 (13.2)	47.5 (12.2)	
Gender			.017
Female	102 (69.4%)	60 (54.1%)	
Dialysis time (mo)			.257
Median (IQR)	4.8 (0, 11.2)	4.76 (2.92, 7.97)	
Preemptive			<.001
	40 (27.2%)	1 (0.901%)	
HLA mismatch (ABDR)			.208
Mean (SD)	3.63 (1.32)	3.41 (1.46)	
Living donor			<.001
	143 (97.3%)	8 (7.21%)	
Retransplant			.035
	55 (37.4%)	57 (51.4%)	
Immunodominant DSA			<.001
Median (IQR)	2433 (909, 5265)	6562 (3050, 11643)	
Year 1 eGFR			.611
Mean (SD)	49.3 (17.2)	48.1 (20.7)	
Year 1 eACR			.805
Median (IQR)	2.93 (1.79, 13.3)	6.3 (2.78, 18.4)	
Immunodominant DSA			<.001
Class I	93 (68.4%)	34 (32.4%)	
Class II	43 (31.6%)	71 (67.6%)	
Five year survival			.448
Survival	114 (77.6%)	87 (78.4%)	
Allograft failure	30 (20.4%)	19 (17.1%)	
Death	3 (2.04%)	5 (4.5%)	

eACR, estimated albumin:creatinine ratio; eGFR, estimated glomerular filtration rate; g, glomerulitis in Banff score; ci, chronic interstitial fibrosis Banff score; DSA, donor specific antibody.

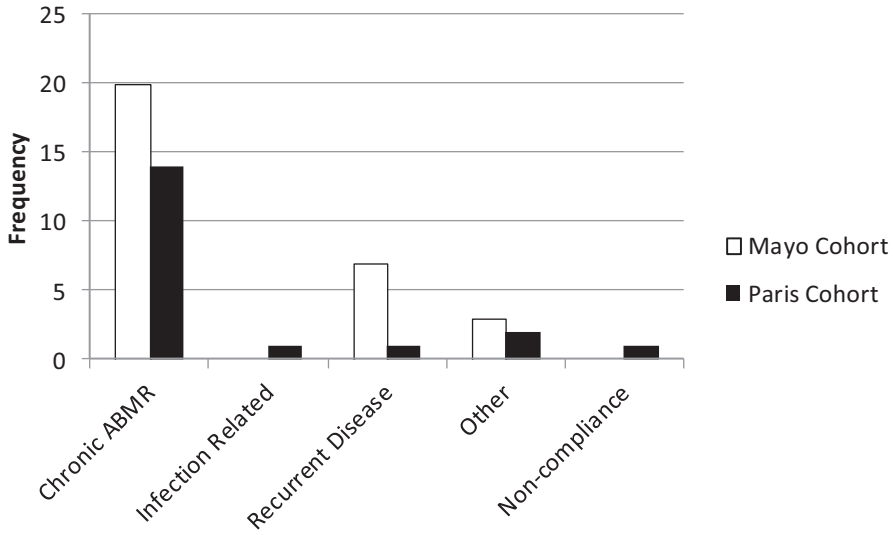


FIGURE 1 Cause of graft failure. The majority of allografts are lost due to chronic antibody mediated rejection (ABMR), where recurrent disease is the next highest risk factor for allograft loss within the first 5 years in the Mayo Cohort

Mayo cohort than the Paris cohort in 1-year biopsies (Table S2).There was, on a population level, a higher MFI in the Mayo cohort at baseline compared to Paris (Figure S1A), but conversely, the change in the MFI in the immunodominant DSA remained at a higher level 1-year posttransplant in the Paris cohort, than in the Mayo cohort (Figure S1B and S2C).

3.2 | Performance of "low-risk" predictor models in a high-risk DSA+ Mayo cohort

The Birmingham Risk model, based on serum and clinical variables, performed well for both overall graft failure and DCGF with the C-statistics for prediction of failure were 0.754 and 0.758, respectively. The BirMay histology-based risk model demonstrated similar predictive value for both graft loss and death. For overall graft failure, the C-statistic was 0.751 and for DCGF the C-statistic was 0.784. The latter being a slight improvement on the Birmingham model. This model including histology, as a variable was better calibrated to the DSA+ cohort with a Hosmer-Lemeshow *P*-value of 0.105 for overall graft failure and 0.001 for DCGF (compared to < 0.001 in the Birmingham Model). Both prediction models are shown in Table 1. The ability to predict graft loss and death, based on eGFR or eACR alone was compared to the current models under investigation and results as seen in Figure 2. Both models improved prediction, as compared to eGFR alone in the Mayo Clinic DSA+ Cohort for both DCGF and overall graft failure, with a net classification index (NRI) of around 30% for preexisting models, and an improvement of 30% to 50% in the new model (Table 3).

3.3 | Validation in the Paris cohort DSA+ cohort of existing models

Applying the existing Birmingham model, without histology to the Paris cohort demonstrated good prediction of both DCGF (C-stat 0.843), and this was slightly improved by adding the BirMay model with the histological parameters to give a C-statistic of 0.860. The Hosmer-Lemeshow statistic was 0.042 and 0.003, respectively, for both of these models in the Paris cohort (Table 4).

The application of the BirMay model risks scores were then grouped in each cohort and 5-year survival is then shown to be significantly different in those with ≥10% risk, which comprises of 54/147 patients in Mayo cohort and 46/111 patients in Paris cohort (Figure 3A -Mayo cohort and Figure 3B Paris cohort).

3.4 | Creating a new model for DSA+ patients—Independent predictors of transplant failure risk

All variables with clinical significance were included and the univariable are shown in Table S1. The factors associated at a univariate level were Banff scores of glomerulitis, tubulitis, peritubular capillaritis, chronic glomerulopathy, interstitial fibrosis and tubular atrophy, renal function, and albuminuria. In addition, recurrent glomerulonephritis risk, Hepatitis B and rejection in 1 year were other risk factors. Factors associated at a level of $P < .1$ were included in the multivariate model shown in Table S1. For Overall Graft Survival, eACR and eGFR were strongest predictive factors and these same factors were predictive in the death-censored graft survival model, albeit with different hazard ratios of effect – and thus only these two factors were included in the model for DSA+ patients (Figure 2C and D).

Using the new DSA+ model in the Paris cohort, this again performed very well and compared equivalently to the BirMay Predictor model with DCGF (C-stat 0.859). Overall graft failure (C-stat 0.788) also performed well, but this is due largely to the strong influence of both eGFR, and particularly, ACR in all of the prediction models. The benefit of risk reclassification was seen mainly in comparison to eACR in the Paris cohort for both existing models DCGF (NRI 30-46%) and the new model added reclassification benefit with the benefit of 33% over eACR alone.

3.5 | A "disease-specific" cause model for graft loss due to cABMR

Not all patients with DSA at the time of transplantation lose their graft to cABMR, even though this is an enriched population due to inclusion criteria. From 1 to 5 years, graft loss to other reasons was

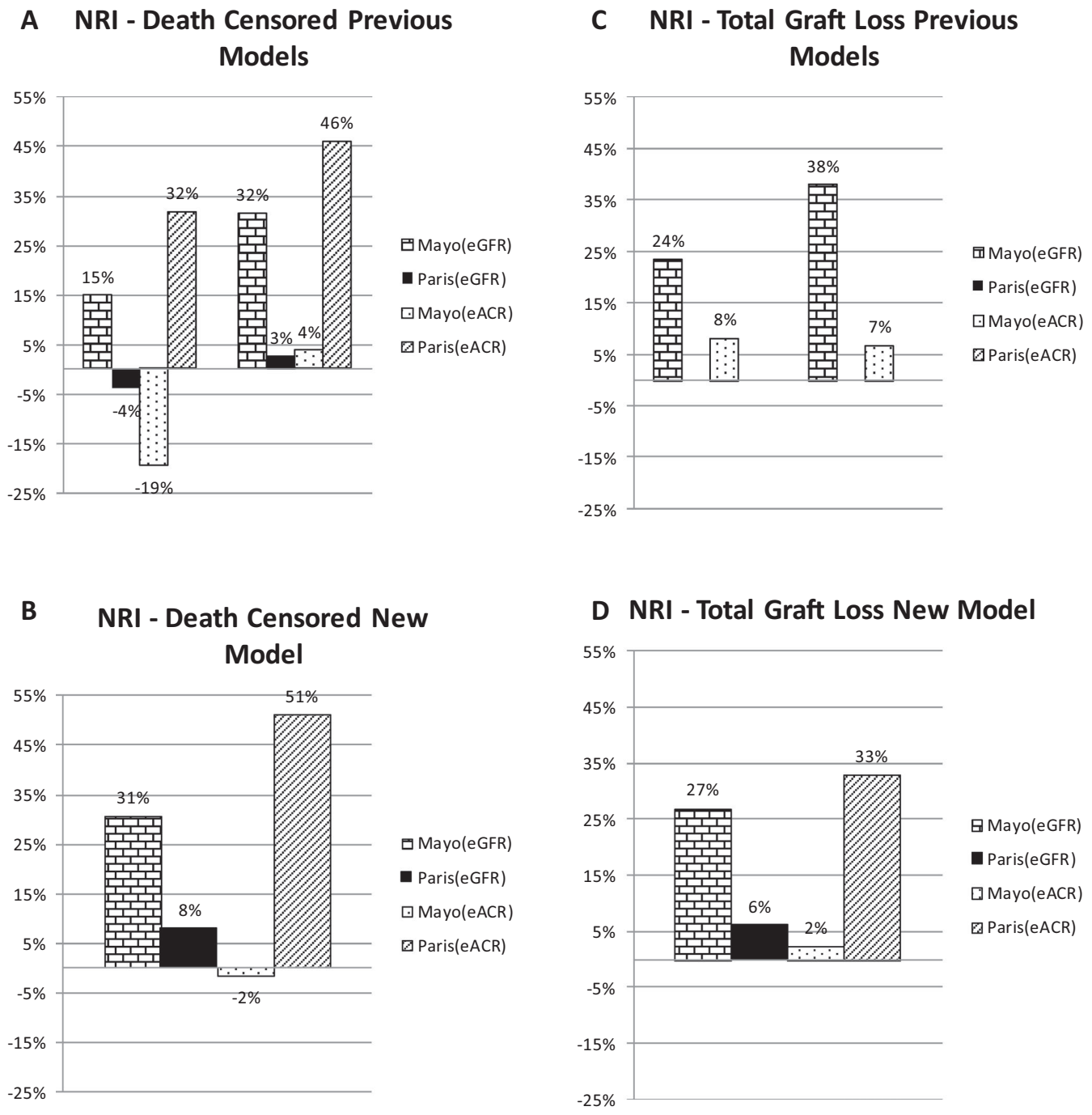


FIGURE 2 Net reclassification index (NRI) comparison between existing Birmingham Model and new Birmingham-Mayo model for both death censored and total graft loss, compared to both estimated glomerular filtration rate (eGFR) and estimated albumin:creatinine ratio (eACR) in both cohorts in A and B. A similar comparison is made for the new model in C and D

6.8% (10/147) in the Mayo cohort and 4.5% (5/111) in the Paris cohort. Graft losses from cABMR was 20/147 in the Mayo Cohort and 14/111 in the Paris cohort but these did represent 69.4% (34/49) of all total graft losses combining the two cohorts.

Using the BirMay model with ≥15% risk stratification in each cohort, we correctly identified 19/30 in Mayo failures 12/19 in Paris cohort, allowing significant improvement in risk identification of patients at higher risk of graft loss between 1 and 5 years. Thus, this model had a negative predicted value was 90% in Mayo Cohort and 92% in the Paris cohort using a cut off value of risk of 15%. Sensitivity of cABMR was 63% in each cohort and specificity of 84.6% and 88.0%, respectively, for Mayo and Paris (Figure 4).

3.6 | Potential clinical trial impact of Modeling

The clinical impact of applying the BirMay model to the Mayo Cohort in order to risk stratify for an intervention study would exclude 89 patients who would not have had allograft failure by 5 years after

1-year assessment (Figure 4). The use of the BirMay model would reduce the at-risk population to 37 out of 147 total. Based on any DCGF rate 20.4% (30/147), this proportion would be increased to 51.4% (19/37). The enrichment of the population means that for the 50% reduction in graft failure, a power calculation would need a study population of 133 are initially if using the prevalent population, but only 36 study participants if the BirMay model is applied—approximately a 75% population reduction in recruitment. If specific focus on cABMR as a cause of failure is the goal of study population, using the original incident patient rate of 13.6% (20/147) in the whole cohort compared to 37.8% (14/37) if applying the risk model; then recruitment of 214 patients would have been initially required to participate in a study, compared to only 59 recruited patients if only using BirMay risk selected population—again approximately a 75% population reduction in recruitment—for specific cABMR treatment.

Using similar calculations for the Paris cohort, the impact on trial power calculations are equivalent. The DCGF cause of failure rate of 17.1% (19/111) in the overall population is compared to 52.2% (12/23) graft failure in the enrichment of the population (post-BirMay model). Thus for an effective 50% reduction due to treatment, 167 participants would have been initially required, but only 35 study participants if the BirMay model is applied—approximately an 80% population reduction in recruitment. Again, specifically focusing on cABMR as a cause of failure, the rate of failure was 12.6% (14/111) in the whole cohort compared to 43.5% (10/23) if applying the BirMay risk model. Meaning that for an efficacy of 50% reduction in failure, 233 participants for total population would have been required to be recruited compared to 48 patients if only using BirMay risk selected population—again approximately an 80% population reduction in recruitment—and importantly would be exposing more patients to drug that would have no significant benefit to them and may increase harm. The latter being a significant factor in stopping many clinical trials before benefit to the diseased population is achieved.

TABLE 3 New risk factors for model of high immunological risk in Mayo cohort

Overall			Death censored		
Variable	Forward stepwise variable selection hazard ratio	*P*-value	Variable	Forward stepwise variable selection hazard ratio	*P*-value
Discrete DSA					
log(ACR)-0.72	3.45 (2.42, 4.92)	<.0001	log(ACR)-0.72	4.05 (2.72, 6.02)	<.0001
(eGFR - 51)/10	0.74 (0.62, 0.89)	.0014	(eGFR - 51)/10	0.71 (0.57, 0.88)	.0014

Model	C-statistic	Hosmer-Lemeshow *P*-value Rochester
Overall		
BirMay	0.751 (0.669-0.833)	.105
New DSA+ model	0.741 (0.668-0.814)	.055 (0.023; 0.042)
Death censored		
BirMay	0.784 (0.702-0.866)	.001
New DSA+ model	0.776 (0.698-0.854)	.133 (0.001; 0.017)

ACR, albumin creatinine ratio; BirMay, Birmingham-Mayo; DSA, donor specific antibodies; eGFR, estimated glomerular filtration rate.

TABLE 4 Validation of models in comparison in Paris cohort

	Censor type	C-statistic	Hosmer-Lemeshow statistic Paris
Birmingham	Death-censored graft failure	0.843 (0.694, 0.929)	<0.0001
BirMay		0.860 (0.676, 0.933)	0.0033
Birmingham	Overall graft failures		
BirMay			
New DSA+ model	Death-censored graft failure	0.830 (0.712-0.947)	0.0003
New DSA+ model	Overall graft failures	0.788 (0.684-0.893)	<0.0001

BirMay, Birmingham-Mayo; DSA, donor specific antibodies.

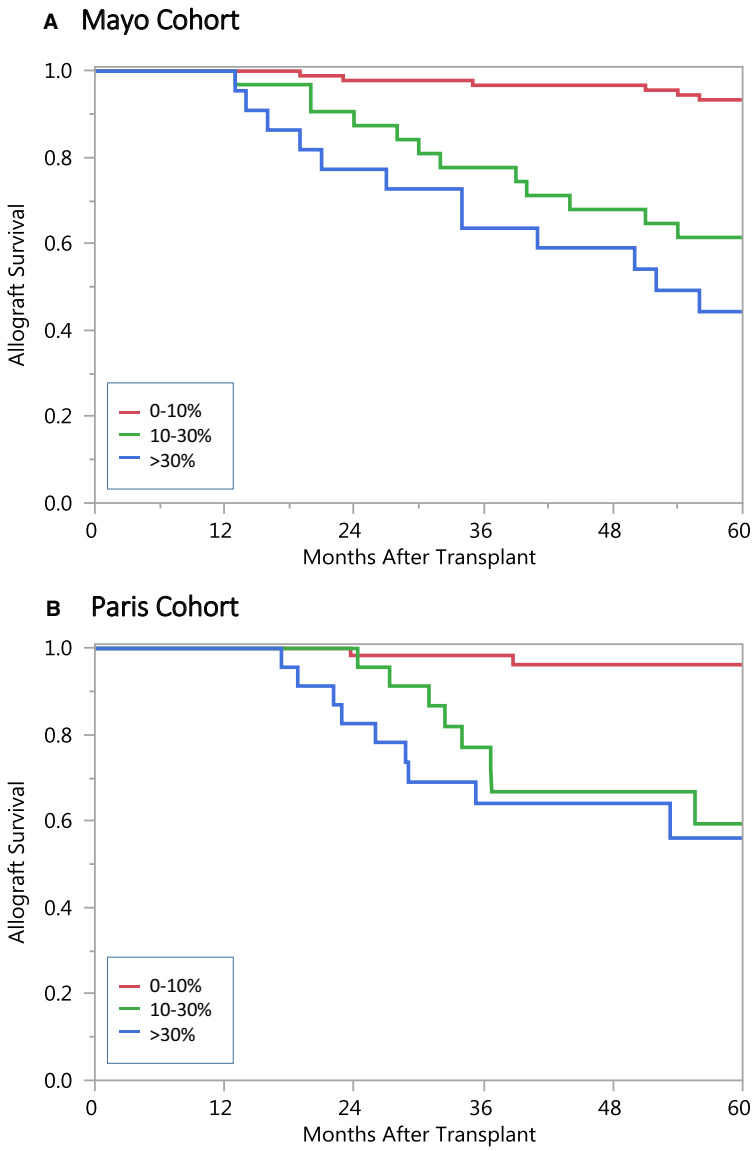


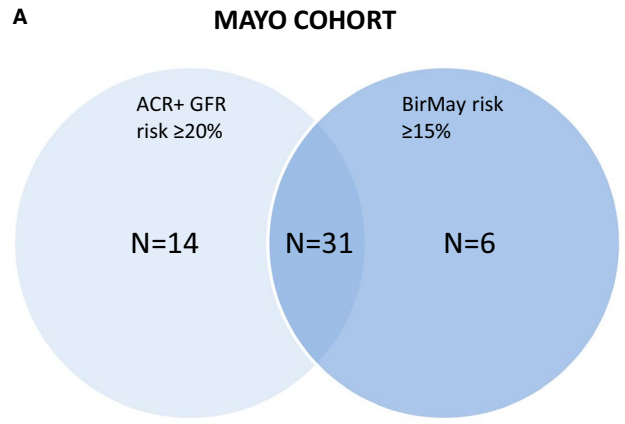
FIGURE 3 Kaplan-Meier allograft survival by BirMay risk score in Mayo cohort (A) and in Paris cohort (B) demonstrating similar risk assessment in both cohorts for 5-year allograft death censored allograft survival

4 | DISCUSSION

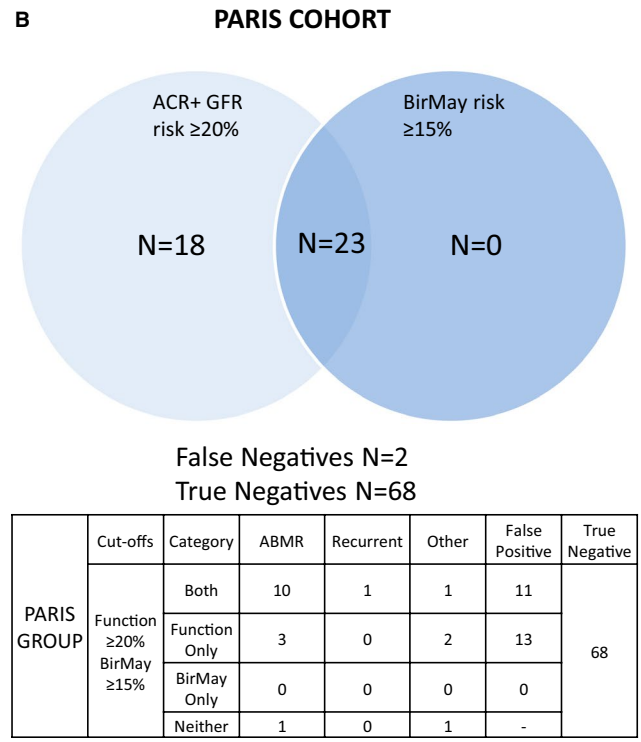
This study demonstrates that a mathematical model of graft loss based primarily on clinical parameters present at 1 year (the BirMay Predictor) predicts all-cause graft loss by 5 years well in both low-risk DSA- kidney transplant recipients and high-risk DSA+ recipients. In both cohorts, the addition of histologic data present on 1-year biopsies improved predictability slightly. However, the inclusion of histology allowed the ability to identify grafts that would fail due to a specific disease—cABMR. In addition, the model allowed us to exclude almost all grafts that would either not fail or would fail from other causes. Simulations suggest that this approach might markedly decrease the number of patients needed to show efficacy in intervention trials of cABMR.

While the current study focused on DSA+ recipients, it is likely that the inclusion of histology will be an important means in Modeling disease-specific graft loss in low-risk populations. The biology of cABMR is such that any allograft with glomerulitis at 1 year is likely to have a greater risk of graft loss as a result of progressive disease than other grafts without glomerulitis. This is true whether they had pretransplant DSA or developed DSA posttransplant. Since cABMR is a major cause of graft loss in both patients with and without DSA at the time of transplantation, the major difference between the two groups is the incidence of cABMR in the population and the relatively low rate of graft loss in the low-risk population overall. The validation of the BirMay model between the Mayo DSA+ and the Paris DSA+ groups demonstrates the effectiveness of using histology, particularly glomerulitis and interstitial fibrosis, as a biomarker for 5-year allograft survival—despite heterogeneities in the treatments between cohorts and the difference of living and deceased donation, not significantly affecting the model's performance.

Composite scores of graft loss could be used in clinical trials in two ways. First, in the setting of de novo therapy (ie, therapy started at the time of transplantation), a composite surrogate endpoint such as BirMay Prediction Score could be used to determine if the therapies lead to a differential effect on the composite endpoint, ie, a



	Cut-offs	Category	ABMR	Recurrent	Other	False Positive	True Negative
MAYO CLINIC	Function ≥20% BirMay ≥15%	Both	13	3	2	13	89
		Function Only	1	3	0	10	
		BirMay Only	1	0	0	5	
		Neither	5	1	1	-	



	Cut-offs	Category	ABMR	Recurrent	Other	False Positive	True Negative
PARIS GROUP	Function ≥20% BirMay ≥15%	Both	10	1	1	11	68
		Function Only	3	0	2	13	
		BirMay Only	0	0	0	0	
		Neither	1	0	1	-	

FIGURE 4 Venn diagram of causes of failure by predictive models in Mayo cohort. ABMR, antibody-mediated rejection; ACR, Albumin creatinine ratio; BirMay, Birmingham-Mayo; GFR, glomerular filtration rate

reduced BirMay score at 1 year posttransplant. Such a result will likely be more clinically meaningful in assessing the treatment effect on long-term graft survival than the current standard combined endpoint of biopsy-proven rejection in the first year combined with 1-year patient and graft survival. However, given the relatively low incidence of the factors that drive these models (low eGFR, high albumin/creatine ratio, fibrosis, and glomerulitis) means that it will be difficult to show an effect using de novo therapy. Conversely, these composite risk scores of graft loss could also be used to identify at 1 year patients at high risk for graft loss by 5 years. However, in this setting, assessing all-cause graft loss might not be as useful in this setting as identifying disease-specific graft loss type as treatments should be biologically targeted. For example, a study that includes a drug likely to be effective in ABMR should be used in trial in which the study population is enriched for graft loss due to cABMR. Thus, excluding grafts lost to other reasons using a process similar to the one we applied in the current study is clearly a more powerful approach to study design. Importantly, a disease-specific approach could be applied to a low-risk population in an intervention trial that sought to begin treatment at 1 year to improve clinical efficacy in randomized controlled trials, excluding patients for whom no clinical disease was likely to occur. The study by Viglietti et al demonstrated an effective predictive model in patients with ABMR diagnosed, predominantly with de novo DSA (63% of study population, 167/278)—but differs in the application to the incident ABMR population, to this study which looks to screen the at large transplant population.

Limitations of the current study are that the time point studied is only 1 year and it would be important to study if the model was applicable to data obtained at other time points. An underappreciated problem with these models is the fact that any composite score that is critically dependent on a Banff score is subject to miscalculation of the composite score due to the poor reproducibility of individual Banff scores.[12,13] One study suggested that agreement on g score was less than 50% among pathologists.[14] Using multiple pathologists to score a biopsy has been suggested as a means of decreasing the error rate and would be an important part of any clinical trial. Multiple readings of the biopsies were not possible in the current study but should be a part of future validation of any composite scoring system and likely would improve their correlation with outcomes. Yet another problem with this and similar studies is defining the cause of graft loss which depends on the judgement of the person assigning the cause and thus will always be problematic despite clear criteria. Finally, the number of graft losses in the current study is still relatively small and needs to be validated in larger cohorts, however in these higher risk cohort, collaboration between international multicenter alliances is necessary given the low frequency these are now performed. No novel molecular biomarkers were included in this study, which may help with differentiation further of higher risk cohorts in a risk Modeling strategy, however currently these are restrictive in widespread application—one of the goals of a model in its international application.

In conclusion, we have demonstrated that our existing model with histology has good predictive value in higher immunological risk patients, and a newer model did not show improvement. Research is required into specific pathways that can help clinicians understand causes of allograft failure, and thus lead to disease specific

interventions to improve allograft survival. This study demonstrates that prediction for graft failure and high negative predictive values can greatly enrich study populations with preexisting DSA for clinical trials in chronic antibody-mediated rejection.

ACKNOWLEDGMENTS

AB was funded by a National Institute for Health Research (Clinical Lecturership award). This article presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. This publication was made possible by CTSA Grant Number KL2 TR002379 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH.

AUTHOR CONTRIBUTIONS

Participated in research design: MMG, AB, BHS, MDS, RB. Participated in the writing of the paper: MMG, AB, BHS, MDS, RB, AL. Participated in the performance of the research: BHS, MMG, AB, KB, MDS, RB, AL, CL. Participated in data analysis: BHS, AB, MDS, RB.

DISCLOSURE

The authors of this manuscript have no conflicts of interest to disclose as described by the *American Journal of Transplantation*.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Andrew Bentall https://orcid.org/0000-0002-8495-9782
Byron H. Smith https://orcid.org/0000-0001-9318-2437
Manuel Moreno Gonzales https://orcid.org/0000-0001-6971-3971
Keisha Bonner https://orcid.org/0000-0001-7695-399X
Carrie A. Schinstock https://orcid.org/0000-0002-3274-3277

REFERENCES

1. Lefaucheur C, Loupy A, Hill GS, et al. Preexisting donor-specific HLA antibodies predict outcome in kidney transplantation. *J Am Soc Nephrol*. 2010;21(8):1398-1406.
2. Bentall A, Cornell LD, Gloor JM, et al. Five-year outcomes in living donor kidney transplants with a positive crossmatch. *Am J Transplant*. 2013;13(1):76-85.
3. Schinstock CA, Gandhi M, Cheungpasitporn W, et al. Kidney transplant with low levels of DSA or low positive B-flow crossmatch: an underappreciated option for highly-sensitized transplant candidates. *Transplantation*. 2017;101(10):2429-2439.
4. Bentall A, Cohney SJ. Overcoming preexisting alloantibody in renal transplantation-improving outcomes while reducing needs and costs. *Am J Transplant*. 2017;17(12):3003-3005.
5. Shabir S, Halimi JM, Cherukuri A, et al. Predicting 5-year risk of kidney transplant failure: a prediction instrument using data available at 1 year posttransplantation. *Am J Kidney Dis*. 2014;63(4):643-651.
6. Gonzales MM, Bentall A, Kremers WK, Stegall MD, Borrows R. Predicting individual renal allograft outcomes using risk models with 1-year surveillance biopsy and alloantibody data. *J Am Soc Nephrol*. 2016;27(10):3165-3174.
7. Viglietti D, Loupy A, Aubert O, et al. Dynamic prognostic score to predict kidney allograft survival in patients with antibody-mediated rejection. *J Am Soc Nephrol*. 2018;29(2):606-619.
8. Stegall MD, Diwan T, Raghavaiah S, et al. Terminal complement inhibition decreases antibody-mediated rejection in sensitized renal transplant recipients. *Am J Transplant*. 2011;11(11):2405-2413.
9. Amrouche L, Aubert O, Suberbielle C, et al. Long-term outcomes of kidney transplantation in patients with high levels of preformed DSA: the Necker high-risk transplant program. *Transplantation*. 2017;101(10):2440-2448.
10. Halimi JM, Matthias B, Al-Najjar A, et al. Respective predictive role of urinary albumin excretion and nonalbumin proteinuria on graft loss and death in renal transplant recipients. *Am J Transplant*. 2007;7(12):2775-2781.
11. Ix JH, Wassel CL, Stevens LA, et al. Equations to estimate creatinine excretion rate: the CKD epidemiology collaboration. *Clin J Am Soc Nephrol*. 2011;6(1):184-191.
12. Gough J, Rush D, Jeffery J, et al. Reproducibility of the Banff schema in reporting protocol biopsies of stable renal allografts. *Nephrol Dial Transplant*. 2002;17(6):1081-1084.
13. Marcussen N, Olsen TS, Benediktsson H, Racusen L, Solez K. Reproducibility of the Banff classification of renal allograft pathology. *Inter- and intraobserver variation. Transplantation*. 1995;60(10):1083-1089.
14. Smith B, Cornell LD, Smith M, et al. A method to reduce variability in diagnosing antibody mediated rejection in renal allograft biopsies using Banff 2013 [abstract]. *Am J Transplant*. 2017;17(S3):411-815.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bentall A, Smith BH, Gonzales MM, et al. Modeling graft loss in patients with donor-specific antibody at baseline using the Birmingham-Mayo (BirMay) predictor: Implications for clinical trials. *Am J Transplant*. 2019;19:2274-2283. https://doi.org/10.1111/ajt.15312