Received: 14 May 2020 | Revised: 3 August 2020 | Accepted: 4 September 2020

DOI: 10.1111/ajt.16311

ORIGINAL ARTICLE

AJT

A noninferiority design for a delayed calcineurin inhibitor substitution trial in kidney transplantation

Peter W. Nickerson[1,2,3] | Robert Balshaw[4] | Chris Wiebe[1,2,3] | Julie Ho[1,2,3] | Ian W. Gibson[2,5] | Nancy D. Bridges[6] | David N. Rush[1,2] | Peter S. Heeger[7]

[1]Department of Internal Medicine, Max Rady College of Medicine, University of Manitoba, Winnipeg, Canada

[2]Health Sciences Centre, Shared Health Services Manitoba, Winnipeg, Canada

[3]Department of Immunology, Max Rady College of Medicine, University of Manitoba, Winnipeg, Canada

[4]George and Fay Yee Centre for Healthcare Innovation, University of Manitoba, Winnipeg, Canada

[5]Department of Pathology, Max Rady College of Medicine, University of Manitoba, Winnipeg, Canada

[6]Division of Allergy, Immunology and Transplantation, National Institute of Allergy and Infectious Disease, Bethesda, Maryland

[7]Translational Transplant Research Center, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York

Correspondence
Peter Nickerson
Email: peter.nickerson@umanitoba.ca

Funding information
Canadian Institutes of Health Research, Grant/Award Number: 133636; National Institute of Allergy and Infectious Diseases, Grant/Award Number: U01-AI063594, R34-AI150361

Improving long-term kidney transplant outcomes requires novel treatment strategies, including delayed calcineurin inhibitor (CNI) substitution, tested using informative trial designs. An alternative approach to the usual superiority-based trial is a noninferiority trial design that tests whether an investigational agent is not unacceptably worse than standard of care. An informative noninferiority design, with biopsy-proven acute rejection (BPAR) as the endpoint, requires determination of a prespecified, evidence-based noninferiority margin for BPAR. No such information is available for delayed CNI substitution in kidney transplantation. Herein we analyzed data from recent kidney transplant trials of CNI withdrawal and "real world" CNI- based standard of care, containing subjects with well-documented evidence of immune quiescence at 6 months posttransplant—ideal candidates for delayed CNI substitution. Our analysis indicates an evidence-based noninferiority margin of 13.8% for the United States Food and Drug Administration's composite definition of BPAR between 6 and 24 months post-transplant. Sample size estimation determined that ~225 randomized subjects would be required to evaluate noninferiority for this primary clinical efficacy endpoint, and superiority for a renal function safety endpoint. Our findings provide the basis for future delayed CNI substitution noninferiority trials, thereby increasing the likelihood they will provide clinically implementable results and achieve regulatory approval.

KEYWORDS
clinical research / practice, clinical trial design, immunosuppressant - calcineurin inhibitor (CNI), immunosuppression / immune modulation, kidney transplantation / nephrology, rejection: acute

Abbreviations: ABMR, antibody-mediated rejection; BLA, biological license application; BPAR, Biopsy-proven acute rejection; CI, confidence interval; CNI, calcineurin inhibitor; CSA, cyclosporin A; CTOT, Clinical Trials in Organ Transplantation; DSA, donor-specific antibody; eGFR, estimated glomerular filtration rate; FDA, United States Food and Drug Administration; HLA, Human Leukocyte Antigen; MMF, mycophenolate mofetil; NDA, new drug application; NI, Noninferiority; RCT, Randomized control trial; SOC, Standard of care; TAC, tacrolimus; TCMR, T cell–mediated rejection; TTC, Transplant Therapeutics Consortium.

1 | INTRODUCTION

The transplant community is captive to its success. While long-term outcomes following organ transplantation remain suboptimal,[1] excellent short-term outcomes and the absence of validated surrogate endpoints effectively preclude the use of a superiority-based randomized control trial (RCT) design to achieve regulatory approval of investigational agents.[2–4] The last drugs approved by the United States Food and Drug Administration (FDA) based on RCTs demonstrating superiority were cyclosporin (CSA) in 1983,[5] using the clinical endpoint of patient and kidney graft survival; and sirolimus in 1999,[6] using the FDA's composite definition for biopsy-proven acute rejection (BPAR) (all BPAR events, whether or not clinically suspected, as well as death, graft loss, or lost to follow-up in those without BPAR).[7]

An alternative design strategy is a noninferiority (NI) RCT, which tests the hypothesis that an investigational agent is not unacceptably worse than the standard of care (SOC) already in use (and could be equivalent or superior).[8] Indeed, everolimus (2010), belatacept (2011), and extended-release tacrolimus (2015) were FDA approved on the basis of NI RCTs.[9–11] Noninferiority designs directly compare an investigational agent to an active control arm (i.e., SOC), but do not include a placebo arm (generally unethical in transplantation); may be accomplished using short-term clinical outcomes (i.e., BPAR); and because the interest is one-sided (not unacceptably worse than active control), can generally be informative using a smaller sample size, all of which enhances feasibility.

A "successful" NI trial will show a predetermined, acceptably small difference between the investigational agent and the active control; however, it must assume the superiority of the active control arm (i.e., SOC) to a historical placebo control by a prespecified margin (M1).[8] In other words, had the NI trial included a placebo, an active control to placebo difference of at least M1 would have been detected (Figure 1). Defining M1 is based on (a) historical evidence of sensitivity to SOC drug effects (i.e., consistently superior to placebo); (b) similarity of the new NI trial design being proposed to the historical trials, particularly with respect to event rates under SOC (i.e., constancy assumption); and (c) the quality of the new trial (i.e., removing elements that would minimize differences between treatments).[8] Failure to take these factors into consideration could lead to false conclusions—a trial could conclude that an investigational agent is noninferior to SOC when neither is superior to a placebo control (i.e., outcomes in the active control arm were worse than anticipated).

It is reasonable to consider a NI trial if the investigational agent offers a better side-effect profile, quality of life, improved medication dosing or adherence versus the SOC, or a reduced total cost of care.[8] For example, the BPAR efficacy failure rate for belatacept was elevated compared to a CSA-based regimen, but remained within an FDA specified 20% NI margin for the FDA's composite definition of BPAR (Figure 1).[7,12,13] Together with the finding that the belatacept-based regimen resulted in significantly better renal function, avoided calcineurin inhibitor (CNI) toxicity, and allowed for a simplified dosing schedule, meeting this NI endpoint led to regulatory approval of belatacept for the prophylaxis of BPAR.[7,10]

While the belatacept trials met the NI M1 endpoint, the results have not translated to clinical transplant practice; in real world cohorts BPAR rates and severity are greater over the initial 6 months posttransplant in CNI-free belatacept-based regimens initiated at the time of transplantation.[14–16] While alemtuzumab induction with belatacept can minimize the observed increase in BPAR,[17] CNI-free belatacept-based immunosuppression initiated at transplantation is not a widely accepted treatment regimen. Still, the transplant community remains interested in posttransplant CNI substitution trials to evaluate novel agents,[18,19] potentially through NI designs. Regulatory approval on the basis of a CNI substitution NI RCT requires that an evidence informed NI margin for BPAR be established relative to today's SOC (i.e., tacrolimus-based therapy)[7,8] and no such analysis has been reported. Even once-daily extended-release tacrolimus, approved for the conversion from twice-daily tacrolimus, required a de novo NI RCT to document noninferiority; a NI margin for the conversion NI RCT could not be justified to the FDA.[11] Herein

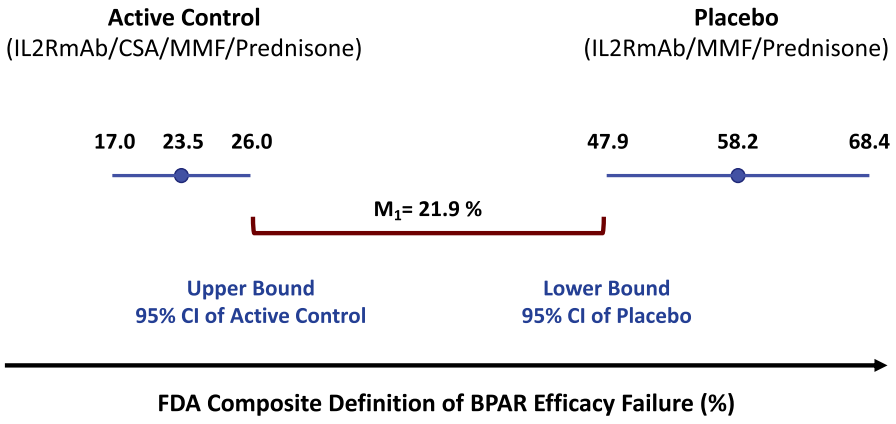

FIGURE 1 Noninferiority margin for the FDA composite definition of BPAR in a CSA-based CNI-free RCT design. Blue circles and lines represent efficacy failure point estimates and 95% confidence intervals (CI). The red bracket represents the noninferiority margin (M1) derived from the FDA's analysis of the historical CSA trial literature[7,33]

we make use of data from recently reported delayed CNI withdrawal and real world SOC kidney transplant trials to derive a NI margin for the FDA's composite definition of BPAR, in the context of a delayed CNI substitution RCT design.

2 | METHODS

The study was conducted with University of Manitoba Institutional Review Board approval (H2020:115).

2.1 | Selection of cohorts for analysis

2.1.1 | Direct comparison of efficacy failure of active control versus placebo

To construct a BPAR NI margin two items are required: (a) the BPAR efficacy failure rate on active control (i.e., SOC) and (b) the BPAR efficacy failure rate on placebo. Although frequently not available in transplant trials for ethical reasons, the FDA guidance report on noninferiority trial design prefers that estimation of a NI margin be constructed with formal RCTs directly comparing active control to placebo.[8] Therefore, a systematic review was conducted to identify RCTs that directly compared active control (i.e., SOC, tacrolimus [TAC]/mycophenolate mofetil [MMF]/steroid-based immunosuppression) versus placebo (i.e., MMF/steroid) in patients who were immune quiescent, as defined by a surveillance biopsy free of BPAR at the time of randomization (Data S1). By definition this required the systematic review to consider only CNI withdrawal RCTs to derive a placebo arm after immune quiescence had been established.

2.1.2 | Enhancing the accuracy of defining efficacy failure on active control

As the size of the RCTs directly comparing active control to placebo were small, to accurately set the BPAR efficacy failure rate for active control additional studies, without a placebo control, were included in the analysis. In this context, the FDA guidance recommends consideration of the historical evidence of sensitivity to the drug effect of the active control, and ensuring similarity of the proposed noninferiority trial to the historical trials (i.e., constancy assumption).[8] Therefore, a literature search was conducted for RCTs where the study mirrored the design of the aforementioned TAC withdrawal RCTs and the proposed delayed CNI substitution NI trial (Data S1). Specifically, RCTs that had a TAC/MMF/steroid active control arm, included a surveillance biopsy at 6 months posttransplant, and at least an 18-month follow-up period. Identified RCTs, with access to patient level data (publicly available or through contact with the study authors), further restricted the analysis to adult kidney transplant recipients meeting key inclusion criteria of the proposed delayed CNI substitution NI trial: (a) BPAR-free up to and including a 6-month surveillance biopsy, (b) a 6-month estimated glomerular filtration rate (eGFR) >20 ml/min/1.73 m^2, and (c) tacrolimus trough level target between 5 and 8 ng/ml beyond 6 months posttransplant. Finally, given the limited number of such highly characterized active control arm RCTs, we further supplemented the analysis with available patient level data from a single-center consecutive real-world cohort study meeting the aforementioned stringent inclusion criteria.

2.2 | BPAR diagnosis and definition

BPAR was diagnosed by either a "for cause" or a "surveillance" biopsy. BPAR was defined as (a) T cell–mediated rejection (TCMR) if histology met the Banff 1997 definition of Borderline rejection or higher (i.e., Banff lesion scores i ≥1 and t ≥1) or (b) antibody-mediated rejection (ABMR) if it met the Banff 2013 criteria for ABMR, including C4d-negative ABMR.[20,21]

2.3 | Efficacy failure definition

The primary clinical endpoint of the proposed CNI substitution noninferiority trial design is the BPAR efficacy failure rate between 6 and 24 months posttransplant. BPAR efficacy failure utilizes the FDA composite definition of BPAR (i.e., all BPAR events, whether or not clinically suspected, as well as death, graft loss, or lost to follow-up in those without BPAR).[7]

2.4 | Noninferiority margin estimation

Efficacy failure rates and 95% confidence intervals were determined for active control (Tac/MMF/steroid) and placebo (MMF/steroid) by aggregating event rates from the previously reported studies using meta-analytic methods suggested by the FDA guidance document.[8] The DerSimonian and Laird method, as demonstrated by the FDA, implied separate analyses of the two placebo arms and then of the five active control arms of the studies. However, analysis of event counts also suggests the use of a binomial regression model, with the possibility of between-study heterogeneity (i.e., a quasi-binomial regression model accounting for overdispersion relative to the binomial model). These regression models allowed a natural combination of the data from both active and the placebo arms from the five studies. Results from all three methods are reported, and this study adopted the results generating the most conservative (smallest) noninferiority margin as defined by the difference between the lower 95% confidence interval bound of the placebo control and the upper 95% confidence interval bound of the active control. This follows the FDA guidance document's approach to set a pre determined fixed noninferiority margin (M1).[7,8]

3 | RESULTS

3.1 | Direct comparison of efficacy failure of active control versus placebo

There are only two RCTs directly comparing active control (Tac/MMF/steroid, i.e., SOC) versus placebo (MMF/steroid, i.e., in the context of CNI withdrawal) in recipients with rigorously established histologic and serologic evidence of immune quiescence (i.e., BPAR negative and donor-specific antibody [DSA] negative) prior to CNI withdrawal.[22,23] The Clinical Trials in Organ Transplant (CTOT)-09 study enrolled immune quiescent recipients at 6 months posttransplant and observed a BPAR rate of 42.9% (6/14) in the CNI withdrawal arm (i.e., placebo, MMF/steroid) versus 0% (0/7) in the SOC arm (i.e., active control, TAC/MMF/steroid) (Table 1).[22] None of the patients in the CTOT-09 RCT experienced death, graft loss, or were lost to follow-up during the 18-month follow-up period. The Nantes tacrolimus withdrawal RCT enrolled immune quiescent recipients late posttransplant (i.e., ≥4-years).[23] Observed BPAR rates were 60% (3/5) in the CNI withdrawal arm (i.e., placebo, MMF/steroid) versus 0% (0/5) in the SOC arm (i.e., active control, TAC/MMF/steroid) (Table 1). None of the patients in the Nantes RCT experienced death, graft loss, or were lost to follow-up during the study. In both of these RCTs the data safety monitoring boards (DSMB) stopped the studies due to the high rates of alloimmune events in the CNI withdrawal arms.

3.2 | Enhancing the accuracy in defining efficacy failure on active control

Two Canadian multicenter RCTs (FKC-008 and FKC-014) and the Manitoba Consecutive Real World Cohort Study met the prespecified inclusion criteria to accurately set the efficacy failure rate for the active control arm (Tac/MMF/steroid); that is access to patient level data, a surveillance biopsy at 6 months posttransplant and a minimum of 18 months of follow-up.[24–26] In both the FKC-014 RCT and the Manitoba Consecutive Real World Cohort Study a DSA test result was also available 6 months posttransplant and individuals were included if the DSA was negative up to that point. In the FKC-008 RCT, a 6-month posttransplant DSA evaluation was not available. However, none of the FKC-008 individuals included in the current analysis had ABMR on the 6-month surveillance biopsy and none developed ABMR during follow-up including on a 24-month surveillance biopsy—the latter providing inferential evidence of absence of DSA at 6 months posttransplant.

In the FKC-008 RCT, 57% (119/210) of kidney transplant recipients were considered immune quiescent at 6 months and they experienced a BPAR rate of 7.56% (9/119) between 6 and 24 months posttransplant (Table 1). None of these patients experienced death or graft loss, and none were lost to follow-up during the study. In the FKC-014 RCT, 47% (102/217) of kidney transplant recipients were immune quiescent at 6 months and they experienced a BPAR rate of 10.78% (11/102) between 6 and 24 months posttransplant (Table 1). None of these patients in the FKC-014 RCT experienced death or graft loss, and only 1.96% (2/102), who were without BPAR during the study were lost to follow-up. In the Manitoba Consecutive Real World Cohort Study, 47% (152/322) of kidney transplant recipients were immune quiescent at 6 months; they experienced a BPAR rate of 5.26% (8/153); 1.32% (2/152) experienced death; 0.67% (1/152) experienced graft loss; and 0.67% (1/152) were lost to follow-up without BPAR between 6 and 24 months posttransplant (Table 1).

3.3 | Noninferiority margin estimation

Three analytic approaches were used to evaluate aggregate efficacy failure point estimates and 95% confidence intervals (Table 2). The DerSimonian and Laird approach, the most conservative in this instance when used for this purpose, identified an efficacy failure rate for the FDA composite definition for BPAR to be 8.6% (5.9, 11.4) for active control (Tac/MMF/steroids) versus 47.4% (25.2, 69.6) for placebo (MMF/steroids). This allows for the derivation of an evidenced-informed fixed noninferiority margin (M1) for the FDA composite definition of BPAR of 13.8% (Table 2, Figure 2).

TABLE 1 Efficacy failure rates on placebo versus tacrolimus-based therapy

Published study	Immunosuppression	BPAR (B-TCMR+)	Death	Graft loss	Lost to follow-up	FDA efficacy failure
Active control versus placebo RCT						
CTOT–09 [22]	MMF/Pred	42.9% (6/14)	0% (0/14)	0% (0/14)	0% (0/14)	42.9% (6/14)
	Tac/MMF/Pred	0% (0/7)	0% (0/7)	0% (0/7)	0% (0/7)	0% (0/7)
Nantes [23]	MMF/Pred	60.0% (3/5)	0% (0/5)	0% (0/5)	0% (0/5)	60.0% (3/5)
	Tac/MMF/Pred	0% (0/5)	0% (0/5)	0% (0/5)	0% (0/5)	0% (0/5)
Active control RCT						
FKC–008 [24]	Tac/MMF/Pred	7.56% (9/119)	0% (0/119)	0% (0/119)	0% (0/119)	7.56% (9/119)
FKC–014 [25]	Tac/MMF/Pred	10.78% (11/102)	0% (0/102)	0% (0/102)	1.96% (2/102)	12.75% (13/102)
Consecutive Real-World Cohort Study						
Manitoba [26]	Tac/MMF/Pred	5.26% (8/152)	1.32% (2/152)	0.67% (1/152)	0.67% (1/152)	7.89% (12/152)

3.4 | Clinical trial sample size estimations

To estimate the sample size required to conduct a delayed CNI substitution NI RCT between 6 and 24 months posttransplant the following parameters were set: noninferiority margin for a composite BPAR efficacy failure rate of 13.5%; power 80%; and alpha 0.025 one-sided. Under the usual assumption of equality under the noninferiority alternative hypothesis, the composite BPAR efficacy failure rate was explored over a range of 6% to 12%. The calculations were performed with a 1:1 randomization (Table 3a) and a 2:1 randomization (Table 3b) to an investigational versus active control arm.[27] Under these conditions a CNI substitution NI RCT between 6 and 24 months posttransplant would require up to 182 total recipients in a 1:1 randomization and 206 total recipients in a 2:1 randomization.

In both CNI-free and CNI substitution RCT study designs, one of the common endpoints evaluated is improvement in eGFR in the investigational arm versus the active control arm.[12,13,15,28,29] A sample size of 200 total recipients in a 1:1 randomization and 225 total recipients in a 2:1 randomization would be sufficient to yield 80% power to detect a clinically relevant difference in eGFR of 8 ml/min/1.73 m^2, assuming a conservative standard deviation of 20, using a two-sided alpha of 0.05.

In the current analysis 47-57% of individuals receiving SOC therapy met the definition of immune quiescence on the basis of a surveillance biopsy ±DSA evaluation at 6 months posttransplant. Therefore, at least 500 transplant recipients on Tac/MMF/steroid would need to be screened (conservatively assuming only 45% [225/500] of those screened will qualify for enrollment into the RCT) to achieve the number required to evaluate both a primary outcome of noninferiority for BPAR efficacy failure and superiority for a key secondary safety outcome (i.e., renal function, eGFR).

TABLE 2 Point estimates, 95% confidence intervals, and noninferiority (NI) margins for efficacy failure

Analysis method	Active control (Tac/MMF/steroids)	Placebo (MMF/steroids)	NI margin
DerSimonian & Laird	8.64% [5.85, 11.44]	47.44% [25.24, 69.63]	13.80%
Binomial	8.83% [6.38, 12.11]	47.37% [26.78, 68.89]	14.67%
Quasibinomial	8.83% [6.62, 11.69]	47.37% [28.81, 66.69]	17.12%

4 | DISCUSSION

The transplant community is actively calling for innovative trial designs, biomarkers/surrogate endpoints, and the development of research consortia to encourage biopharmaceutical companies to reinvest in drug development to address the unmet needs of transplant recipients.[3,4,30] To this end, the Paris Transplant Group and the Transplant Therapeutics Consortium (TTC) are actively developing and validating early (eg, 1- year) composite surrogate endpoints in kidney transplantation to allow for short-term accelerated drug approval.[31,32] However, as mandated by the regulatory authorities, such a trial must be continued (e.g., 5^+years) to demonstrate superiority for a clinical endpoint (i.e., patient and graft survival) to achieve

TABLE 3 Sample size estimates for a delayed CNI substitution noninferiority trial

NI margin =13.5%	Composite BPAR efficacy failure			
a				
Randomized 1:1	6%	8%	10%	12%
Investigational agent	49	64	78	91
Active control arm	49	64	78	91
Power 80%, alpha =0.025 one-sided				
b				
Randomized 2:1	6%	8%	10%	12%
Investigational agent	73	95	117	137
Active control arm	37	48	59	69
Power 80%, alpha =0.025 one-sided				

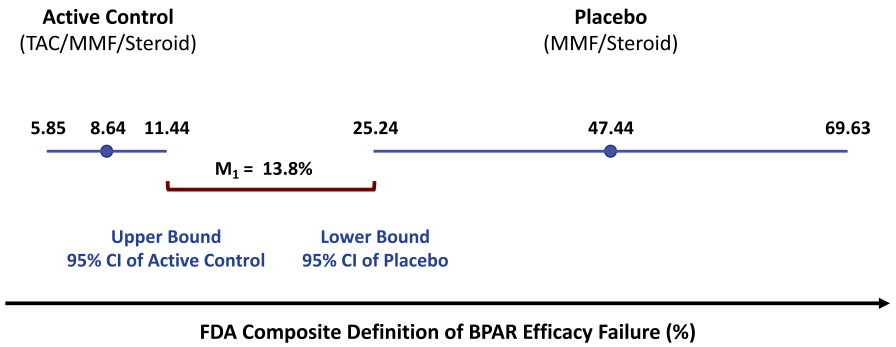

FIGURE 2 Noninferiority margin for the FDA composite definition of BPAR in a delayed Tac-based CNI substitution RCT design. Blue circles and lines represent efficacy failure point estimates and 95% confidence intervals (CI). The red bracket represents the noninferiority margin (M1) derived from the study's analysis

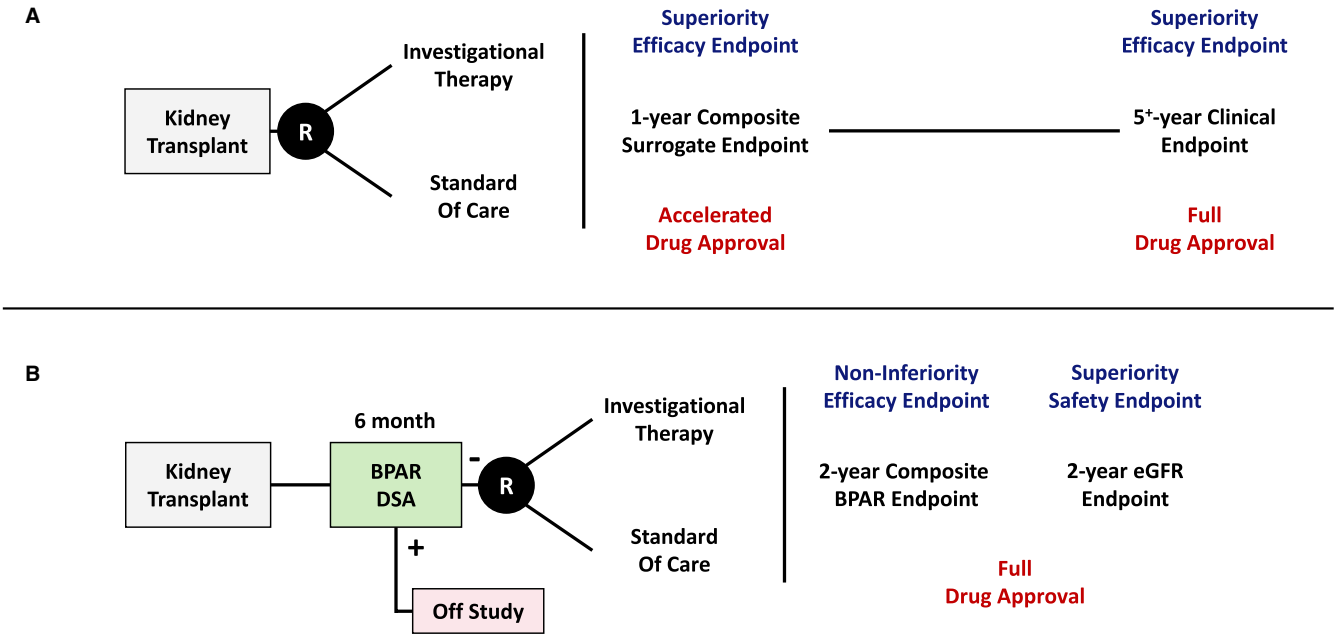

FIGURE 3 Transplant trial designs with the potential to achieve full regulatory drug approval. (A) CNI-free RCT design using a 1-year composite surrogate endpoint and a 5^{+}-year clinical endpoint (i.e., patient and graft survival); (B) delayed CNI substitution RCT design with the requirement for immune quiescence at 6 months posttransplant prior to randomization, and an 18-month follow-up period using a 2-year posttransplant composite BPAR efficacy endpoint and a 2-year eGFR safety endpoint. R = randomized

full drug approval (Figure 3A). Our current analysis outlines a potential alternative path to full regulatory drug approval in a 2-year time frame (Figure 3B).

A BPAR noninferiority RCT has two clear advantages: (a) prophylaxis of BPAR is recognized by the FDA as a primary clinical endpoint for the evaluation of efficacy using the composite BPAR definition; and (b) an evidenced-based NI margin in a well-designed noninferiority RCT can serve as the basis for establishing drug effectiveness in an FDA biologics license application (BLA) or new drug application (NDA).[8] In the current analysis the NI margin for the FDA composite definition of BPAR (13.8%, Figure 2) was justified in the context of a delayed CNI substitution RCT design initiated 6 months posttransplant with an 18-month follow-up period to 24 months posttransplant (Figure 3B). We derived this BPAR NI margin based on stringent selection criteria and patient level data. We included two CNI withdrawal RCTs that directly compared the current SOC active control (Tac/MMF/steroid) to placebo (MMF/steroid) and three SOC active control (Tac/MMF/steroid) studies that all used similarly stringent inclusion criteria—a surveillance biopsy in all subjects, and a DSA assessment in >70% of subjects—to determine immune quiescence prior to the start of the follow-up period. This is a critical point; without such an assessment one could be substituting an investigational agent when BPAR is already subclinically present, which could confound the intended comparison between the active control and the investigational agent.

The FDA requires that a NI margin "should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative".[8,33] In this regard the aggregate point estimate and 95% confidence interval (CI) of the BPAR efficacy failure rate calculated for placebo (MMF/steroid) is derived from the withdrawal trial's placebo arms, which had a small number of predominantly low risk (i.e., living donor, Caucasian) transplant recipients.[22] This likely led to a lower point estimate and a wider 95% CI than would be seen in a real world cohort (i.e., it is conservative). Similarly, the point estimate for the active control arm does not rely solely on the small numbers in the CNI withdrawal RCTs, but rather includes multicenter RCTs plus a consecutive real world cohort study. By including these latter three studies the observed BPAR rate in the active control arm is increased. Finally, the DerSimonian & Laird analytical approach, which the FDA has used for NI margin estimation,[33] led to the widest confidence intervals resulting in the smallest NI margin (Table 2). Together, our approach for deriving a NI margin for the FDA composite definition of BPAR adheres to the FDA guidance to be "suitably conservative."

Two design issues of the proposed delayed CNI substitution NI RCT deserve comment. *Why not initiate enrolment earlier than 6 months?* Adams et al. reported a higher rate of BPAR when CNI withdrawal was initiated at 3 months as compared to after 6 months posttransplant.[15] One could postulate that immune quiescence is more established by 6 months (i.e., further away from early posttransplant inflammatory events, and at a time when donor antigen presenting cells are likely eliminated). *Why a follow-up period of 18 months rather than 1-year following substitution?* In the Adams' study a slow, rather than a rapid CNI taper (over 3-4 months rather than 1-2 months), was met with less efficacy failure.[15] Hence, an 18-month follow-up period allows for a slower CNI taper while maintaining a minimum follow-up for 1-year off CNI prior to evaluation. Moreover, a key endpoint of a CNI substitution RCT is superiority in

eGFR in the investigational versus active control arm. In one delayed CNI substitution RCT it took an 18- month follow-up period to see a difference in eGFR between the investigational and TAC-based active control arms.[29] Thus the proposed 6- to 24-month delayed CNI substitution non-inferiority RCT design balances key considerations, while allowing a reasonable time to observe definitive outcomes.

While including Borderline TCMR (i.e., as defined by Banff 1997)[20] in the BPAR definition may be controversial, recent evidence supports its inclusion. In the CNI withdrawal RCTs, Borderline TCMR was included in the DSMB's decision to halt the trials—the DSMB conservatively considered Borderline TCMR after an immune quiescent biopsy to represent loss of control of the primary alloimmune response.[22,23] Moreover, the appearance of Human Leukocyte Antigen (HLA) DR/DQ DSA, which associates with worse graft survival,[26,34] frequently accompanied Borderline TCMR.[22] Together with evidence that Borderline TCMR correlates with the degree of HLA DR/DQ molecular mismatch, is linked to subsequent Banff ≥IA TCMR and/or DSA development, and is associated with allograft loss,[34,35] the aggregate evidence supports including Borderline TCMR in a conservative definition of BPAR.

When considering a noninferiority design, CNI-free RCTs initiated at transplant may be perceived as preferable to a delayed CNI substitution RCT. The stringent enrolment criteria including the need for documenting immune quiescence (via a surveillance biopsy) are limiting (i.e., in our analysis only ~50% of the kidney transplant recipients met the inclusion criteria for the CNI substitution study), hence regulatory approval on the basis of a delayed CNI substitution RCT results in a restricted label indication. On the other hand, emerging noninvasive biomarkers may ultimately be able to detect eligible patients without a biopsy,[22,36,37] and stratification based on risk enriches study populations so as to increase the likelihood of trial success.[38] Moreover, a CNI-free NI RCT initiated at transplant does not ensure real world translation to equivalent efficacy when it does not also reflect the stringent inclusion criteria of the CNI-free NI RCT,[12–16] and delayed CNI substitution can be more effective than a CNI-free strategy initiated at transplant in retaining efficacy for BPAR prophylaxis.[15,28,29] Thus, well-designed CNI substitution NI RCTs should not be dismissed by the transplant or biopharmaceutical communities—regulatory approval of an agent, even for a substantial subset of kidney transplant recipients, is a superior outcome to no approval for all recipients. In fact, identification of responsive patient subsets is a specific goal of personalized medical care.

Interpreting the results of a NI trial requires careful consideration even when it rules out a difference between the SOC and the investigational agent larger than M1, which is critical to support a conclusion of effectiveness.[8,39] Indeed, a trial can demonstrate noninferiority, but the test drug can be found to be superior, equivalent, noninferior with loss of efficacy in comparison to the SOC, or even inferior in comparison to the SOC (Figure 4). This has led the FDA to develop in their guidance document the concept of a more conservative NI margin M2 (e.g., M2 being 50% of M1), which is defined as "the largest clinically acceptable difference (degree of inferiority) of the test drug compared to the active control"—it is based on clinical judgement.[8] Conversely, the FDA acknowledges that a larger M2 (<M1) may be clinically justified, if the investigational agent "were shown to have some important advantage (e.g. safety or on a secondary endpoint)."[8] For example, regulatory approval of belatacept was on the basis of the belatacept arms remaining within the 20% NI margin (M1) set by the FDA to conclude an effect over placebo.[7,33] At the FDA advisory committee, the debate was whether belatacept's

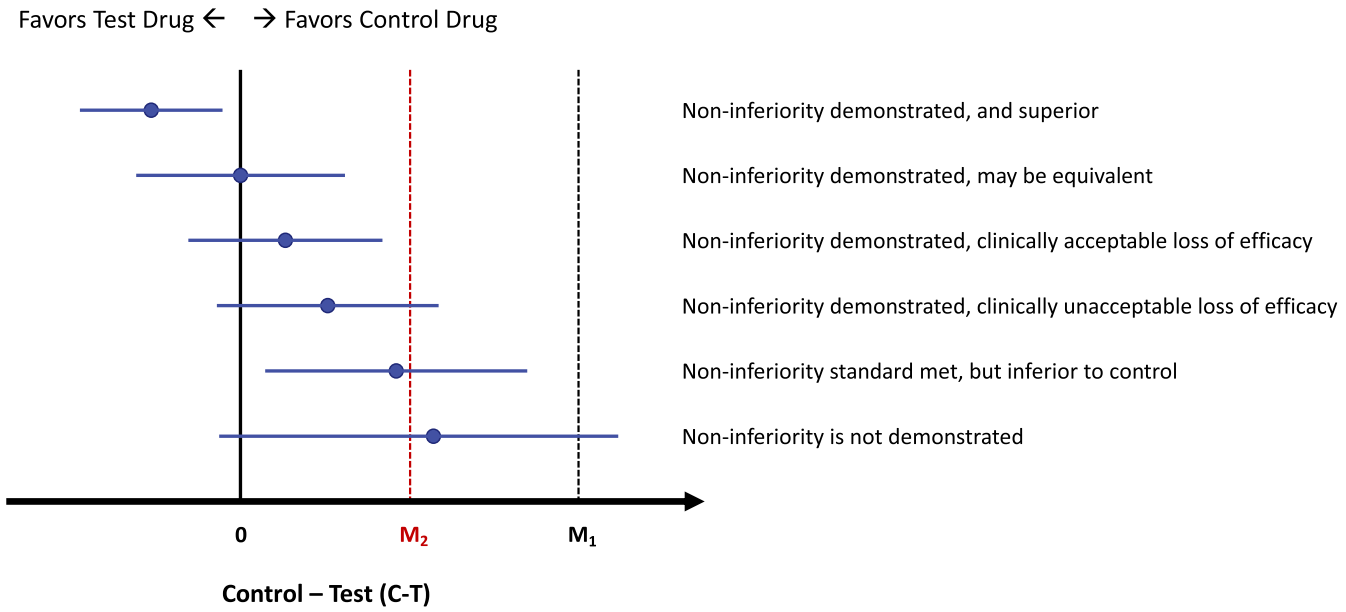

FIGURE 4 Potential outcomes in a noninferiority trial depicting control drug – test drug differences. Blue circles and lines represent control minus test point estimates and 95% confidence intervals (CI). M1—the entire effect of the active control drug assumed to be present in the noninferiority trial. M2 —the largest clinically acceptable difference (degree of inferiority) of the test drug compared with the active control drug[8]

loss of efficacy in BPAR prophylaxis as compared to SOC was clinically acceptable. In the end the committee decided the benefits of belatacept balanced the loss of BPAR prophylaxis efficacy observed in the trials.

In the context of the current study, had there been larger patient numbers in the CNI withdrawal RCT placebo arms, the placebo confidence interval would have narrowed significantly, which in turn would have supported a larger M1 than was calculated in our analysis (Figure 2). As such the study's M1 of 13.8% for the FDA's composite definition of BPAR is conservative. Whether or not one should design a delayed CNI substitution NI trial with an even more conservative BPAR NI margin (e.g., M2 of 7%) to require the test drug to retain a greater proportion of the SOC efficacy will need to be weighed against the increase in sample size it will demand versus the potential benefits of the test drug in comparison to the SOC.

4.1 | Limitations

The size of the RCTs directly comparing active control versus placebo is limiting; however, as discussed this results in a conservative NI margin estimation. The number of RCTs available to accurately set BPAR efficacy failure of the active control arm are few; CNI-based RCTs and CNI substitution RCTs generally do not include a surveillance biopsy to be able to rigorously document a state of immune quiescence.[29,40–42] Where a CNI substitution RCT included a pre-enrollment surveillance biopsy, it did so at 3 months posttransplant, and allowed early TCMR prior to enrolment (i.e., it does not mirror the proposed delayed CNI substitution RCT design).[43] A single-arm prospective cohort study may be regarded as suboptimal for inclusion in a formal NI margin estimation, but this is not unique. The FDA used a single-arm study to establish the efficacy failure rate for placebo when establishing a NI margin for CSA.[7,44] The limited availability of RCTs with patient level data that met the stringent inclusion criteria make the use of a single-center consecutive cohort on SOC reasonable, and as it reflects a real-world recipient population it contributes to a conservative NI margin estimation.

5 | CONCLUSION

The transplant community urgently requires viable short-term strategies for regulatory drug approval to address the unmet needs of kidney transplant recipients. Our analysis derives an evidence informed NI margin for the FDA composite definition of BPAR in the context of a delayed CNI substitution RCT between 6 and 24 months posttransplant. Sample size estimation further determined that a CNI substitution NI RCT is viable to evaluate both a primary clinical efficacy endpoint (FDA composite definition of BPAR) as well as superiority for a key FDA safety endpoint (renal function [eGFR]). This may enable a CNI substitution NI RCT that has the potential to rapidly lead to regulatory drug approval and positively impact patient care.

ACKNOWLEDGMENTS

The authors are indebted to Dr. Renata Albrecht and Dr. Paolo Cravedi for their critical review. PSH, PWN, and IWG are funded by the Clinical Trials in Organ Transplantation (CTOT), NIH U01-AI063594. PWN, RB, and PSH are funded by an Clinical Trial planning grant, NIH R34-AI150361. PWN, JH, CW, IWG, and DNR are funded by an operating grant from CIHR, 133626. PWN holds the Flynn Family Chair in Renal Transplantation. JH holds a CIHR New Investigator award.

DISCLOSURE

The authors of this manuscript have conflicts of interest to disclose as described by the *American Journal of Transplantation*. DNR is a consultant with Astellas Pharma. PWN is a consultant with Vitaeris Inc. and Renalytix AI Inc.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Peter W. Nickerson https://orcid.org/0000-0002-7393-7799
Robert Balshaw https://orcid.org/0000-0002-2455-8792
Chris Wiebe https://orcid.org/0000-0002-1006-3545
Julie Ho https://orcid.org/0000-0002-8342-9093
Ian W. Gibson https://orcid.org/0000-0003-3505-0262
Nancy D. Bridges https://orcid.org/0000-0002-1140-5201
David N. Rush https://orcid.org/0000-0002-3451-9447
Peter S. Heeger https://orcid.org/0000-0003-4673-6913

REFERENCES

1. Hart A, Smith JM, Skeans MA, et al. OPTN/SRTR 2018 Annual Data Report: Kidney. *Am J Transplant*. 2020;20(Suppl. 1):20–130.
2. Leas BF, Uhl S, Sawinski DL, et al. *Calcineurin Inhibitors for Renal Transplant*. Rockville, MD: Agency for Healthcare Research and Quality (US); 2016. https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/calcineurin-inhibitor_research.pdf. Accessed February 10 2020
3. Stegall MD, Morris RE, Alloway RR, Mannon RB. Developing new immunosuppression for the next generation of transplant recipients: the path forward. *Am J Transplant*. 2016;16(4):1094–1101.
4. O'Connell PJ, Kuypers D, Mannon RR, et al. Clinical trials for immunosuppression in transplantation; the case for reform and change in direction. *Transplantation*. 2017;101(7):1527–1534.
5. FDA Center for Drug Evaluation and Research – Sandimmune approval letter 11/15/1983. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=050574. Accessed April 15 2020
6. FDA Center for Drug Evaluation and Research – Sirolimus approval letter 09/15/1999. https://www.accessdata.fda.gov/drugsatfda_docs/nda/99/21083A_Rapamune_appltr.pdf. Accessed April 15 2020

7. Archdeacon P, Dixon C, Belen O, Albrecht R, Meyer J. Summary of the US FDA approval of belatacept. *Am J Transplant*. 2012;12(3):554–562.
8. FDA. Non-inferiority clinical trials to establish effectiveness – Guidance for Industry Nov 2016. https://www.fda.gov/media/78504/download. Accessed January 30 2020
9. FDA Center for Drug Evaluation and Research – Everolimus approval letter 04/20/2010. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2010/021560s000_zortress_toc.cfm. Accessed April 15 2020
10. FDA Center for Drug Evaluation and Research – Belatacept approval letter 06/15/2011. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/125288Orig1s000TOC.cfm. Accessed April 15 2020
11. FDA Center for Drug Evaluation and Research – Envarsus XR approval letter 10/07/2015. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2015/206406Orig1_toc.cfm. Accessed April 17 2020
12. Vincenti F, Charpentier B, Vanrenterghem Y, et al. A phase III study of belatacept-based immunosuppression regimens versus cyclosporine in renal transplant recipients (BENEFIT Study). *Am J Transplant*. 2010;10(3):535–546.
13. Durrbach A, Pestana JM, Pearson T, et al. A phase III study of belatacept versus cyclosporine in kidney transplants from extended criteria donors (BENEFIT-EXT Study). *Am J Transplant*. 2010;10(3):547–557.
14. Mannon R, Armstrong B, Stock PG, et al. Avoidance of CNI and steroids using belatacept – Results of the Clinical Trials in Organ Transplantation 16 trial. *Am J Transplant*. 2020;00:1–10. https://doi.org/10.1111/ajt.16152
15. Adams AB, Goldstein J, Garrett C, et al. Belatacept combined with transient calcineurin inhibitor therapy prevents rejection and promotes improved long-term renal allograft function. *Am J Transplant*. 2017;17(11):2922–2936.
16. De Graav GN, Baan CC, Clahsen-van Groningen MC, et al. A randomized controlled clinical trial comparing belatacept with tacrolimus after de novo kidney transplantation. *Transplantation*. 2017;101(10):2571–2581.
17. Woodle ES, Kaufman DB, Shields AR, et al. Belatacept-based immunosuppression with simultaneous calcineurin inhibitor avoidance and early corticosteroid withdrawal: A prospective, randomized multicenter trial. *Am J Transplant*. 2020;20(4):1039–1055.
18. Darres A, Ulloa C, Brakemeier S, et al. Conversion to belatacept in maintenance kidney transplant patients: A retrospective multicenter European study. *Transplantation*. 2018;102(9):1545–1552.
19. Sethi S, Najjar R, Peng A, et al. Outcomes of conversion from calcineurin inhibitor to belatacept-based immunosuppression in HLA-sensitized kidney transplant recipients. *Transplantation*. 2020;104(7):1500–1507.
20. Racusen L, Solez K, Colvin RB, et al. The Banff 97 working classification of renal allograft pathology. *Kidney Int*. 1999;55(2):713–723.
21. Haas M, Sis B, Racusen LC, et al. Banff 2013 meeting report: Inclusion of C4d-negative antibody-mediated rejection and antibody-associated arterial lesions. *Am J Transplant*. 2014;14(2):272–283.
22. Hrick D, Formica RN, Nickerson P, et al. Adverse outcomes of tacrolimus withdrawal in immune-quiescent kidney transplant recipients. *J Am Soc Nephrol*. 2015;26(12):3114–3122.
23. Dugast E, Soulillou J-P, Foucher Y, et al. Failure of calcineurin inhibitor (tacrolimus) weaning randomized trial in long-term stable kidney transplant recipients. *Am J Transplant*. 2016;16(11):3255–3261.
24. Rush D, Arlen D, Boucher A, et al. Lack of benefit of early protocol biopsies in renal transplant patients receiving TAC and MMF: A randomized study. *Am J Transplant*. 2007;7(11):2538–2545.
25. Cockfield SM, Wilson S, Campbell PM, et al. Comparison of the effects of standard vs. low- dose prolonged-release tacrolimus with or without ACEi/ARB on the histology and function of renal allografts. *Am J Transplant*. 2019;19(6):1730–1744.
26. Wiebe C, Kosmoliaptsis V, Pochinco D, et al. HLA-DR/DQ molecular mismatch: A prognostic biomarker for primary alloimmunity. *Am J Transplant*. 2019;19(6):1708–1719.
27. Hylown Consulting LLC."Compare 2 proportions 2-sample non-inferiority or superiority: Power and sample size calculators : Hylown", Retrieved from: http://www.poerandsamplesize.com/Calculators/Compare-2-Proportions/2-Sample-Non-Inferirority-or-superiority. Accessed February 09 2020.
28. Rostaing L, Massari P, Garcia VD, et al. Switching from calcineurin inhibitor-based regimens to a belatacept-based regimen in renal transplant recipients: A randomized phase II study. *Clin J Am Soc Nephrol*. 2011;6(2):430–439.
29. Grinyo J, Alberu J, Contieri FLC, et al. Improvement in renal function in kidney transplant recipients switched from cyclosporine or tacrolimus to belatacept: 2-year results from the long-term extension of a phase II study. *Transplant Int*. 2012;25(10):1059–1064.
30. Stegall MD, Smith B, Bentall A, Schinstock C. The need for novel trial designs, master protocols, and research consortia in transplantation. *Clin Transplant*. 2020;34(1):e13759.
31. Stegall MD, Troy Somerville K, Everly MJ, et al. The importance of drug safety and tolerability in the development of new immunosuppressive therapy for transplant recipients: The Transplant Therapeutics Consortium's position statement. *Am J Transplant*. 2019;19(3):625–632.
32. Loupy A, Aubert O, Orandi B, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation system. *BMJ*. 2019;366:l4923.
33. Cardiovascular and Renal Drugs Advisory Committee meeting. March 1, 2010. Briefing materials: https://wayback.archive-it.org/7993/20170114003119/http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM201858.pdf. Accessed February 02 2020.
34. Wiebe C, Rush DN, Gibson IW, et al. Evidence for the alloimmune basis and prognostic significance of borderline T cell-mediated rejection. *Am J Transplant*. 2020;00:1–10. https://doi.org/10.1111/ajt.15860
35. Nankivell BJ, Agrawal N, Sharma A, et al. The clinical and pathological significance of borderline T cell–mediated rejection. *Am J Transplant*. 2019;19(5):1452–1463.
36. Zhang W, Yi Z, Keung KL, et al. A peripheral blood gene expression signature to diagnose subclinical acute rejection. *J Am Soc Nephrol*. 2019;30(8):1481–1494.
37. Friedewald JJ, Kurian SM, Heilman RL, et al. Development and clinical validity of a novel blood-based molecular biomarker for subclinical acute rejection following kidney transplant. *Am J Transplant*. 2019;19(1):98–109.
38. FDA. Enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products: Guidance for industry March 2019. https://www.fda.gov/media/121320/download. Accessed January 16 2020.
39. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials*. 2011;12:106. https://doi.org/10.1186/1745-6215-12-106
40. Berger SP, Sommerer C, Witzke O, et al. Two-year outcomes in de novo renal transplants receiving everolimus-facilitated calcineurin inhibitor reduction regimen from the TRANSFORM study. *Am J Transplant*. 2019;19(11):3018–3034.
41. De Fijter JW, Holdass H, Oyen O, et al. Early conversion from calcineurin inhibitor- to everolimus-based therapy following kidney transplantation: results of the randomized ELEVATE trial. *Am J Transplant*. 2017;17(7):1853–1867.
42. The 3C study collaborative group, Haynes R, Blackwell L, et al. Campath, calcineurin inhibitor reduction, and chronic allograft nephropathy (the 3C study) – results of a randomized controlled clinical trial. *Am J Transplant*. 2018;18(6):1424–1434.

43. De Sandes-Freitas TV, Felipe CR, Campos EF, et al. Subclinical lesions and donor-specific antibodies in kidney transplant recipients receiving tacrolimus-based immunosuppressive regimen followed by early conversion to sirolimus. *Transplantation*. 2015;99(11):2372–2381.
44. Vincenti F, Ramos E, Brattstrom C, et al. Multicenter trial exploring calcineurin inhibitors avoidance in renal transplantation. *Transplantation*. 2001;71(9):1282–1287.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Nickerson PW, Balshaw R, Wiebe C, et al. A noninferiority design for a delayed calcineurin inhibitor substitution trial in kidney transplantation. *Am. J. Transplant*. 2021;21:1503–1512. https://doi.org/10.1111/ajt.16311