Generic immunosuppression in solid organ transplantation: systematic review and meta-analysis

Amber O Molnar,1,2 Dean Fergusson,1,3 Anne K Tsampalieros,1,2 Alexandria Bennett,1,2 Nicholas Fergusson,1,2 Timothy Ramsay,1 Greg A Knoll1

ABSTRACT

OBJECTIVE
To compare the clinical efficacy and bioequivalence of generic immunosuppressive drugs in patients with solid organ transplants.

DESIGN
Systematic review and meta-analysis of all studies comparing generic with innovator immunosuppressive drugs.

DATA SOURCES
Medline and Embase from 1980 to September 2014.

REVIEW METHODS
A literature search was performed for all studies comparing a generic to an innovator immunosuppressive drug in solid organ transplantation. Two reviewers independently extracted data and assessed quality of studies. Meta-analyses of prespecified outcomes were performed when deemed appropriate. Outcomes included patient survival, allograft survival, acute rejection, adverse events and bioequivalence.

RESULTS
1679 citations were screened, of which 50 studies met eligibility criteria (17 randomized trials, 15 non-randomized interventional studies, and 18 observational studies). Generics were compared with Neoral (cyclosporine) (32 studies), Prograf (tacrolimus) (12 studies), and Cellcept (mycophenolate mofetil) (six studies). Pooled analysis of randomized controlled trials in patients with kidney transplants that reported bioequivalence criteria showed that Neoral (two studies) and Prograf (three studies) were not bioequivalent with generic preparations according to criteria of the European Medicines Agency. The single Cellcept trial also did not meet bioequivalence. Acute rejection was rare but did not differ between groups. For Neoral, the pooled Peto odds ratio was 1.23 (95% confidence interval 0.64 to 2.36) for kidney randomized controlled trials and 0.66 (0.40 to 1.08) for observational studies. For kidney observational studies, the pooled Peto odds ratios were 0.98 (0.37 to 2.60) for Prograf and 0.49 (0.09 to 2.56) for Cellcept. Meta-analyses for non-renal solid organ transplants were not performed because of a lack of data. There were insufficient data reported on patient or graft survival. Pooling of results was limited by inconsistent study methods and reporting of outcomes. Many studies did not report standard criteria used to determine bioequivalence. While rates of acute rejection seemed similar and were relatively rare, few studies were designed to properly compare clinical outcomes. Most studies had short follow-up times and included stable patients without a history of rejection.

CONCLUSIONS
High quality data showing bioequivalence and clinical efficacy of generic immunosuppressive drugs in patients with transplants are lacking. Given the serious consequences of rejection and allograft failure, well designed studies on bioequivalence and safety of generic immunosuppression in transplant recipients are needed.

Introduction

With the recent expiry of patents for commonly prescribed immunosuppressive drugs such as Prograf (tacrolimus) and Cellcept (mycophenolate mofetil), the use of generics in solid organ transplantation has become controversial.1-5 Generic substitution has the potential for huge cost savings and is therefore an essential component to maintaining comprehensive and equitable healthcare, especially within public healthcare systems, where limited resources must be fairly distributed.16 For certain classes of drugs, such as cardiovascular drugs, published studies show no difference in outcomes between the generic and innovator preparations, and generic substitution is therefore not a concern.7 The substitution of drugs with a narrow therapeutic index, such as immunosuppressants, however, is more controversial. There is considerable concern among physicians and patients that generic and innovator immunosuppressants are not equivalent to one another.8-11 Before approval, each generic drug must show bioequivalence to the innovator version in healthy adults, but there is no requirement to show bioequivalence or clinical efficacy in patients with transplants. Many argue that the current criteria are not sufficient as such patients often have comorbidities and are taking multiple drugs, which could alter the pharmacokinetics of a drug.13-14 Also, bioequivalence might not...
necessarily mean equivalence in terms of clinical effectiveness, which could lead to catastrophic consequences in a patient with a transplant (for instance, loss of the graft). Based on current guidelines, however, clinical effectiveness is assumed if bioequivalence can be shown.12

The concerns raised in the transplant community have led to the recommendation that patients and healthcare providers pay careful attention to drug formulations and monitor drug concentrations more often if a patient is switched to a generic preparation.12,13,14 Europe and Canada have also adopted more stringent bioequivalence criteria for tacrolimus and cyclosporine.12,15 Certain countries in Europe have gone one step further, with the United Kingdom now requiring that the brand of tacrolimus be specified on all prescriptions to avoid inadvertent switching and Denmark banning the generic substitution of tacrolimus and cyclosporine products.16,17 These recommendations are not based on high quality evidence, and many, including regulatory agencies, argue that the methods of determining bioequivalence are reliable and sufficient.18,19 If patients and physicians remain doubtful of the equivalence of generic immunosuppressive drugs, this will limit the cost saving potential of these drugs from underprescribing and more frequent laboratory monitoring when a generic is prescribed.1,20

We investigated the clinical efficacy, safety, and bioequivalence of generic immunosuppressive drugs compared with innovator drugs in solid organ transplant recipients. Our focus was on clinically important outcomes, such as patient survival, transplant allograft survival, and acute rejection.

Methods
Search strategy
We performed a comprehensive systematic search of articles published in peer reviewed journals using Medline and Embase (from 1980 to 4 September 2014). The search was carried out with the assistance of a librarian experienced in systematic reviews. A structured search strategy (appendix 1) was conducted with controlled vocabulary and relevant key terms to enhance sensitivity. Reference lists of included papers and previous reviews were hand searched for additional relevant studies. The search was not restricted by study design or language.

Study selection
One investigator (AOM) performed an initial screen of identified titles and abstracts. Those deemed to be clearly irrelevant were removed on the initial screen. Two independent reviewers (AOM and AKT) performed a second screen to identify potentially relevant studies. If no abstract was available, the full text was obtained unless the article could be confidently excluded by title alone. If there was any doubt as to whether or not a study could be excluded, a full text screen was performed to reduce the likelihood of incorrect exclusion of a relevant study. Two reviewers (AOM and AB) obtained and independently screened full text versions of potentially eligible studies to determine their eligibility based on the selection criteria. Any disagreements during the screening process were resolved through discussion among the authors in accordance with the selection criteria.

We included randomized controlled trials, nonrandomized interventional studies, and observational studies if they evaluated the innovator version of an immunosuppressive drug compared with at least one generic version of the same drug in recipients of solid organ (heart, lung, liver, pancreas, kidney, small bowel, or combinations of these organs) transplants. The comparative evaluation had to include at least one clinical efficacy/safety outcome (death, transplant failure, acute rejection, marker of graft function (such as serum creatinine concentration), use of healthcare (such as admission to hospital), infection, drug concentration, or other serious adverse event) or the determination of bioequivalence. There are different definitions of bioequivalence depending on the jurisdiction. In the United States, the Food and Drug Administration (FDA) requires that the 90% confidence interval of the mean ratio (generic/innovator) for the area under the curve (AUC) of the concentration-time curve and the peak concentration (Cmax) should be between 80% and 125%.12,21,22 For drugs with a narrow therapeutic index (such as cyclosporine and tacrolimus), Health Canada (HC) and the European Medicines Agency (EMA) have tighter limits for the AUC acceptance interval (90% to 112% and 90% to 111%, respectively).12,15 For cyclosporine, the EMA has also imposed tighter limits for the Cmax acceptance interval (90% to 111%).15 In this analysis, we assessed bioequivalence using both the FDA definition and the tighter standards from the EMA and Health Canada. We excluded comparative studies that evaluated Sandimmune as this formulation is no longer used in clinical practice. We also excluded case reports, case series, studies including children, and studies performed on animals or conducted in vitro. Studies with a small number of patients were not specifically excluded.

Data extraction and synthesis
Three investigators (AOM, AB, NF) abstracted data. Two different investigators independently abstracted data from each eligible study (see appendix 1 for data abstraction form). Assessed variables related to the organization and outcome of the studies included study design, setting (country), characteristics of the population studied, organ transplanted, number of study participants, immunosuppressive drug studied, and reporting of relevant outcomes. The primary clinical efficacy outcome was acute rejection, and the primary bioequivalence outcome was the mean ratio (and 90% confidence interval) for the Cmax and AUC. The methodological quality of eligible randomized trials (parallel and cross over designs) was evaluated with the Cochrane risk of bias assessment tool.20 The methodological quality of observational and non-randomized experimental studies was evaluated with a checklist
outlined by Wells and colleagues. Three domains were assessed with this checklist: study design, confounding, and selective reporting. Quality assessment was performed for studies as a whole and separately for each prespecified primary outcome. When data were available only in figures, we used the GNU image manipulation program (GIMP 2.8; www.gimp.org/) to extract data.

**Patient involvement**
There was no patient involvement in this study.

**Statistical analysis**
Descriptive methods were used to present the data by type of immunosuppressive drug, type of organ transplant, and outcome. For the randomized trials and non-randomized interventional studies, we pooled the mean ratio and the 90% confidence interval for the $C_{\text{max}}$ and AUC. Data were analyzed with the inverse variance method with a random effects model and presented as a pooled mean ratio with a 90% confidence interval. The standard errors of the AUC and $C_{\text{max}}$ mean ratios were calculated with the 90% or 95% confidence intervals and T statistic of the study. Continuous efficacy outcomes (such as serum creatinine concentration) were pooled when deemed appropriate with the inverse variance method and presented as weighted mean differences. Dichotomous efficacy outcomes (such as acute rejection) were pooled with the Peto method and are presented as the Peto odds ratio. We chose this as it is the preferred estimate when cells contain 0 events. Crossover trials were treated as parallel group trials in the analysis if individual patient level data, sequence specific data, or correlation coefficients were not available. We performed a prespecified sensitivity analysis for cyclosporine that excluded studies involving SangCya as it was recalled in 2000 and is no longer available. Heterogeneity was assessed with the $I^2$ statistic. Meta-analyses were performed with RevMan 5.3. Data from observational studies were not pooled for the outcome of bioequivalence because of concerns about the validity of the results. Data from crossover trials and before/after studies were not pooled for the outcome of acute rejection because of concerns about the statistical and clinical validity of the results. The reporting of this systematic review is in accordance with PRISMA guidelines.

**Results**

**Eligible studies**
The electronic database search identified 2558 records, and a further six records were identified from reference lists. After independently reviewing the title and abstract of all potentially relevant records, 201 articles were retrieved and reviewed in full text. Of these, 50 studies were found to meet inclusion criteria. Figure 1 outlines the study selection.

**Characteristics of patients and studies**
The characteristics of the 50 eligible studies are outlined in tables A–C in appendix 2. Eligible studies included recipients of kidney, heart, and liver transplants. Study designs included randomized controlled...
trials (n=17; eight crossover and nine parallel), non-randomized interventional studies (n=15), and observational studies (n=18; cohort and before/after designs). Innovator drugs studied included Neoral (cyclosporine; n=32 studies,71-82 28 in kidney transplants, three in heart transplants, and one in liver transplants), Prograf (tacrolimus; n=12 studies,63-74 seven in kidney transplants, one in heart transplants, one in liver transplants, and three in a mixture of liver, kidney, or heart transplants), and Cellcept (mycophenolate mofetil; n=6 studies,75-80 five in kidney transplants and one in liver transplants). Neoral was compared with 12 different generics (Imnoloral, Equroral, Gengraf, Cysporin, Zinografi-ME, Neoplanta, Consupren, SangCya (Sang-35), Sigmasporin Microral, Pliva, Cicloral, and Arpimune); Prograf was compared with four different generics (Tacni, Tacrobell, Adoport, and Sandoz-tacrolimus); and Cellcept was compared with five different generics (Myfenax, Medis, Linfonex, Mycept, and Mycomol).

Neoral studies—Sample sizes ranged from 11 to 221 patients, though one study did not report the number of included patients.43 The average age in most studies was 40-50 years. Eight studies included incident transplants59-60 50-62 (table A in appendix 2).

Prograf studies—Sample sizes ranged from 25 to 234 patients. The average age in most studies was between 50 and 60, except for the study by Robertson and colleagues, which included patients aged 60 or older.76 Five studies included incident transplant recipients64 65 68 71 76 (table B in appendix 2).

Cellcept studies—The sample sizes ranged from 5 to 56 patients. Three studies included incident transplant recipients75 77 79 (table C in appendix 2).

Assessment of study quality

Randomized trials—The methodological quality of the randomized controlled trials was generally poor (fig A in appendix 3). Only four trials reported on the methods used for randomization,44 62 68 74 and allocation concealment was poorly reported. There were only two double blind trials,79 74 for one of which there were concerns about selective reporting of outcomes.39

Non-randomized interventional studies and observational studies—The quality assessment of the non-randomized studies is presented in appendix 4. All non-randomized interventional studies had a before/after design, with patients serving as their own controls. Observational studies were a mixture of retrospective and prospective designs. Most cohort studies identified patients as receiving generic or innovator drug based on era (for example, the innovator was used in 2007 and the generic in 2008). Many studies did not account for potential confounders, such as dose adjustments, in their analyses.

Outcomes

Pharmacokinetic and clinical outcomes are summarized in appendices 5 and 6, respectively. Where applicable, most studies explicitly stated that there was a mg:mg conversion from innovator to generic, though some studies allowed dose adjustments after the initial conversion while others did not clearly state whether or not dose adjustments were allowed.

Neoral studies

Ten studies (nine in kidney transplants; one in liver transplant) reported 90% confidence intervals for the primary pharmacokinetic outcomes of the Cmax and AUC mean ratios.32 37 38 44 46 55 56 58-60 All reported 90% confidence intervals for the Cmax and AUC mean ratios fell within the FDA guidelines for bioequivalence. Seven studies, however, reported 90% confidence intervals for the Cmax and AUC mean ratios that did not meet the stricter EMA bioequivalence criteria,32 37 38 56 58-60 and five studies did not meet the Health Canada bioequivalence criteria, based on the 90% confidence intervals of the AUC mean ratio.36 56 58-60 When we pooled results for the two randomized kidney trials,38 56 the FDA criteria for bioequivalence were met while the EMA/Health Canada guidelines were not met (table 1). Pooling of the results for non-randomized interventional kidney studies met FDA/EMA/Health Canada bioequivalence criteria.32 37 44 46 56 58-60 There was variable reporting of other secondary pharmacokinetic outcomes (table A in appendix 5).

For clinical outcomes, follow-up ranged from one week to a year. Acute rejection was reported in 16 studies (13 included kidney transplants, three included heart transplants). Eight studies (three in heart transplants, five in kidney transplants) reported no episodes of acute rejection,31-41 44 48 49 55-57 while one study had a large number of acute rejections (n=59).53 Two studies reported a significant increase in acute rejection for those receiving generic drugs compared with Neoral (39% v 25% (P=0.045)53 and 60% v 25% (P<0.055)52).

<p>| Table 1 | Meta-analysis of bioequivalence between Neoral (cyclosporine) and generic immunosuppressive drugs in patients undergoing kidney transplant |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>No of studies</th>
<th>No of patients</th>
<th>Point estimate (pooled 90% CI)</th>
<th>I² (%)</th>
<th>Meets FDA criteria</th>
<th>Meets EMA criteria</th>
<th>Meets HC criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AUC mean ratio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized controlled trials</td>
<td>2</td>
<td>60</td>
<td>0.93 (0.89 to 0.98)</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Non-randomized interventional studies</td>
<td>7</td>
<td>251</td>
<td>1.00 (0.98 to 1.02)</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Cmax mean ratio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized controlled trials</td>
<td>2</td>
<td>60</td>
<td>0.90 (0.85 to 1.02)</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Non-randomized interventional studies</td>
<td>7</td>
<td>251</td>
<td>0.98 (0.95 to 1.01)</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Cmax=maximum concentration; AUC=area under curve; FDA=Food and Drug Administration; EMA=European Medicines Agency; HC=Health Canada.
Acute rejection was pooled for the randomized parallel group kidney trials that measured the outcome at six months and longer.\textsuperscript{52,57,62} We found no significant difference in acute rejection (pooled Peto odds ratio 1.23, 95% confidence interval 0.64 to 2.36) (fig 2). We also found no significant difference in acute rejection when we pooled results from observational kidney studies (0.66, 0.40 to 1.08) (fig 2). Eight studies reported graft loss, and nine reported patient survival. Of the studies that reported graft loss, only two episodes occurred, and these were in the Neoral arm\textsuperscript{50,62} (table A in appendix 6). Data on serum creatinine concentration were pooled for the randomized parallel group kidney trials that measured creatinine at six months and longer.\textsuperscript{52,61,62} Serum creatinine was not significantly different between the Neoral and generic arms (mean difference 6.45 µmol/L, 95% confidence interval −0.67 to 13.57 µmol/L; fig A in appendix 3).

**Pragrof studies**

Three kidney randomized controlled trials reported the primary pharmacokinetic outcome of the $C_{\text{max}}$ and AUC.\textsuperscript{63,68,74} When we pooled data, the 90% confidence intervals for both the AUC and $C_{\text{max}}$ mean ratios did not meet FDA, EMA, or Health Canada bioequivalence criteria (table 2). When each study was examined individually, two studies did not meet FDA, EMA, or Health Canada bioequivalence (table B in appendix 5).

For clinical outcomes, follow-up ranged from four weeks to a year, with only one study following up patients beyond six months. Of the 10 studies (six in kidney, two in kidney/liver, one in kidney/liver/heart, one in liver transplants) that reported the outcome of acute rejection, five reported no events. The study by Yu and colleagues was the only one that reported a difference (not significant) in the incidence of acute rejection (8% v 0% in the Prograf and generic arms, respectively; $P=0.08$).\textsuperscript{71} We pooled acute rejection for the observational kidney studies that measured the outcome at six months and longer.\textsuperscript{64,66,67} No significant difference in rejection was found (Peto odds ratio 0.98, 95% confidence interval 0.37 to 2.60) (fig 3). The study with the most acute rejection (eight in the Prograf arm and nine in the generic arm) was also the only study that reported patients with graft loss (six in the Prograf arm and eight in the generic arm; $P=0.78$).\textsuperscript{64} The outcome of serum creatinine concentration was measured at different time points in each study and therefore could not be pooled. There was no significant difference in serum creatinine concentration between the Prograf or generic arms in any of the studies that reported the outcome (table B in appendix 6).

**Cellcept studies**

One study (crossover trial in kidney transplants) reported the AUC and $C_{\text{max}}$ mean ratios and 90% confidence intervals (0.959 (0.899 to 1.023) and 0.873 (0.787 to 0.968), respectively).\textsuperscript{78} These values did not fulfill the FDA, EMA, or Health Canada requirements for bioequivalence (table C in appendix 5).

For clinical outcomes, follow-up ranged from three months to two years. One randomized controlled trial in kidney transplants reported acute rejection; there was only one event in each arm.\textsuperscript{73} We pooled acute rejection for the observational kidney studies that measured the outcome at six months or greater and found no significant difference in rejection (Peto odds ratio 0.49, 95% confidence interval 0.09 to 2.56) (fig 4). One study in liver transplants reported acute rejection; there were no events. One study in kidney transplants reported graft loss, but there were no events\textsuperscript{79} (table C in appendix 6).

**Sensitivity analysis**

We repeated the Neoral bioequivalence meta-analysis after removing kidney studies that included the generic Sang-Cya,\textsuperscript{58,60} which did not significantly change the results.

**Discussion**

**Principal findings**

This analysis included 50 studies comparing generic with innovator immunosuppressive drugs in more than 3130 recipients of solid organ transplants. When we pooled the results of randomized trials, we found that generic cyclosporine met FDA but not EMA or Health Canada criteria for bioequivalence; neither generic
tacrolimus nor mycophenolate mofetil met any agency’s criteria for bioequivalence. There was no significant difference in acute rejection for generic cyclosporine, tacrolimus, or mycophenolate mofetil compared with innovator versions, but the methodological quality of most studies was poor, limiting inferences that can be made from this data. Other important clinical outcomes, such as survival of patients and grafts, were inconsistently reported with few events occurring.

Less than a third of studies on cyclosporine reported the standard criteria needed to determine bioequivalence. As well, the single published study conducted in liver transplants examined the generic SangCya, which is no longer approved for use. Available data suggest that the cyclosporine generics studied are bioequivalent to Neoral when used in kidney transplants based on FDA criteria. It remains unclear if Neoral and generic cyclosporine are bioequivalent based on EMA and Health Canada criteria and if bioequivalence exists in recipients of a non-renal solid organ transplant.

Only three of 12 tacrolimus studies reported bioequivalence criteria. All three were randomized controlled trials and conducted in kidney transplants, with differing bioequivalence results. Included patients differed between the studies, which could explain the inconsistent results. Another potential explanation is that each study compared a different generic preparation with Prograf. Only one of six studies on mycophenolate mofetil reported bioequivalence criteria. This study was a randomized controlled trial conducted in stable kidney transplants and did not show bioequivalence. Overall, it remains unclear if generic tacrolimus or mycophenolate mofetil are bioequivalent to Prograf and Cellcept, respectively, and there is a complete lack of data in recipients of a non-renal solid organ transplant.

Acute rejection was found to be no different for all generics. For generic cyclosporine, we included studies on only kidney transplant in the meta-analysis for acute rejection because of a lack of data for other solid organ transplants. Two cyclosporine kidney studies found a higher rate of acute rejection in the generic arm than in the Neoral arm. These studies, however, had overall high rates of acute rejection, included only incident transplants, were single centre, and were retrospective with historical controls. For generic tacrolimus, a low number of events significantly limited the ability to pool data, and acute rejection could be pooled only for kidney observational studies. The confidence intervals for two of the three meta-analyzed tacrolimus studies were extremely wide because of the low number of events. Five tacrolimus studies reporting acute rejection included recipients of liver and/or heart transplant, four of which reported no events while one study with incident liver transplants reported a greater number of events (although not significant) in the Prograf arm. This study was retrospective, with the Prograf arm composed of historical controls; era effect could therefore potentially explain the greater number of events in the Prograf arm. For generic mycophenolate mofetil, three kidney studies and one liver study reported acute rejection, with no differences being found. Because of limited data, we included only kidney observational studies in the meta-analysis. Once again, a small number of events resulted in wide confidence intervals. Overall, the data for acute rejection must be interpreted with caution given the low number of observed events and largely observational nature of the data. There is also a paucity of data for non-renal solid organ transplants.

Comparison with other studies
To our knowledge, this study is the first systematic review and meta-analysis published on this topic. The number of generic immunosuppressants available on the market continues to rise. As a result, the prescription and safety of these drugs has become a topic of concern, leading to the publication of multiple narrative reviews and editorials.13 15 18-20 81-83 Our findings of an overall lack of high quality data supporting the bioequivalence and clinical efficacy of generic immunosuppressants in solid organ transplantation agree with that of a recently published narrative review.17 The results of our review do not refute the current general consensus in the literature that any generic substitution should be performed with caution and that drug monitoring should be increased after substitution, with the recognition that none of these recommendations is based on high quality evidence.13 19 108 Our results also do not refute concern that the current
Strengths and limitations
Our review is comprehensive with the inclusion of all types of comparative peer reviewed published studies for three of the most commonly used immunosuppressants, and all of our primary outcomes of interest were prespecified. Unfortunately, the conclusions of our review are limited for several reasons. Included studies inconsistently reported outcomes, limiting our ability to pool results. Although included studies were designed to examine the equivalence between an innovator drug and a generic, only a minority reported the 90% confidence intervals for the AUC and C max mean ratios, which are the standard criteria needed to determine bioequivalence. Many studies reported only drug concentrations, which are not sufficient to comment on the bioequivalence of two drug preparations and can actually be misleading. In a recent study by Robertsen and colleagues that examined the bioequivalence of Prograf to a generic version, the generic was not found to be bioequivalent based on AUC and C max mean ratio criteria, though trough drug concentrations were identical. This highlights the need for formal pharmacokinetic studies when bioequivalence in commented on. Of the minority of studies that reported criteria for bioequivalence, the results were inconsistent and inconclusive potentially because of varying study methods and sample sizes.

Studies measured acute rejection at various time points, the methods of determining acute rejection were inconsistent (clinical judgment v biopsy), and most meta-analyzed studies were observational, making the results potentially more prone to bias and confounding. Several studies that measured acute rejection were not included in the meta-analysis because of real concerns about validity and clinical applicability of the data. About a third of all studies included in this review were interventional before/after or conversion studies and about half of included trials were crossover design. These study designs can be useful when pharmacokinetic outcomes, such as drug concentrations, are examined but not informative for clinical outcomes, such as acute rejection. Crossover trials generally have a short follow-up time and can have a carryover effect; before/after studies are subject to an era effect bias, which is also a concern with many of the published cohort studies because of the use of historical controls. Published interventional before/after studies all specified stable graft function as an inclusion criterion, and many also specified no recent episodes of acute rejection, which creates selection bias. The inclusion of only stable patients in most studies is likely a contributing factor to the low number of observed events.

Overall, the quality of studies and study reporting were poor. The ideal standard for determining bioequivalence and for comparing clinical outcomes is a randomized crossover trial and a randomized parallel group trial, respectively, which most of the studies were not. Of the third of studies that were randomized trials, most were open label with unclear methods of randomization and allocation concealment. Also, many studies either allowed dose adjustments to occur before drug concentrations were measured or did not clearly report the timing of measurements in relation to any dose adjustments or if dose adjustments were allowed to occur. This is obviously a concern as a patient should receive the same dose of both innovator and generic in comparisons of any sort of pharmacokinetic outcome.

Conclusions
In conclusion, high quality data showing bioequivalence and clinical efficacy of generic immunosuppressants in solid organ transplants are lacking. There is insufficient evidence to provide reassurance that generics are equivalent to innovator immunosuppressants, but there are also no data to firmly suggest that generics are not equivalent and therefore unsafe. Well designed randomized controlled trials comparing clinical endpoints are unlikely to be performed because of the large sample sizes that would be needed and the additional cost. Given the concerns in the transplant community, well conducted bioequivalence studies that include transplant recipients would be a reasonable alternative. In addition, switching from innovator to generic or from one generic formulation to another without physician input should be limited.

Contributors: All authors fulfill the ICMJE criteria for authorship. AOM conceived and designed the study; acquired, analysed, and interpreted the data; and drafted and finally approved the manuscript. DF designed the study, interpreted the data; and drafted and finally approved the manuscript. AKT, AB, and NF acquired and interpreted the data and drafted and finally approved the manuscript. GAK conceived and designed the study; interpreted the data; and drafted and finally approved the manuscript. TR helped with the analysis, interpreted the data and finally approved the manuscript. All authors agree to be accountable for all aspects of the work. AOM and GAK are guarantors.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. AOM received salary support from the KRESCENT Program.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: GAK has received research grants from Astellas Canada, Roche Canada, Novartis Canada, and Pfizer Canada outside the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

Transparency: The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.


© BMJ Publishing Group Ltd 2015

Appendix 1: Search strategy and data abstraction form
Appendix 2: Characteristics of eligible studies (tables A-C)
Appendix 3: Supplementary figures A and B
Appendix 4: Quality assessment of non-randomized studies
Appendix 5: Pharmacokinetic outcomes
Appendix 6: Clinical outcomes