Transplant International

REVIEW

Trial design and endpoints in clinical transplant research

Simon R. Knight[1,2,*], Peter J. Morris[1,2], Stefan Schneeberger[3] & Liset H. M. Pengel[1,2,*]

1 Centre for Evidence in Transplantation, Clinical Effectiveness Unit, Royal College of Surgeons of England and the London School of Hygiene and Tropical Medicine, University of London, London, UK
2 Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK
3 Department of Visceral Transplant and Thoracic Surgery, Innsbruck Medical University, Innsbruck, Austria

Correspondence
Simon Knight, Nuffield Department of Surgical Sciences, Oxford Transplant Centre, Churchill Hospital, Old Road, Oxford OX3 7LE, UK.
Tel.: 01865 227131;
e-mail: simon.knight@nds.ox.ac.uk

* These two authors contributed equally to the research and preparation of this review.

SUMMARY

The number of clinical trials in solid organ transplantation is progressively increasing year on year, but the quality of design and reporting still varies considerably. The constraints on organ availability, improving short-term outcomes, ethics and timescales involved in organ transplantation present unique challenges for trials in this field. An understanding of the methodology and potential pitfalls in clinical research is essential both to interpret trial results and to design robust studies. This review summarizes the scope and quality of reporting in existing transplant clinical trials and details aspects of clinical trial methodology with particular relevance to transplantation. We highlight initiatives designed to improve the quality of this process to ensure that the results of clinical trials are robust, well reported and of use in everyday clinical practice.

Transplant International 2016; 29: 870–879

Key words
clinical trials, evidence, outcomes, PICO, randomized controlled trial

Received: 20 August 2015; Revision requested: 1 October 2015; Accepted: 30 December 2015; Published online: 17 February 2016

Introduction

A large part of the increasing success of organ transplantation stems from a drive to improve the evidence base with which we make treatment decisions for our patients. The adoption of evidence-based practice in surgical specialities has historically lagged behind that seen in medical specialities, with patient management less likely to be supported by evidence from randomized controlled trials (RCTs) [1]. Trials in surgical specialities face unique challenges relating to recruitment, equipoise, learning curves and ethics [2]. The mix of surgical and medical interventions used in transplant recipients mean that only part of transplant practice has developed around evidence from RCTs.

Randomized controlled trials are the backbone of evidence-based surgical practice. It is therefore important to realize that not all clinical trials are of equal methodological quality, and indeed, clinical trials are not always the optimal way to answer a given clinical question. This review examines the design and reporting of clinical trials in solid organ transplantation, with particular attention to the peculiarities of transplantation that can make conducting quality trials challenging.

Clinical trials in transplantation – the status quo

The number of randomized controlled trials published over the past 5 years in solid organ transplantation is summarized in Table 1. Of the 738 publications, nearly

doi:10.1111/tri.12743

Table 1. Total number of publications from randomized controlled trials (RCTs) published between January 2010 and September 2015.

Organ	Total publications	%
Kidney	576	64
Liver	181	20
Heart	75	8
Lung	50	6
SPK	4	1
Pancreas	8	1
Total	894*	100

*Total number of RCTs is 738; some publications cover more than one organ type. Data from the Transplant Library database (www.transplantlibrary.com).

two-thirds recruited renal transplant recipients, with studies focussing on other organ types lagging behind. As a point of reference, 68% of the total number of transplants performed in the United Kingdom in the year 2014/2015 were kidneys, whilst liver transplants made up 18.8% of the total [3]. These figures therefore suggest that trial activity, at least for kidney and liver transplantation, is in proportion to overall transplant activity. In contrast, very few trials have been published in pancreas transplantation (<2% of the total) despite these making up around 5% of all transplants.

These figures highlight the problem with generating evidence for rare interventions. To demonstrate incremental benefits in an already-successful treatment such as transplantation, large numbers of patients are required, putting adequately powered clinical trials out of the reach of individual institutions. Multicentre, and even international, collaboration is therefore essential to deliver robust trials of new interventions, particularly for the less frequently transplanted organs.

Identifying research topics

Systematic literature reviews are designed to identify gaps in the existing evidence and thus help to inform researchers of future research priorities. Such syntheses will help to avoid duplication of existing research but can also help to guide trial design in terms of sample size calculations, selection of appropriate outcomes and treatment regimens. An increasing number of funding bodies now require a formal systematic review of the subject area to be submitted as part of the application process. The use of the Transplant Library of randomized clinical trials (www.transplantlibrary.com) allows a quick examination of the evidence that is available. All members of ESOT have access to this library.

There is also a growing recognition that the research priorities of clinical researchers may be biased, and not reflect the priorities of patients or their carers [4]. To address this, there has been a great deal of interest in ways to involve patients and carers in clinical research and, in particular, in helping to set the research agenda [5]. In the UK, a project is currently ongoing to identify and prioritize areas for research in renal transplantation (www.transplantpsp.org/kidney).

Table 2. The PICO structure for defining a research question.

	Description	Example
Population	Patient characteristics to be included/ excluded from the study. May include age, underlying condition, comorbidity.	*De novo* adult renal transplant recipients
Intervention	The intervention, exposure or test to be studied. May include dose, procedural technique or duration.	Withdrawal of corticosteroids at day 7 post-transplant.
Comparator	The control comparator, exposure or test to be studied. May be no intervention, placebo or current standard therapy.	Continuation of maintenance steroids at 5 mg/day.
Outcomes	Outcomes of clinical importance, including efficacy, survival and safety.	Acute rejection, graft function, graft and patient survival, steroid-related side effects (serum lipids, hypertension, new-onset diabetes, leucopenia, infection, cataracts)

The importance of a good question

Defining a clear, concise research question is essential when planning a clinical trial. For randomized and other comparative trial designs, a commonly used tool for defining the research question is the 'PICO' structure: population, intervention, comparator and outcomes (Table 2). The specific question has implications

for inclusion and exclusion criteria, the exact nature of the study intervention and comparisons, and the outcomes recorded, all of which can affect the generalizability and applicability of the results of the trial when completed.

Population

The trial population must be defined clearly, and the patients recruited into the trial must be representative of the population that we see in everyday practice. Here, a distinction is often made between 'efficacy' trials and 'effectiveness' trials. Efficacy trials recruit participants under ideal, highly controlled circumstances with the aim of increasing internal validity and minimize confounding variables. Effectiveness trials recruit participants under pragmatic, real-world conditions to test the impact of an intervention in day-to-day clinical practice and allow a more heterogeneous trial population. This difference can be illustrated by looking at two trials comparing once-daily versus twice-daily tacrolimus in renal transplant recipients. The OSAKA trial aimed to include patients representative of the European transplant population and included a large proportion of deceased donors and extended criteria donor (ECD) transplants [6]. An earlier European trial used more stringent eligibility criteria, with exclusion of donation after circulatory death (DCD) recipients and sensitized patients, and as a result, the generalizability of results is much more limited [7]. In reality, the distinction is not dichotomous, with efficacy and effectiveness at two ends of a continuous spectrum, ideally presenting a balance between acceptable internal validity and high generalizability.

Many transplant trials, even effectiveness trials, limit the inclusion to certain patient populations such as those with high immunological risk (usually defined by panel reactive antibodies or previous graft loss), the elderly, or recipients of high-risk organs such as DCD, long cold ischaemia times or elderly donors. Whilst these decisions are often guided by ethics committees and safety concerns, they may result in trial populations that are not reflective of our increasingly high-risk recipient and more marginal donor populations. In an analysis of 573 randomized controlled trials in renal transplantation, Blosser and colleagues found that 30% of trials had an exclusion criteria based on age [8]. The mean age of the recipients enrolled in the trials was significantly lower than the prevailing US renal transplant population, questioning the external validity of existing trials in older populations.

Intervention

Careful consideration must be given as to the exact definition of the intervention to be used in the trial. Differences in procedural technique, drug dose and timing of the intervention can all have important implications for the treatment effect recorded. Restricting aspects of normal care, such as baseline immunosuppression, can also reduce the external validity of the results. Where the optimum dose or timing has not yet been determined, randomization to multiple intervention groups may be of use [9]. Particular attention must be paid to the detailed description of complex interventions (those containing multiple interacting components) to ensure reproducibility. In the transplant setting, examples include behavioural interventions aimed to increase donation rates [10] or to improve compliance [11].

Comparator

The comparator should be clinically meaningful and is determined by the research question. In transplantation, the comparator is often the prevailing standard of care, which may vary between centres or countries. Whilst, in an efficacy trial, the comparator may be very rigidly defined, more pragmatic effectiveness trials may allow for greater variation [12]. For a large trial with medium- to long-term follow-up (3–5 years or more), it is conceivable that standard of care will change during the duration of the trial, making the results at best difficult to interpret, and at worst irrelevant by the time, the trial results are published. A good example is the BENEFIT trial – at trial inception, standard of care was cyclosporine-based immunosuppression, but by the time of trial completion, almost all centres had switched to tacrolimus-based regimens [9]. This leaves us wondering whether the excess early rejection rates seen in the belatacept arms would be even greater when compared to a tacrolimus-based regimen.

Outcomes

Selecting an appropriate primary endpoint for a clinical trial can often be difficult. There is a conflict between what we actually want to measure, and the practicalities and costs of doing so. For example, in trials of immunosuppressive drug therapy, one could argue that one of the most important outcomes for the patient is long-term graft survival. With current 5-year kidney graft survival at around 85%, demonstrating an (perhaps unrealistic) improvement in graft survival to 90%

with 80% certainty would require recruitment of over 1300 patients (chi-squared test, alpha = 0.05). The rarer the event of interest, and the smaller the anticipated treatment effect, the larger the number of patients required becomes. Thus, the majority of trials focus on short-term outcomes and this leaves trials with a limited ability to detect differences in survival outcomes and rare adverse events of interest, such as post-transplant lymphoproliferative disease, infections and other malignancies.

It is also important that the outcomes reported in clinical trials reflect those that are important to patients, not just those treating them, as there is some evidence that priorities may differ. In a nominal group study of renal transplant recipients, only 12% of participants ranked their own survival as more important than transplant survival, in stark contrast to the traditional clinician-led view that the survival of the patient is most important [13]. Outcomes reported directly by patients (patient reported outcomes; PROMs) are increasingly recognized as important and are essential if a health economic analysis of a new intervention is planned. These usually consist of questionnaires or surveys that record patient's activities, symptoms and quality of life. PROMs can be generic, working across different conditions, or disease specific. Both generic and disease-specific tools have been evaluated for use in transplant recipients and are likely to be complementary [14].

Developing alternative endpoints

With one-year graft survival rates reaching 90–95% and the rate of biopsy proven acute rejection (BPAR) decreasing to below 10%, the established study endpoints are challenging as a framework for development of new interventions. Not only are numbers required to establish differences in the outcome high and the burden of cost for any trial significant, but early cellular BPAR alone as an established endpoint – if reversible – may also be less significant in the context of modern salvage options and the characteristics of rejection [15]. Whilst graft loss during the first year has been reduced to a minimum in nonsensitized patients, the long-term outcome and deterioration of graft function remain unsolved challenges. Hence, the development of early alternative endpoints has gained attention in order for the field to develop a framework for development of novel treatment strategies and/or drugs to address the *unmet need* for improving long-term graft survival in solid organ transplantation.

Transplant clinical trials often resort to the use of surrogate endpoints [16,17]. Surrogate endpoints are those that are not necessarily of direct importance to the patient, but have been shown to predict long-term outcomes such as survival. Most research has been in the field of renal transplantation, with little literature on outcomes in other organ types [18]. In kidney transplantation these include (*de novo*) donor-specific antibodies and/or complement binding subgroups, graft histological features such as inflammation at an early stage after transplantation, biomarkers and gene or protein expression profiles indicative for chronic damage to the grafts, adherence with immunosuppression, glomerular filtration rate (GFR) and/or composite endpoints including a number of the candidates mentioned above and balanced in relevance to each other following mathematical modelling. A multidimensional approach seems to have the advantage of incorporating measures of the damage caused by different agents, but holds the limitation of driving a cause-effect relationship. Markers of chronic graft injury can predict graft outcomes in renal transplantation [19], but biopsy is an invasive procedure with associated cost and risk of complications meaning that there is a reluctance for protocol biopsies in clinical trials. In the liver, scores of early function have been developed and can predict survival [20,21].

An initiative commenced by major transplant medical associations and facilitated by the Food and Drug Administration (FDA) in the USA is currently being developed further and carries the hope of driving the steps required to establish surrogate endpoints and developing them to the point of robustness to serve as endpoints for future trials [22].

Study design

Whilst the randomized controlled trial is arguably the current gold standard for assessing new interventions, there are variations in quality and methodology. An analysis of the reporting methodology of RCTs included in the Transplant Library (www.transplantlibrary.com) demonstrates that just over half of the trials published between 2010 and 2015 are considered good quality (Fig. 1). Around half reported an adequate method of allocation concealment, and around 60% report some form of intention-to-treat analysis. Whilst these figures appear to be relatively static over the past 5 years, this is a considerable improvement over a previous analysis of trials reported between 2004 and 2006, in which only around one-third were considered good quality and a

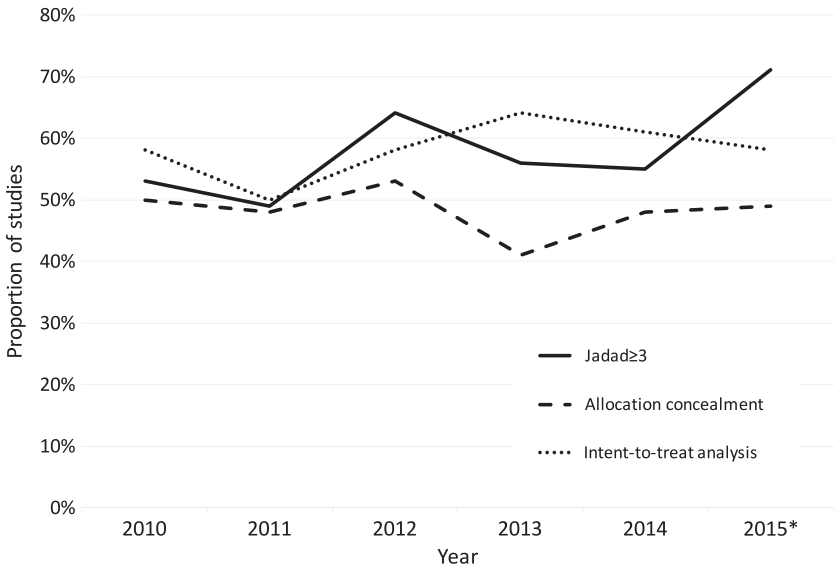

Figure 1 Trends in trials quality in transplantation 2010–2015. Randomized controlled trials with a Jadad score of at least 3 of 5 are considered good-quality trials. Intention-to-treat analysis includes strict intention-to-treat analysis, available case analysis and modified intention-to-treat analysis. *January–September, 2015.

similar proportion reported adequate allocation concealment [23].

Careful study protocol design is the necessary basis to conduct a trial to a high standard and will also aid adequate reporting of the study once complete. Initiatives such as the standard protocol items: Recommendations for interventional trials (SPIRIT) statement assist researchers by providing a checklist of times required in a high-quality study protocol [24].

Superiority versus noninferiority

Most trials of new interventions employ a superiority design to demonstrate that the new intervention is superior to the existing gold standard by a clinically relevant margin. In contrast, equivalence trials are powered to demonstrate that a new treatment is neither worse or better than existing standard of care, whereas a noninferiority study attempts to demonstrate that a new intervention is not worse than the existing standard of care by more than a predefined margin. Equivalence trials are common in pharmacokinetic studies, where the treatment effect can vary in either direction from the reference, for example investigating whether a generic formulation of a drug is (bio) equivalent to the reference drug [25]. For a noninferiority design to be appropriate, there must be a reason other than clinical efficacy where a new treatment would be preferred – usually relating to cost, route of administration or availability. A good example in the field of transplantation is the use of a noninferiority design to compare once-daily versus twice-daily tacrolimus formulations [6].

Understanding the trial design to be used at an early stage is very important, as it has implications for sample size and analysis plans. In contrast to superiority studies, noninferiority studies usually employ a per-protocol analysis (intent-to-treat analysis biases towards finding no difference) and a single-sided p-value (e.g. 0.025) is used.

Randomized controlled trials can still contain bias

The random assignment of interventions in an RCT aims to reduce the risk of systematic differences between groups of participants (bias). On the face of it, selection bias (the preferential assignment of patients with particular characteristics to one study group) should not be possible with adequate randomization. It can, however, still occur when participants are systematically excluded from the trial on the basis of their randomized group. This can occur before randomization if the trial staff are able to predict the allocation for the next participant (inadequate allocation concealment) or after randomization by the means of withdrawal from the trial (attrition bias). The former scenario can be avoided by the used of centralized or Web-based randomization, sequentially numbered containers or sealed opaque envelopes. Whilst attrition bias can be difficult to avoid, it is important that all randomized patients are followed and a full

description of trial withdrawals and exclusions from the analysis are presented in a flow diagram.

Even in those patients remaining in the trial, bias can be introduced if there are systematic differences in the way that patients are treated or outcomes are measured between groups. This may include differences in care, but also more frequent and intensive monitoring resulting in over-reporting of adverse events in one group compared with another (detection bias). The most robust method for preventing this is adequate double-blinding of participants and investigators. Whilst this is relatively straightforward in drug trials with the use of a placebo, it can be prohibitively expensive in larger trials and difficult if the route of administration of two treatments differs. If it is not possible to fully double-blind study, it is often possible to partially blind either participants (for example in organ preservation studies) or outcome assessors (radiologists, pathologists) to attempt to reduce the risk of bias. Despite these limitations, there are some excellent examples of transplant trials in which blinding has been maintained for up to 5 years post-transplant, such as the Astellas steroid withdrawal study [26].

Blinding can be even more problematic in trials of surgical interventions. Blinding the operating surgeon is impossible. Whilst the use of sham surgery to blind the patients and/or outcome assessors is possible, there are considerable ethical concerns with recruiting patients to control interventions with an associated procedural risk [27]. Even if ethical approval is granted, recruitment to such trials often fails because of reservations from participants and investigators.

Randomization does not always result in the desired effect of equivalent baseline characteristics between groups. Particularly in small trials, chance variation can lead to a difference in characteristics between groups making interpretation of results difficult. Whilst it is possible to correct for imbalances in analysis, this is far from desirable. The risk of chance variation can be minimized by the use of stratified block randomization. Stratification ensures that subgroups of participants in whom the treatment effect may differ are equally allocated between treatment arms. Commonly stratified subgroups in clinical trials include transplant centre (as differing protocols and standard care between centres), immunological risk and donor type (DCD, DBD, living donor). Stratification must be taken into account in the final analysis of the trial data.

Stratification is usually achieved by the use of blocked randomization – each block has an equal number of treatments and controls to ensure that the groups are balanced at all stages of the trial. If the block sizes are known to the investigators, it may become possible to predict the allocation of the next patient within a centre at the end of each block. An example of this is seen in a recent trial of conventional versus piggy-back venous drainage in liver transplantation [28]. Patients were stratified by Child–Pugh score and randomized in blocks of size 2. If an investigator knew which group that the last recipient with the same Child–Pugh score was randomized to, he could predict that the next patient would be in the alternative group. This may then effect the decision to include the patient in the trial. Such situations can be avoided by the use of random block sizes to maintain allocation concealment.

Registry data

The field of transplantation is fortunate to be serviced by a number of robust national and international registries. Whilst these have their own drawbacks, including incomplete data and risk of recall bias, the sheer volume of data collected and ability to follow patients for longer periods of time than is feasible in clinical trials makes them an ideal platform to study rare outcomes and long-term effects of treatments. There are, however, some notable examples of opposing results from registry and clinical trial data. Whilst meta-analysis of randomized controlled trials has demonstrated increased risk of acute rejection in patients undergoing steroid withdrawal following transplantation, registry data from the Collaborative Transplant Study demonstrated no increased risk of rejection, with improved survival in patients avoiding maintenance steroids [29,30].

It is also possible to use registry data to complement the design of clinical trials and facilitate long-term follow-up. An excellent example of this is the ongoing 3C study [31]. Whilst the primary outcomes will be reported at 2 years, participants will be flagged in the national transplant registry to enable longer term follow-up of basic outcomes such as graft function and survival. Other examples utilizing UNOS and ANZDATA registry data have also been reported [32,33].

When randomization is challenging or impossible

Particular aspects of transplantation can make randomization challenging. Deceased donor transplantation often occurs at short notice out of hours, making the potential window to consent participants short. Ethical review panels often find this difficult to comprehend as

it may not allow sufficient time for the recipient to consider the risk and benefits of the trial and make an informed decision. This can be improved by providing patients on the waiting list with information regarding ongoing trials, and even recruiting patients on the waiting list, although this can be resource-intensive.

Not all interventions that we wish to study can be controlled. This is particularly true in transplantation in relation to donor characteristics and interventions. Whilst it may be desirable to investigate the effect of a donor characteristic or intervention on a particular recipient, we have no control over organ offers and allocations. Generating evidence as to which organs we should accept for transplantation, and which organs should be allocated to which recipients, is therefore not possible in the context of a randomized controlled trial. For example, comparing the outcomes of older donor DBD and DCD transplants is not possible in a clinical trial as we have no control over the allocation of these organs. Carefully analysed observational studies, with adequate identification and correction for confounding variables, become essential to generate evidence for practice [34].

Even when randomization is possible, it is not always successful. In nonblinded trials, clinicians may lose equipoise and patients may express a preference for one treatment or another. This can be especially true in studies comparing medical and surgical interventions, or where treatment changes in stable patients are proposed. The result is often either poor recruitment or increased withdrawals following randomization leading to an imbalance between arms. A good example is seen in a recent trial investigating the use of cyclosporine in conjunction with antiviral therapy in liver transplant recipients with recurrent hepatitis C [35]. Only 92 of the planned 355 patients were recruited, at least in part because local investigators were reluctant to switch stable patients from tacrolimus to cyclosporine.

Alternative study designs

Even if recruitment targets are met, patient and physician preferences may result in particular subgroups of eligible patients missing from the trial population. This will affect the external validity of the trial, even when inclusion criteria are broad. To address these issues, a number of variations on the traditional RCT design have been suggested. In the comprehensive cohort study design, all eligible patients that refuse randomization but receive one of the study treatments by preference are included in the trial and followed up in the same way as randomized participants [36]. Analysis of the four resulting groups improves the external validity of the trial. The downside to this approach is that if a large number of patients express a preference for one treatment, then insufficient patients may be randomized.

An alternative strategy, proposed by Zelen, is to seek consent from patients after randomization, once the assigned treatment is known [37]. Those patients receiving standard care need only be consented for data collection. Thus, in situations where equipoise may have been lost but a new intervention is not yet standard of care, recruitment of patients to the control arm becomes easier. There are some ethical concerns, as randomization occurs prior to patient approach, and therefore, patients do not receive information about all available treatment options. Intent-to-treat analysis is essential, and significant crossover can lead to dilution of the treatment effect seen.

A more recently suggested alternative to the pragmatic RCT is the cohort multiple randomized controlled trial design (cmRCT) [38]. This design starts with a large observational cohort in the population of interest, with regular outcome assessments performed. For each randomized controlled trial, eligible patients from within the cohort are identified. A random subset of these patients are selected and offered the trial intervention. Outcomes in these patients are then compared with those eligible patients in the cohort that remained on usual care. This system has the capacity for multiple randomized controlled trials to run within the same cohort and facilitates long-term follow-up of outcomes.

The existence of robust national and international registries makes the use of a cmRCT design a feasible alternative to pragmatic trials in the field of transplantation, as the infrastructure for patient identification and follow-up are already in place. Existing registries concentrate on post-transplant follow-up, meaning that this approach would be most suited to the study of post-transplant interventions in stable patients. Using such a design to study *de novo* recipients would be challenging, but could be made possible by the linking of national transplant waiting lists to national transplant registries for the identification and follow-up of patients respectively.

To date, no published examples of these alternative trial designs exist in the transplant literature.

Dealing with withdrawals

Not all patients entering a clinical trial maintain the study treatment until the end of the trial. This can result from treatment changes because of poor tolerance or

Transplant International 2016; 29: 870–879

adverse events, graft losses, deaths and losses to follow-up. If such events are balanced between the two arms, then this is unlikely to affect the results. A problem arises when one intervention leads to a greater dropout rate than another, resulting in an imbalance between arms. Take, for example, the Spiesser study in which *de novo* renal transplant recipients were randomized to cyclosporine or sirolimus-based immunosuppression. At 8 years, those patients still receiving sirolimus have a significantly better glomerular filtration rate compared with those on cyclosporine (74 vs. 46.9 ml/min) [39]. However, only 52% of the originally randomized patients remained on sirolimus. If patients were analysed in the groups to which they were originally assigned (irrespective of their final treatment), the difference between the arms was much smaller (62.5 vs. 47.8 ml/min). It is therefore important that patients entering superiority trials are analysed in their originally randomized groups (termed intention-to-treat analysis), as this better reflects the real-world use of a new treatment and takes into account differing dropout rates between arms.

Other examples where imbalance can arise between groups include where access to treatment is limited. This can occur when comparing transplantation to another intervention – transplantation has a waiting list, and the effects of remaining on the waiting list when compared to an intervention that is immediately available must be considered. Patients with hepatocellular carcinoma may benefit from liver transplantation compared with surgical resection, but this benefit may be lost if their disease progresses to become unresectable whilst on the transplant waiting list [40].

Whilst a strict intention-to-treat analysis, where all patients are analysed in their randomized group, is the gold standard, difficulties can arise when randomized patients do not receive the randomized intervention and therefore no outcome data are available. This is a particular problem in transplant trials, where recipients are often consented and randomized pretransplant. In a proportion of cases, the transplant may not go ahead because of an unsuitable organ, nonproceeding DCD donor or positive cross-match. It would not seem logical to include these patients in analysis as they have not received the intervention and outcome data will not be available, requiring imputation. In this situation, a modified intention-to-treat analysis may be more appropriate, excluding patients who did not follow the protocol. However, there may be a situation where knowledge of the randomized group leads to withdrawal prior to transplantation, leading to a systematic difference between the groups. Take, for example, a trial of a novel organ perfusion machine versus standard cold storage. If an investigator loses equipoise and decides to only transplant marginal organs that have been allocated to the machine, then a bias is introduced that would be missed if patients not transplanted were excluded from analysis.

Statistical analysis plan

A detailed statistical analysis plan should be included when writing a trial protocol. The plan should specify the primary and secondary outcomes including the methods of analysis and effect measures to compare groups, how missing data will be dealt with, and whether any subgroup or sensitivity analyses will be conducted [24]. If interim analyses are planned, details of these should be provided including a description as to how they will be incorporated into the final analysis. A complete and clear statistical plan allows replication of the analysis and it is important that any changes and deviations from the original plan are explained and justified to avoid reporting bias (see below).

When planning a trial the sample size is a key aspect in terms of statistical power, budget, successful recruitment and recruitment time frame. Larger studies improve precision of the effect estimate and thus more likely to reflect the true effect of a treatment compared with the results of small studies by minimizing random error [41]. Sample size estimation requires both an understanding of the likely primary outcome in the control cohort (event rate for dichotomous outcomes, mean and standard deviation for continuous outcomes) and an estimate of the minimal clinically important treatment effect that the investigator would like to be able to detect. Useful sources of information for these estimates include systematic reviews, registry data and previous/pilot studies. The investigator must also specify the required significance level (usually 5%) and the required power to detect the specified treatment effect (usually 80–90%) [41,42].

Reporting

Regardless of the quality of trial design, accurate and complete reporting of trial outcomes is essential to allow interpretation and adoption into clinical practice. The CONSORT statement, adopted by the majority of medical journals, provides a useful checklist of all components that should be included in a report of a randomized controlled trial [43]. The STROBE statement provides equivalent reporting recommendations of

observational studies [44]. Despite the adoption of the CONSORT statement by the majority of transplant journals, a previous analysis demonstrated that on average, less than half of items on the checklist were reported in randomized controlled trials in organ transplantation [45]. Quality of reporting is directly related to the methodological quality, that is studies of better methodological quality report items more consistently.

Quality of reporting can also be an issue at the outcome level. Missing or incompletely reported outcomes in clinical trials can lead to difficulties in interpretation and lead to biased effect estimates in subsequent meta-analyses. In an analysis of immunosuppression trials in renal transplant recipients, there was significant variation in the reporting of the most common outcomes [46]. About 94% of studies reported a measure of patient survival, and 92%, a measure of graft survival. Serum creatinine was reported in 68%, with estimated glomerular filtration rate (eGFR) reported in 64%. All four outcomes were reported in some form in just 34% studies, with only 16% complete.

Summary

Good-quality trials require careful planning, with adequate consideration given to the population enrolled, intervention employed and comparator selected to ensure external validity. Improving outcomes means that identifying suitable endpoints for adequately powered studies is increasingly challenging. Strategies to facilitate longer term follow-up such as registry linkage are becoming increasingly important.

Whilst the overall quality of trials in transplantation is improving over time, there are still improvements to be made in terms of study design and accuracy and completeness of reporting. Randomized controlled trials remain the gold standard but even these can be subject to bias making rigorous methodology and transparency of reporting essential. There are many situations where parallel randomization is not feasible, and the use of alternative strategies, including newer pragmatic designs with observational components or robust observational registry studies may be more appropriate.

Funding

No external funding was received for the review.

Conflicts of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Howes N, Chagla L, Thorpe M, McCulloch P. Surgical practice is evidence based. *Br J Surg* 1997; **84**: 1220.
2. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002; **324**: 1448.
3. NHS Blood and Transplant. Organ donation and transplantation activity report 2014/15 [Internet]. 2015; [cited 2015 Oct 11] Available from: http://nhsbtmediaservices.blob.core.windows.net/organ-donation-assets/pdfs/activity_report_2014_15.pdf
4. Tallon D, Chard J, Dieppe P. Relation between agendas of the research community and the research consumer. *Lancet* 2000; **355**: 2037.
5. Oliver S, Clarke-Jones L, Rees R, *et al.* Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technol Assess* 2004; **8**: 1.
6. Albano L, Banas B, Klempnauer JL, *et al.* OSAKA trial: a randomized, controlled trial comparing tacrolimus QD and BD in kidney transplantation. *Transplantation* 2013; **96**: 897.
7. Wlodarczyk Z, Squifflet J-P, Ostrowski M, *et al.* Pharmacokinetics for once- versus twice-daily tacrolimus formulations in de novo kidney transplantation: a randomized, open-label trial. *Am J Transplant* 2009; **9**: 2505.
8. Blosser CD, Huverserian A, Bloom RD, *et al.* Age, exclusion criteria, and generalizability of randomized trials enrolling kidney transplant recipients. *Transplantation* 2011; **91**: 858.
9. Vincenti F, Larsen CP, Alberu J, *et al.* Three-year outcomes from BENEFIT, a randomized, active-controlled, parallel-group study in adult kidney transplant recipients. *Am J Transplant* 2012; **12**: 210.
10. Ismail SY, Luchtenburg AE, Timman R, *et al.* Home-based family intervention increases knowledge, communication and living donation rates: a randomized controlled trial. *Am J Transplant* 2014; **14**: 1862.
11. Suhling H, Rademacher J, Zinowsky I, *et al.* Conventional vs. tablet computer-based patient education following lung transplantation–a randomized controlled trial. *PLoS ONE* 2014; **9**: e90828.
12. Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol* 2014; **5**: e45.
13. Howell M, Tong A, Wong G, Craig JC, Howard K. Important outcomes for kidney transplant recipients: a nominal group and qualitative study. *Am J Kidney Dis* 2012; **60**: 186.
14. Cleemput I, Dobbels F. Measuring patient-reported outcomes in solid organ transplant recipients: an overview of instruments developed to date. *Pharmacoeconomics* 2007; **25**: 269.
15. El Ters M, Grande JP, Keddis MT, *et al.* Kidney allograft survival after acute rejection, the value of follow-up biopsies. *Am J Transplant* 2013; **13**: 2334.
16. Hariharan S, Kasiske B, Matas A, Cohen A, Harmon W, Rabb H. Surrogate markers for long-term renal allograft survival. *Am J Transplant* 2004; **4**: 1179.

17. Hariharan S, McBride MA, Cohen EP. Evolution of endpoints for renal transplant outcome. *Am J Transplant* 2003; **3**: 933.
18. Washburn K. Endpoints in trials for clinical liver transplantation. *Curr Opin Organ Transplant* 2008; **13**: 252.
19. Nankivell BJ, Fenton-Lee CA, Kuypers DR, *et al.* Effect of histological damage on long-term kidney transplant outcome. *Transplantation* 2001; **71**: 515.
20. Pareja E, Cortes M, Hervás D, *et al.* A score model for the continuous grading of early allograft dysfunction severity. *Liver Transpl* 2015; **21**: 38.
21. Olthoff KM, Kulik L, Samstein B, *et al.* Validation of a current definition of early allograft dysfunction in liver transplant recipients and analysis of risk factors. *Liver Transpl* 2010; **16**: 943.
22. Food and Drug Administration (FDA) Centre for Drug Evaluation and Research. Surrogate endpoints in clinical trials of kidney transplantation [Internet]. 2015; [cited 2015 Dec 9] Available from: http://www.fda.gov/downloads/Drugs/NewsEvents/UCM470429.pdf
23. Pengel LHM, Barcena L, Morris PJ. The quality of reporting of randomized controlled trials in solid organ transplantation. *Transpl Int* 2009; **22**: 377.
24. Chan A-W, Tetzlaff JM, Altman DG, *et al.* SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013; **158**: 200.
25. Alloway RR, Sadaka B, Trofe-Clark J, Wiland A, Bloom RD. A randomized pharmacokinetic study of generic tacrolimus versus reference tacrolimus in kidney transplant recipients. *Am J Transplant* 2012; **12**: 2825.
26. Woodle ES, First MR, Pirsch J, *et al.* A prospective, randomized, double-blind, placebo-controlled multicenter trial comparing early (7 day) corticosteroid cessation versus long-term, low-dose corticosteroid therapy. *Ann Surg* 2008; **248**: 564.
27. Macklin R. The ethical problems with sham surgery in clinical research. *N Engl J Med* 1999; **341**: 992.
28. Brescia MDG, Massarollo PCB, Imakuma ES, Mies S. Prospective randomized trial comparing hepatic venous outflow and renal function after conventional versus piggyback liver transplantation. *PLoS ONE* 2015; **10**: e0129923.
29. Opelz G, Döhler B, Laux G, Collaborative Transplant Study. Long-term prospective study of steroid withdrawal in kidney and heart transplant recipients. *Am J Transplant* 2005; **5**: 720.
30. Knight SR, Morris PJ. Steroid avoidance or withdrawal after renal transplantation increases the risk of acute rejection but decreases cardiovascular risk. A meta-analysis. *Transplantation* 2010; **89**: 1.
31. Haynes R, Baigent C, Harden P, *et al.* Campath, calcineurin inhibitor reduction and chronic allograft nephropathy (3C) study: background, rationale, and study protocol. *Transplant Res* 2013; **2**: 7.
32. Gallagher M, Jardine M, Perkovic V, *et al.* Cyclosporine withdrawal improves long-term graft survival in renal transplantation. *Transplantation* 2009; **87**: 1877.
33. Lentine KL, Schnitzler MA, Xiao H, Brennan DC. Long-term safety and efficacy of antithymocyte globulin induction: use of integrated national registry data to achieve ten-year follow-up of 10-10 Study participants. *Trials* 2015; **16**: 365.
34. Summers DM, Johnson RJ, Hudson A, Collett D, Watson CJ, Bradley JA. Effect of donor age and cold storage time on outcome in recipients of kidneys donated after circulatory death in the UK: a cohort study. *Lancet* 2013; **381**: 727.
35. Duvoux C, Villamil F, Renner EL, *et al.* Sustained virological response to antiviral therapy in a randomized trial of cyclosporine versus tacrolimus in liver transplant patients with recurrent hepatitis C infection. *Ann Transplant* 2015; **20**: 25.
36. Olschewski M, Scheurlen H. Comprehensive Cohort Study: an alternative to randomized consent design in a breast preservation trial. *Methods Inf Med* 1985; **24**: 131.
37. Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979; **300**: 1242.
38. Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *BMJ* 2010; **340**: c1066.
39. Gatault P, Bertrand D, Büchler M, *et al.* Eight-year results of the Spiesser study, a randomized trial comparing de novo sirolimus and cyclosporine in renal transplantation. *Transpl Int* 2015; **29**: 41.
40. Llovet JM, Fuster J, Bruix J. Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation. *Hepatology (Baltimore, MD)* 1999; **30**: 1434.
41. Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev* 2002; **24**: 39.
42. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009; **338**: b1732.
43. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *J Pharmacol Pharmacother* 2010; **1**: 100.
44. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; **370**: 1453.
45. Liu LQ, Morris PJ, Pengel LHM. Compliance to the CONSORT statement of randomized controlled trials in solid organ transplantation: a 3-year overview. *Transpl Int* 2013; **26**: 300.
46. Masson P, Duthie FA, Ruster LP, *et al.* Consistency and completeness of reported outcomes in randomized trials of primary immunosuppression in kidney transplantation. *Am J Transplantat* 2013; **13**: 2892.