The Randomized Controlled Trial

Vianda S. Stel^a Carmine Zoccali^c Friedo W. Dekker^{a, b} Kitty J. Jager^a

^aEuropean Renal Association – European Dialysis and Transplant Association Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, and ^bDepartment of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; ^cDepartment of Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Institute of Biomedicine and Molecular Immunology, National Research Council, Renal and Transplantation Unit, Ospedali Riuniti, Reggio Calabria, Italy

Key Words

Epidemiology · Patient selection · Randomized controlled trials · Study design

Abstract

Until recently, high-quality randomized controlled trials (RCTs) have rarely been performed in nephrology, because of several difficulties in conducting an RCT. The purpose of this article is to describe several issues related to the design and to the quality of RCTs in their conduct, analysis, interpretation, and reporting of the results. The advantage of an RCT is that, as a result of randomization, selection by prognosis by the clinician is prevented. However, not all RCTs provide a definitive answer on the research question that the investigators try to answer because of potential problems in their design, conduct and analysis. For example, the results of an RCT could be biased if the physicians and patients in the experimental and control groups were not sufficiently 'blind-ed'. Nevertheless, if conducted properly, RCTs remain the gold standard for studying the effects of therapy and other interventions.

Copyright © 2009 S. Karger AG, Basel

Introduction

For a research question on the effect of a therapy, the randomized controlled trial (RCT) is the most appropriate study design [1]. Nevertheless, because of several difficulties in conducting an RCT [2], until recently high-quality RCTs have rarely been performed in nephrology [3, 4]. The purpose of this article is to describe several issues related to the design and quality of RCTs in their conduct, analysis, interpretation, and reporting of the results.

Issues Related to the Design of RCTs

Randomization

To explain the concept of 'randomization', we use the hemodialysis (HEMO) trial of Eknoyan et al. [5] as an example. The HEMO trial was designed to determine whether increasing the dose of dialysis or using a high-flux dialyzer membrane alters survival (or morbidity) among patients undergoing hemodialysis. This RCT comprised a 2×2 factorial design in which patients were randomly assigned to a standard or high dose of dialysis and to a low- or high-flux dialyzer (fig. 1). The investigators tested the effect of flux intervention (or Kt/V intervention) on mor-

Fig. 1. HEMO study randomized interventions: under the 2×2 factorial design of the HEMO study, 1,846 patients were simultaneously randomly assigned to one of two dose interventions (high dose, defined as equilibrated Kt/V target of 1.45, vs. standard dose, defined as equilibrated Kt/V target of 1.05) and to one of two flux interventions (high flux, with mean β_2 -microglobulin clearance ≥ 20 ml/min, vs. low flux, with mean β_2 -microglobulin clearance ≤ 10 ml/min) [reprinted from ref. 26, with permission from Elsevier].

tality, regardless of whether or not the patient received Kt/V intervention (or flux intervention). As a result of randomization, only chance determined to which group patients were allocated and therefore only chance determined the distribution of measured and unmeasured confounders over the experimental and control groups. This prevented selection by prognosis (also known as confounding by indication), i.e. the clinician deciding who would receive which therapy based on their opinion about the patient's prognosis.

Within the HEMO study, randomization was performed with the use of random 'blocks', in order to increase the probability of having equal groups with respect to relevant prognostic factors. The patients were first classified into strata based on characteristics that may influence treatment response: age group, diabetic status and clinical center (fig. 2) [6]. Thereafter, within each stratum, permuted blocks were used to make sure there are the same number of individuals within each block and that these persons were equally distributed across the two treatment conditions. It should be stressed that the groups resulting from any (block) randomization may still not be equal in all prognostic factors (complete equality is even unlikely, given the variety in genotypes among study subjects). After randomization, however, any remaining differences between groups are the result of chance and not of selection by prognosis.

Besides the randomization of people, one could for example randomize on the level of dialysis centers to

compare treatment strategies (clustered randomized controlled trial) [7], or on the level of left and right kidneys of deceased donors to compare methods of preservation on delayed graft failure [8]. Usually computer-generated random numbers or a random number table are used for the randomization procedure. Assigning patients according to date of birth (odd or even years) or hospital record number are not acceptable methods as these assignment methods may have a level of predictability [9]. Also, the use of sealed envelopes may bias allocation, as it can lead to clinicians opening envelopes scheduling patients in such a way that those with particular characteristics would be allocated to the desired treatment [10].

Blinding

Whenever feasible, it is strongly recommended that also after randomization, patients and clinicians do not know who receives the intervention and who does not. One reason for 'blinding' is that the patient and clinician may act differently if they know about treatment allocation. For example, in an unblinded RCT, the clinician may tend to check more often on adverse effects in patients from the experimental group than they would do in the control group, which may influence the number of adverse events discovered in the experimental group. The other reason for blinding is that the outcome variable should be determined independently from the treatment provided in order to prevent what is called information bias, i.e. the prejudice deriving from foreknowledge. For example, if a clinician believes that a new drug provided in the experimental group may increase the risk of myocardial infarction, they may feel this diagnosis is more likely in symptomatic patients in the experimental group than in the control group. In a double-blind study, both the patients and clinicians do not know which patient receives the treatment, whereas in a single-blind study either the patients or the clinicians do not know who receives the treatment. For example in the RCT of DePaul et al. [11], the clinicians but not the patients knew whether the exercise program belonged to the experimental or control group. However, sometimes 'blinding' may not be possible. In the unblinded RCT by Ruggenenti et al. [12], both the patients and the clinicians were aware of the allocation because of the nature of the study, i.e. comparing two different targets of blood pressure control.

It is accepted that usually 'blinding' is desirable, which is certainly true for trials aimed at testing causal research hypotheses ('explanatory' trials). Although there is a continuum rather than a dichotomy between explanatory

Fig. 2. Example of the HEMO study which used block randomization: individuals within a block, i.e. clinical centers (2 of the 15 centers are presented in the figure), age group, and diabetes status, were distributed at random across the treatment conditions.

and pragmatic trials, one could say that in ‘pragmatic’ trials, which are designed to help choosing between options of care and examine whether interventions work under real-life conditions, it can be accepted that participants and clinicians are not blinded, as the expectations of the participants and clinicians reflect normal clinical practice and may influence the size of effect. An example of a pragmatic RCT is the one currently performed by Maione et al. [13] comparing standard clinical practice settings with each other, e.g. the impact of different treatments of renin-angiotensin system inhibition on cardiorenal outcomes. Nevertheless, it is still desirable to blind the assessment of outcome in order to obtain objective outcome variables [13, 14].

Issues Related to the Conduct of RCTs

Sample Size and Recruitment

An RCT must have a sufficient sample size to have adequate statistical power so that if the treatment being studied has a clinically relevant effect, it will not be mistakenly missed by the trial. A higher number of events could be achieved by the use of a composite endpoint (e.g. a composite of fatal and non-fatal first major cardiovascular events) instead of a single endpoint (e.g. cardiovascular death only). Nevertheless, the RCT of Zannad et al.

[15] still lacked statistical power to definitely determine the effect of foscipr on cardiovascular events in patients with end-stage renal disease. This was due to the number of cardiovascular events being lower than anticipated.

To estimate the sample size of a trial, one needs to fix the significance level (e.g. $p < 0.05$), the power level (the ability of the trial to detect an effect, given that the effect actually exists), and the minimal difference in the primary outcome considered as clinically relevant [16]. Knowledge gathered in previous observational studies can be used to estimate the effect size that can be reasonably achieved and tested in the actual trial [17].

Generalizability

Strict inclusion and exclusion criteria can increase the ability to determine effects of the interventions within a trial. However, they can also narrow down the patient population to a (very) selected group, thereby compromising the generalizability of its results. This last limitation applies to almost all ‘explanatory’ trials in which recruitment is most often selective. However, in ‘pragmatic’ trials, all patients who might receive the intervention should be included as these trials reflect routine practice [14]. Therefore, the results of ‘pragmatic’ trials are usually more generalizable than the results of ‘explanatory’ trials.

Fig. 3. Hypothetical example comparing the results of an as-treated analysis with the results from an intention-to-treat analysis. In this RCT, patients would be randomized to the experimental group (treatment) and the control group (no treatment) in order to investigate the effect of the treatment on mortality. In the treatment group, 10 patients were non-compliant because of side effects and died shortly thereafter. In the intention-to-treat analyses, in both the experimental and control group 20 of the 100 patients died

(event rate = 20%). In the as-treated analysis, patients who became non-compliant during the study period were excluded, and 10 of the 90 patients who were compliant to the treatment died (event rate = 11%). However, the rate in the control group would be 20%. In this hypothetical example, if one was interested in the results of the intention-to-treat analysis, the results of the as-treated analyses represent a spurious reduction in the mortality risk for the new treatment compared to the control group.

Issues Related to the Data Analysis of RCTs

Intention-to-Treat and As-Treated Analysis

In RCTs, the primary analysis should be an intention-to-treat analysis whereby all comparisons are made on the basis of the groups to which patients were initially randomized, regardless of any treatment that they subsequently received and of actual compliance. In this case, any differences between groups after randomization can be reasonably ascribed to differences in the intended treatments. An alternative method is the as-treated analysis (also known as per-protocol analysis), in which the patients who no longer receive the actual treatment under investigation, such as non-compliant patients (e.g. non-adherence or withdrawal), and patients who switch treatments are excluded from the analysis or censored in coincidence of the treatment switch. In these analyses, one can determine the efficacy of a treatment (i.e. the effect

of the treatment under ideal conditions). However, in practice, one is often not interested in the results of an as-treated analysis, but only in the results of an intention-to-treat analysis, which takes into account the use of the treatment in clinical practice (fig. 3).

Interim Analysis

Usually an independent safety committee receives the intermediate study results without knowing which data are from the experimental or the control group. They may decide to stop an RCT prematurely when a difference between the groups is achieved before the study end and when this possibility is contemplated in the stopping rules of the study protocol [18]. This was the case in an RCT performed by the GISEN group [19]. At interim analysis after 2 years, the safety committee found that the decline in glomerular filtration rate per month was highly significantly lower in one group (which appeared to be

the group receiving ramipril) than in the other group (placebo group; 0.39 vs. 0.89 ml/min; $p < 0.001$) in patients with baseline urinary protein excretion rates ≥ 3 g/24 h. Therefore, this RCT was terminated prematurely, and the most effective therapy was provided to all patients. An RCT may also be stopped because of an unexpected excess of adverse outcomes in the active arm of the study, e.g. in the study by Besarab et al. [20]. However, the average RCT is too small and has insufficient follow-up to detect adverse events [21].

Interpretation of the Results of RCTs

In an RCT, the observed treatment effect is expected to represent the 'true' difference between the new and comparative treatment. However, this is not always the case. The results of the trial may be biased if the experimental and control groups are not handled equally (besides the difference in treatment), for example if the trial is not sufficiently 'blinded'. Secondly, even after randomization, the groups may not be equal with regard to relevant prognostic factors and the association between exposure and outcome may still be confounded. During subsequent data analysis, it is still possible to reduce confounding by using multivariate adjustment techniques. Furthermore, very strict inclusion and exclusion criteria of an RCT may result in a limited generalizability of its results. Finally, as-treated analyses may lead to biased conclusions [9].

Register and Reporting of RCTs

Each RCT should be registered in the International Standard Randomized Controlled Trial Number Register (<http://isrctn.org>) or in one of the national registers (www.who.int) that meet the requirements of the International Committee of Medical Journal Editors (www.icmje.org). The introduction of this registration of trials may help that the results of all trials will be published, and not mainly those trials with positive results, resulting in 'publication bias' (although within nephrology only few published RCTs have positive results [22]). For a proper interpretation and evaluation of RCTs, results should be adequately described. The Consolidated Standards of Reporting Trials initiative published a checklist of items to be included in the reports on RCTs [14, 23, 24]. In addition, the quality of reports of RCTs with regard to randomization, blinding and withdrawals can be assessed for example with the scale of Jadad [25].

Conclusion

The crucial advantage of an RCT is the prevention of selection by prognosis by the clinician. Not all RCTs provide a definitive answer on the research question being addressed. For example strict inclusion and exclusion criteria may limit the generalizability of the trial results just to a small category of patients. However, if conducted properly, RCTs remain the standard for studying the effects of therapy and other interventions. A further article in this series will describe when observational studies can be a valid alternative for RCTs.

References

- 1 Stel VS, Jager KJ, Zoccali C, Wanner C, Dekker FW: The randomized clinical trial: an unbeatable standard in clinical research? *Kidney Int* 2007;72:539–542.
- 2 Jager KJ, Stel VS, Wanner C, Zoccali C, Dekker FW: The valuable contribution of observational studies to nephrology. *Kidney Int* 2007;72:671–675.
- 3 Strippoli GF, Craig JC, Schena FP: The number, quality, and coverage of randomized controlled trials in nephrology. *J Am Soc Nephrol* 2004;15:411–419.
- 4 Zoccali C, Abramowicz D, Cannata-Andia JB, Cochat P, Covic A, Eckardt KU, Fouque D, Heimbürger O, McLeod A, Lindley E, Locatelli F, Spasovski G, Tattersall J, Van Biesen W, Wanner C, Vanholder R: European best practice quo vadis? From European Best Practice Guidelines (EBPG) to European Renal Best Practice (ERBP). *Nephrol Dial Transplant* 2008;23:2162–2166.
- 5 Eknoyan G, Beck GJ, Cheung AK, Daugirdas JT, Greene T, Kusek JW, Allon M, Bailey J, Delmez JA, Depner TA, Dwyer JT, Levey AS, Levin NW, Milford E, Ornt DB, Rocco MV, Schulman G, Schwab SJ, Teehan BP, Toto R: Effect of dialysis dose and membrane flux in maintenance hemodialysis. *N Engl J Med* 2002;347:2010–2019.
- 6 Greene T, Beck GJ, Gassman JJ, Gotch FA, Kusek JW, Levey AS, Levin NW, Schulman G, Eknoyan G: Design and statistical issues of the hemodialysis (HEMO) study. *Control Clin Trials* 2000;21:502–525.
- 7 Sullivan C, Sayre SS, Leon JB, Machekano R, Love TE, Porter D, Marbury M, Sehgal AR: Effect of food additives on hyperphosphatemia among patients with end-stage renal disease: a randomized controlled trial. *JAMA* 2009;301:629–635.
- 8 Moers C, Smits JM, Maathuis MH, Treckmann J, van Gelder F, Napieralski BP, van Kasterop-Kutz M, van der Heide JJ, Squifflet JP, van Heurn E, Kirste GR, Rahmel A, Leuvenink HG, Paul A, Pirenne J, Ploeg RJ: Machine perfusion or cold storage in deceased-donor kidney transplantation. *N Engl J Med* 2009;360:7–19.
- 9 Wang D, Bakhai A: *Clinical Trials. A Practical Guide to Design, Analysis and Reporting*. London, Remedica, 2006.

- 10 Torgerson DJ, Roberts C: Understanding controlled trials. Randomisation methods: concealment. *BMJ* 1999;319:375–376.
- 11 DePaul V, Moreland J, Eager T, Clase CM: The effectiveness of aerobic and muscle strength training in patients receiving hemodialysis and EPO: a randomized controlled trial. *Am J Kidney Dis* 2002;40:1219–1229.
- 12 Ruggenti P, Perna A, Loriga G, Ganeva M, Ene-Iordache B, Turturro M, Lesti M, Perticucci E, Chakarski IN, Leonardi D, Garini G, Sessa A, Basile C, Alpa M, Scanziani R, Sorba G, Zoccali C, Remuzzi G: Blood-pressure control for renoprotection in patients with non-diabetic chronic renal disease (REIN-2): multicentre, randomised controlled trial. *Lancet* 2005;365:939–946.
- 13 Maione A, Nicolucci A, Craig JC, Tognoni G, Moschetta A, Palasciano G, Pugliese G, Procaccini DA, Gesualdo L, Pellegrini F, Strippoli GF: Protocol of the Long-term Impact of RAS Inhibition on Cardiorenal Outcomes (LIRICO) randomized trial. *J Nephrol* 2007;20:646–655.
- 14 Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D: Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.
- 15 Zannad F, Kessler M, Leher P, Grunfeld JP, Thuilliez C, Leizorovicz A, Lechat P: Prevention of cardiovascular events in end-stage renal disease: results of a randomized trial of foscinopril and implications for future studies. *Kidney Int* 2006;70:1318–1324.
- 16 Altman DG: *Practical Statistics for Medical Research*. New York, Chapman & Hall/CRC, 1991.
- 17 Port FK: Role of observational studies versus clinical trials in ESRD research. *Kidney Int* 2000;57:S3–S6.
- 18 O'Brien PC, Fleming TR: A multiple testing procedure for clinical trials. *Biometrics* 1979;35:549–556.
- 19 Randomised placebo-controlled trial of effect of ramipril on decline in glomerular filtration rate and risk of terminal renal failure in proteinuric, non-diabetic nephropathy. The GISEN Group (Gruppo Italiano di Studi Epidemiologici in Nefrologia). *Lancet* 1997;349:1857–1863.
- 20 Besarab A, Bolton WK, Browne JK, Egrie JC, Nissenson AR, Okamoto DM, Schwab SJ, Goodkin DA: The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. *N Engl J Med* 1998;339:584–590.
- 21 Vandembroucke JP: When are observational studies as credible as randomised trials? *Lancet* 2004;363:1728–1731.
- 22 Covic A, Gusbeth-Tatomir P, Goldsmith D: Negative outcome studies in end-stage renal disease: how dark are the storm clouds? *Nephrol Dial Transplant* 2008;23:56–61.
- 23 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF: Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637–639.
- 24 Moher D, Schulz KF, Altman D: The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987–1991.
- 25 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ: Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.
- 26 Greene T: Randomized and observational studies in nephrology: how strong is the evidence? *Am J Kidney Dis* 2009;53:377–388.