How should a clinician interpret results of randomized controlled trials?

Interpretation of randomized controlled trials

Virginijus Šapoka1, Vytautas Kasiulevičius1, Janina Didžiapetrienė1, 2

1 Faculty of Medicine, Vilnius University, Vilnius, Lithuania
2 Institute of Oncology, Vilnius University, Vilnius, Lithuania

Correspondence to: Vytautas Kasiulevicius, Faculty of Medicine, Vilnius University, M. K. Čiurlionio 21/27, 03101 Vilnius, Lithuania. E-mail: vytautas.kasiulevicius@gmail.com

Randomized controlled trials (RCTs) and systematic reviews are the most reliable methods of determining the effects of treatment. The randomization procedure gives a randomized controlled trial its strength. Random allocation means that all participants have the same chance of being assigned to each of the study groups. The choice of which end point(s) to select is critical to any study design. Intention-to-treat is the preferred approach to the analysis of clinical trials. Sample size calculations and data analyses have an important impact on the planning, interpretation, and conclusions of randomized trials. In this article, we discuss the problematic areas that can affect the outcome of a trial, such as blinding, sample size calculation, randomization; concealment allocation; intention of treating the analysis; selection of end points; selection of traditional versus equivalence testing, early stopped trials, selective publications.

Key words: randomized controlled trials, sample size, outcomes, type of analyses

INTRODUCTION

Randomized controlled trials (RCTs) and systematic reviews are the most reliable methods of determining the effects of treatment. Ideally, trials are designed and conducted both to minimize the bias (i.e. have a high internal validity) and to be relevant to a wide but defined population (i.e. have a high external validity, also termed generalizability). There are problematic areas that can affect the outcome of a trial: blinding; sample size calculation, randomization; concealment allocation; intention to treat the analysis (the analytic method used); selection of end points; selection of traditional versus equivalence testing, early stopped trials, selection of publications. In our review, we address the questions such as what it is that leads the RCT to the highest level of evidence and what the features of the RCT that render it so useful are. In the article, we discuss a number of principles that answer these questions.

RESULTS AND DISCUSSION

Blinding in a clinical trial. The term “blinding” or “masking” refers to withholding information about the assigned interventions from people involved in the trial who may potentially be influenced by this knowledge. Blinding is an important safeguard against bias, particularly when assessing subjective outcomes. Blinding in a clinical trial can be defined as withholding information about treatment allocation from those who could potentially be influenced by this information. Unblinded studies exhibit an increased effect of treatment compared with blinded studies. In the section of methods, the authors should describe in some detail who was blinded, how they were blinded, and the success of blinding. Certainly participants and investigators should be blinded. Less commonly recognized is that data collectors and analysts should be blinded. Participants should be blinded because they may use other effective interventions, may report symptoms differently, or may drop out if they perceive they have received a placebo therapy. Investigators should be blinded because they may prescribe effective co-interventions, influence the follow-up, or patient reporting. Data collectors and
analysts should be blinded because they may exhibit different encouragement during performance testing, exhibit variable recordings of outcomes, or different timing and frequency of outcome measurements. There is no universal agreement on how to assess blinding or even whether it should be assessed. Study authors often ask investigators and participants to guess their treatment allocation and report the results. Some would suggest looking for bias-generating consequences instead of contamination and co-interventions. The measurement bias is defined as an inaccurate measurement due to either the accuracy of the measurement instrument or a bias based upon the expectations of participants and investigators. Blinding will help to limit measurement bias (1, 2).

Randomization. The randomization procedure gives a randomized controlled trial its strength. Random allocation means that all participants have the same chance of being assigned to each of the study groups. The allocation, therefore, is not determined by the investigators, the clinicians, or the study participants. The purpose of random allocation of participants is to assure that the characteristics of the participants are as likely to be similar as possible across groups at the start of the comparison (also called the baseline). If randomization is done properly, it reduces the risk of a serious imbalance in the known and unknown factors that could influence the clinical course of the participants. No other study design allows investigators to balance these factors (1, 2).

Concealment allocation. After the randomization sequence is generated, the list may be given to the investigator responsible for enrolling participants in the study. This is referred to as unconcealed participant allocation. The investigator may steer participants to certain treatment arms based upon prognostic factors either consciously or unconsciously. Concealment allocation can be defined as the process by which the physician is blinded to the randomized sequence which was generated. The person who enrolls participants in the trial should not be the same person who generates the allocation sequence. In RCTs where concealment allocation has not been utilized, there is an overestimation of treatment effect compared to trials which conceal the allocation sequence. The treatment effect may increase by 20 to 30%. The average bias associated with the lack of adequate concealment allocation was less for outcomes that were evaluated objectively (death, ulcer closure) rather than subjectively (pain, patient-reported outcomes) (1, 2). The allocation concealment should not be confused with blinding. Allocation concealment seeks to prevent selection bias, protects the assignment sequence until allocation and can always be successfully implemented. In contrast, blinding seeks to prevent performance and ascertainment bias, protects the sequence after allocation, and cannot always be implemented. Without adequate allocation concealment, however, even random, unpredictable assignment sequences can be subverted.

Discrepancies in sample size calculations. Sample size calculations and data analysis have an important impact on the planning, interpretation, and conclusions of randomized trials. Statistical analysis often involves several subjective decisions about which data to include and which tests to use, producing potentially different results and conclusions depending on the decisions taken. The methods of analysis that are chosen or altered after preliminary examination of the data can introduce bias if a subset of favorable results is then reported in a publication. The study protocol plays a key role in reducing such bias by documenting a pre-specified blueprint for conducting and analyzing a trial. Explicit descriptions of methods before a trial starts help identify and deter unacknowledged, potentially biased changes made after reviewing the study results. To evaluate the completeness and consistency of reporting, we reviewed a comprehensive cohort of randomized trials and compared the sample size calculations and data analysis methods described in the protocols with those reported in the publications (3, 4).

Superiority versus equivalence trials. Most trials test whether a new treatment is superior to control (placebo) group or conventional standard of care. A superiority trial aims to demonstrate the superiority of a new therapy compared to an established therapy or placebo. In contrast, some trials are designed to show that a new treatment is not inferior to standard therapy by a predefined acceptable amount. Several problems challenge the design, conduct, analysis, reporting, and interpretation of noninferiority trials, and recent meta-analyses confirm that the majority of published trials have substantial methodologic flaws (5). As a result, potentially suboptimal treatments might be introduced into routine clinical practice. Other issues that are crucial to ensuring the validity of noninferiority inference, such as ethical considerations, adequate power, the quality of trial conduct, the choice of analytic strategy (intention-to-treat versus per-protocol), and an alternative Bayesian approach to analysis, are beyond the scope of this paper and have been detailed previously (6, 7). In conclusion, if noninferiority trials are to be applied to regulatory and clinical decisions about the marketing and use of new treatments, their assumptions must be made explicit, the criteria on which they are based must be sufficiently justified, and their influence on the resultant conclusions must be assessed rigorously and expressed unambiguously in published reports (8, 9).

Intention-to-treat or on treatment analyses. There are three general analytic approaches in clinical trials: analysis as randomized (referred to as intention-to-treat analysis, or ITT), compliers-only analysis (in which only those patients randomized to a treatment who completed the trial and complied with treatment are analyzed), and as-treated analysis (in which only those who received a given treatment are counted, whether or not the patient was initially assigned to that treatment). Intention-to-treat analysis is a method of analysis for randomized trials in which all patients randomly assigned to one of the treatments are analyzed together, regardless of whether or not they completed or received that treatment. Intention-to-treat analysis prevents a bias caused by the loss of participants, which may disrupt the baseline equivalence
established by random assignment and which may reflect non-adherence to the protocol. Intention-to-treat (ITT) analysis is commonly accepted as more conservative than the per-protocol (PP) restricted to the analysis of data on subjects who completed the study. Commonly, the within-groups differences being smaller in ITT than in PP, their statistical comparison leads to a smaller risk of type I error (i.e. inappropriately concluding a difference while there is not any). It also allows for keeping the randomization scheme (i.e. the balanced distribution of confounding factors) and thus not lead to a differential distribution of confounding factors among the groups if more subjects are withdrawn from the study in a given group (10).

**Surrogate outcomes.** The choice of which end point(s) to select is critical to any study design. Two additional areas require particular attention: the use of surrogate measures and the use of composite end points. The most persuasive trials require particular attention: the use of surrogate measures and select is critical to any study design. Two additional areas are built on tests within subgroups. Even with a significant interaction test, readers should base the interpretation of the findings on biological plausibility, on prespecification of analyses, and on the statistical strength of the information. Generally, adjustments for multiplicity are unnecessary when investigators use interaction tests. However, in view of the frequently frivolous data-dredging pursuits involved, the argument for statistical adjustments is stronger than that for multiple endpoints. Moreover, if investigators do not use interaction tests and report tests on every individual subgroup, multiplicity adjustments are appropriate. Most subgroup findings tend to exaggerate reality. Be especially suspicious of investigators highlighting a subgroup treatment effect in a trial with no overall treatment effect (15–19).

**RCTs stopped early for benefit.** When randomized clinical trials (RCTs) identify larger than expected treatment effects, investigators may conclude, before completing the trial as planned, that one treatment is superior to the other. Such trials often receive considerable attention. Clinicians face challenges when interpreting the results of truncated RCTs. Taking the point estimate of the treatment effect at face value will be misleading if the decision to stop the trial resulted from catching the apparent benefit of treatment at a “random high”. When this occurs, data from future trials will yield a more conservative estimate of treatment effect, the so-called regression to the truth effect. Thus, clinicians must attend not only to the usual methodological safeguards against bias, but also to the characteristics that affect the decision to stop a trial early. Such characteristics include the plausibility of the treatment effect, the planned sample size, the number of interim analyses after which the investigators stopped the RCT, and the statistical methods used to monitor the trial and to adjust estimates, p values, and confidence intervals for interim analyses. While RCTs stopped early for reasons other than benefit might share some characteristics with RCTs stopped early for benefit, their implications are very different. Trials stopped early because of harm or futility tend to result in a decreased use or prompt discontinuation of useless or potentially harmful interventions. In contrast, trials stopped early for benefit may result in a rapid identification, approval and dissemination of promising new treatments (20–23).

**Selective publications.** Another common problem is that the pharmaceutical industry can choose which data to publish and which to leave unavailable. Much has been written on eye-catching stories, such as the difficulties in getting clear information about the number of suicide attempts in industry trials of SSRI antidepressants, or the number of heart attacks in patients on rofecoxib. Equally concerning is the routine grind of publication bias, where disappointing negative results on the benefits of treatments quietly disappear (24). Medical decisions are based on the understanding of publicly reported clinical trials. If the evidence base is biased, then decisions based on this evidence may not be the optimal
decisions. For example, selective publications of clinical trials and the outcomes within those trials, can lead to unrealistic estimates of drug effectiveness and alter the apparent risk–benefit ratio. Attempts to study selective publications are complicated by the unavailability of data from unpublished trials. Researchers have found evidence for selective publication by comparing the results of published trials with information from surveys of authors, registries, institutional review boards, and funding agencies, and even with published methods. Numerous tests are available to detect a selective-reporting bias, but none are known to be capable of detecting or ruling out bias reliably (25–31).

CONCLUSIONS

Although RCTs remain a gold standard proof of efficacy, there are many aspects of trial design that must be appropriately incorporated to ensure the value of a study. An inappropriate use of any tool (including RCTs) compromises the ability to meaningfully interpret the resulting information. We have presented several aspects should be considered by a user of the information when establishing the credence to attach to the information from a RCT.

Received 9 April 2010
Accepted 27 May 2010

References

KAIP KLINICISTAI TURĖTŲ VERTINTI
ATSITIKTINIŲ IMČIŲ KONTROLIUOJAMUS TYRIMUS?

Santrauka

Raktažodžiai: atsitiktinis atrankos tyrimas, imties dydis, rezultatai, analizės tipas