

Why RCTs?

A fundamental question that any researcher may ask is, why is it that evidence based on RCTs is considered to be of the highest quality. The main reason for this is that evidence based on observational data is prone to bias. Bias is defined as the systematic tendency of any factors associated with the design, conduct, analysis, evaluation and interpretation of the results of a study to make the estimate of the effect of a treatment or intervention deviate from its true value. If two or more groups are being compared in an observational study, there are often systematic differences between the groups, so much so that the outcome of the groups may be different because of these differences rather than actual exposure or intervention. This is known as confounding. The only way to eliminate these differences is to allocate each individual to one or the other intervention at random. Therefore, the probability of any individual receiving one intervention or the other is decided solely by chance. In this scenario, if sample size is sufficient, all the factors influential in the outcome are likely to be distributed equally between the groups, because the allocation was at random. Therefore, any difference in the observed outcome between the groups is likely to be due to the intervention rather than any other factors. Although randomization is the best way to minimize the risk of confounding by unmeasured factors, RCTs could still be confounded. A recent article by Howards provides several suggestions on how to address confounding [5](#).

What questions are suitable for an RCT?

An RCT is a study design that is generally used in experiments testing the effectiveness and/or safety of one or more interventions. The intervention being tested is allocated to two or more study groups that are followed prospectively, outcomes of interest are recorded, and comparisons are made between intervention and control groups. The control group may receive no intervention, a standard treatment, or a placebo. The intervention can be therapeutic or preventive and does not necessarily have to be a pharmaceutical agent or a surgical intervention.

Randomized controlled trials are suitable both for preclinical and clinical research. There should be sufficient uncertainty about the utility of an intervention. This is referred to as equipoise. For clinical trials, the proposed intervention is sometimes based on logic, but mostly on data obtained from *in vitro* laboratory studies, animal experiments or preliminary serendipitous/planned observation in an uncontrolled setting. Observational (case–control or cohort) studies may suggest the benefit of an intervention, but they are prone to bias. Important and relevant gaps in the scientific knowledge sometimes come to light in the process of developing guidelines, and such gaps need to be addressed by producing robust evidence. The question to be answered by RCT design should also be safe from participants' point of view. Investigators cannot subject participants to any undue risk/harm. For clinical trials, it can be argued that it is unethical to expose participants to any risk or burden if the research lacks social value [6](#). There are conditions/situations or statuses for which there is unlikely to ever be an RCT; for example, the effectiveness of a parachute in preventing death while sky-diving. One needs to be practical in understanding such dilemmas and not be blind-sided by proponents of RCTs [7](#). Finally, the research question also needs to be ethically appropriate to be answered in an RCT setting. Experiments conducted on human subjects before or during World War II provide striking reminders of the importance of obtaining voluntary informed consent from participants and ensuring their well-being in the design of RCTs [8](#). What is considered ethical may change with time and accumulation of knowledge. As an example, clinical trials of intravenous administration of alcohol to pregnant women as a tocolytic agent to halt preterm labor were performed from the 1960s to the late 1980s [9](#), [10](#). However, administering alcohol to pregnant women is no longer ethical in light of its harmful effects on the mother and developing fetus.

Randomized controlled trials are equally useful to study effectiveness and/or safety of diagnostic and screening tests. However, RCTs are not suitable for investigating etiology or natural history of disease. Outcomes that are extremely rare or take a very long time to develop are also impractical to study using an RCT.

How should an RCT be designed?

Detailed guidance on planning of clinical trials is available [11](#), and should be followed. Here, we will only consider the basic principles. The first step is to assess if an RCT is the best research design for the research question. Then, it is necessary to confirm that the question has not already been answered by an appropriately powered (please see later) RCT. This requires a systematic literature search.

Research question formulation

A hypothesis has to be formed for a research question to be answered using an RCT. The hypothesis has to be precise. The key components of a sound research question should include: P (population of interest), I (Intervention to be studied), C (comparator intervention), O (outcomes to be evaluated) and T (is there a time duration for intervention/outcome ascertainment time). Adequate time needs to be devoted to converting a “free form” question arising from a clinical or nonclinical context to convert it into a properly answerable “PICOT” format question. This is best demonstrated using an example of tranexamic acid that is shown to reduce blood loss during cesarean section [12](#), [13](#). To address the free flowing research question “Does administration of tranexamic acid prevent postpartum hemorrhage?”, investigators need to convert this into an answerable and specific question. The population needs to be defined clearly, such as whether one would investigate women delivering vaginally, by cesarean section (which could also be emergency or elective), or both. Next, the intervention requires specific stipulation as to how much, how frequently, what dose, what time and what route would be used for administering tranexamic acid. For a comparator, it needs to be clearly outlined whether this intervention is compared with placebo or no treatment or any other currently used measures to prevent postpartum hemorrhage. The outcome also needs to be clearly defined with regard to what definition of postpartum hemorrhage will be used for the trial (for example, blood loss >500 mL or >1000 mL). As the outcome of interest here occurs within a fairly short time frame of completion of intervention, the time component may not be very relevant to this question. Hence, a refined form of the question may

look like this: “Does 1.0 g of intravenous tranexamic acid given 4 h prior to elective cesarean section compared with placebo reduce postpartum hemorrhage diagnosed as estimated blood loss of >500 mL within 24 h of birth?”

As evident from the above exercise, an ample amount of time must be devoted to carefully honing the research question. This should result in a very well-defined, clear, feasible, specific, measurable, ethical and clinically important question before one initiates the experiment.

How should an RCT be conducted?

Once a research question is generated, the next step in the conduct of an RCT will be to clearly define the target population, inclusion and exclusion criteria, process of randomization, allocation, blinding of intervention, treatment and control delivery, outcomes assessment, definitions of outcomes, sample size required, ethical requirements, consent process and, finally, data management. These topics should be written as a well-defined protocol. The protocol should be reviewed and approved by an independent ethics committee before starting the trial. For clinical trials, it is now obligatory that protocols are registered with a publicly available trial registry before recruiting any participants to the trial. Apart from promoting transparency, most journals have this as a mandatory requirement and would not publish the results of an RCT unless trial registration details were provided. Retrospective registration may not be acceptable. Some medical journals also accept RCT protocols for publication. This is to make sure that protocols are not altered as a result of trial results. It also helps to ensure that negative trials are not left unpublished. Nevertheless, only two-thirds of RCTs in women's health published in 2015 were prospectively registered, and more than half did not achieve the planned sample size [14](#). However, this is expected to change in future. Moreover, the International Committee of Medical Journal Editors (ICMJE) has recently suggested that clinical trials that will begin enrolling participants on or after 1 January 2019 must include a data-sharing plan when registering the trial [15](#).

An independent data monitoring committee is usually established to periodically oversee the progress of a clinical trial, safety data and

critical efficacy variables, and to recommend to the sponsor (funder) whether to continue, modify or terminate a trial.

The key components of design of an RCT are highlighted below.

Random allocation

Each of the eligible participants should have an equal chance to be allocated the intervention or not. The simplest way of achieving this is by parallel group design, in which each group of participants is exposed to only one of the study interventions. In a crossover design, all the trial participants receive both interventions in a sequential manner and only the order of intervention is randomly assigned. In this way, each participant serves as his/her own control, thereby eliminating individual participant differences. However, this design is more vulnerable to drop out and attrition. If a particular baseline characteristic is of such fundamental importance as to have a big influence on the outcome, it can be taken into account at randomization. Participants with or without that baseline characteristic are randomized separately (stratified randomization). Block randomization is used to maintain a balance between the intervention group and control, so that the numbers are not too dissimilar, which could rarely happen by chance. Cluster randomization can be used when randomization of individual participants is not feasible/practical, in which case hospitals, clinics, geographic areas etc. can be used as units for the allocation of intervention or control groups. Generation of random sequence should be done by some independent personnel, usually a statistician, who is not going to be involved in the conduct of the RCT. The access to this sequence should be restricted to only a few individuals who absolutely need to have access (such as the pharmacist who will be preparing the medication) and not the investigators or personnel involved in ascertaining outcome. The sequence should be opened by this individual only on a case-by-case basis and specified sequence should be followed.

Allocation concealment

One of the key components of an RCT is allocation concealment. This means that neither front-line care providers, investigators or participants are aware of whether the next eligible participant will be receiving treatment or control intervention. This should be masked

until the time when participants are ready to receive intervention. By this virtue, unnecessary adjustments as to whether to enroll a participant or not (such as after knowing that the prognosis is not good and the patient is randomized to an experimental treatment, the investigator changes her/his mind and decides not to include the participant in the study) can be avoided. This is very important in situations when blinding of intervention is not possible.

Blinding

The focus of conducting an RCT is elimination of bias. Unconscious information bias may be introduced if the investigators or participants are aware of who is getting the intervention and who is not. The procedure of blinding the participants (single blind) or both investigators and participants (double blind) helps to eliminate this unconscious information bias. Whenever possible, blinding should be used in an RCT. It is not always possible to blind either the participants or investigators due to the nature of the RCT. For example, Jozwiak et al. [16](#) conducted an RCT comparing vaginal prostaglandins with trans-cervical balloon catheter for induction of labor (PROBAAT study). It was possible to randomize participants to the two types of interventions, but it was not possible to blind the participant or the investigator. Therefore, this trial was conducted as an open-label RCT.

Conduct

An RCT can be conducted at a single site or at multiple sites. RCTs conducted according to a single protocol but at more than one site are referred to as a multi-center trials. Including several sites has the advantage of reaching the required sample size within a shorter time and may also improve generalizability of findings.

The main premise of conducting an RCT is that the participants should be treated exactly the same way in both arms except for the intervention/control treatment. All other procedures of treatment, diagnosis, investigations, alterations etc. should follow the routine process and no undue advantage or testing should be performed on patients in the trial. These data should be collected to identify issues of contaminations, crossover of intervention and co-interventions.

Outcome ascertainment

The prespecified primary and secondary outcomes should be collected by independent observers who are unaware of the allocation and treatment arms of participants. As far as possible, it is advisable that objective measures are used for ascertaining outcome so that personal bias on the part of the collector does not come into play. This is particularly important when the intervention cannot be masked (such as surgical scars). It is also important that the outcome is collected in all randomized patients. The number of patients with missing outcome data should be minimized as far as possible. A high rate of attrition will lead to reduced confidence in the results and may lead to biased estimates.

Sample size

One would always like to conduct a study that has adequate sample size and power so that the conclusions generated from the experiment can be applied to the broader population with ample confidence. The required sample size to test a hypothesis is governed by the effect size [17](#). In a superiority trial, one would like to detect difference in the effect of intervention vs. placebo, which is a minimum clinically important difference, that is required to detect between two groups and convince users of the information to utilize the intervention. This number is usually derived from previous experiments/observations, previous trials or by consensus opinion. In general, the more widely the two groups are separated from each other and the smaller the variability in each group, the fewer participants are necessary in each group to show that the difference is unlikely to be due to chance and more likely to be due to the intervention. The inclusion of too few patients in a study increases the risk that a significant treatment benefit will not be shown, even if such an effect exists (a type 2 error). The details of actual calculations are beyond the scope of this manuscript but published literature on this topic is abundant and standard formulae are available to calculate sample size [18](#), [19](#). Expert statistical help should be sought when needed. However, some key information should be gathered before seeking statistical help. First, one needs to know the baseline estimate of outcome rate in the placebo/control arm, i.e. how many patients are expected to benefit from the control intervention. Second, it is important to have an expected estimate as to what percentage of patients are expected to benefit

from the intervention. This is where one needs to be careful to not overestimate the benefit (as this will need lower sample size) or to underestimate the benefit (as one may end up experimenting on more patients than necessary). Two more aspects that must be kept in mind are: how many type 1 and type 2 errors are we willing to accept before rejecting the null hypothesis (as described below).

Power of a study

If there are significant differences in the primary outcome between the two groups, one can conclude that the difference is likely to be due to the intervention. Typically, the difference is thought to be 'significant' if the probability of this difference arising solely due to chance is <0.05 . This is the well-known probability (p)-value. Therefore, the chance that a difference will be found even if there is no real difference is 1:20 (0.05). This is known as a type 1 error, and is usually fixed at 0.05 due to convention. However, it is not necessary to fix this at 0.05, and other levels of significance (0.01 or 0.005) may be chosen. It is possible that a significant difference may not be observed even if this is present. The chance that the study will be able to demonstrate a significant difference if it is present, is known as the power of the study. By convention it is fixed at 0.8 to 0.95 level (80–95%). Inability to demonstrate a significant difference even when one does exist is known as a type 2 error. The probability of type 2 error is conventionally set at 0.2 to 0.05. With these four values, it will be easy for a statistician to calculate sample size in an experiment where effect size is measured in proportion. If the effect size is measured as a continuous variable (such as difference in blood pressure) then one would need the mean and standard deviation of the variable in each group in addition to type 1 and 2 error values to calculate the sample size. This provides adequate information for most routine types of RCTs. For different and complicated statistical parameters to be evaluated in an RCT, it is advisable to consult an experienced statistician before designing the experiment, as prohibitive sample size may jeopardize research efforts. RCTs can be designed and conducted to evaluate the superiority, equivalence or noninferiority of an intervention compared with another, and power calculations are different for these different types of RCTs. Power calculations

may be based on simulations performed by a statistician, the details of which are beyond the scope of this review.

Trial phases

Clinical trials evaluating safety and efficacy of pharmaceutical agents generally have to go through a series of studies before they can be used in clinical practice. There are four sequential phases of clinical trial that have the objective of: studying human pharmacology of the agent (phase I); exploring therapeutic potential (phase II); confirming therapeutic effect (phase III); and evaluating it for therapeutic use (phase IV). Phase I trials are conducted in a small number of healthy participants (20–80) to determine the absorption, distribution, metabolism and toxicity of a new drug in humans for the first time. Phase II trials are designed to estimate dose and test the safety and therapeutic efficacy in a slightly larger population (100–300) afflicted with the condition for which the drug was developed. Phase III is a definitive study of efficacy of the drug after sample size estimation for proper evaluation. Data on side effects are collected meticulously. Phase IV trials are post marketing studies after a drug has been approved by a regulatory body such as the US Food and Drug Administration in the USA or the European Medicines Agency in Europe. Such trials provide additional information including the benefits, optimal dose, effectiveness and adverse events of the drug in different patient populations.

Improving generalizability

Generalizability should be considered as a very important criterion in designing an RCT. The answer provided by an RCT is applicable to the patient population similar to that used in the trial. Extrapolating it to other patients is not strictly valid. Therefore, the inclusion criteria should be as wide as is practically possible, while at the same time maintaining scientific rigor. None of the aforementioned steps should be so strict that once the RCT is over, replication of the intervention is impossible in a practical setting. The process of administering intervention to a group of subjects should be relatively easy, and collection of outcome data should also not be too onerous. For example, the investigators of WOMAN trial [13](#) used a clinical definition of postpartum hemorrhage, and

supplemented it with a quantitative estimation of blood loss: >500 mL for vaginal births, and >1000 mL for operative deliveries. This implied that the trial result was valid both for spontaneous and operative delivery.

Interim analysis

Randomized controlled trials are designed with an anticipated incidence of the primary outcome in the control arm. The observed incidence may be lower, making the trial underpowered, or higher, making the trial unnecessarily prolonged. Interim analysis is a useful way to make sure that the observed incidence is not too different from the expected incidence. However, interim analyses should be preplanned and stated in the protocol. Analysis should be performed by an independent statistician blinded to the identity of either group. Interim analysis may sometimes show that differences in the two groups are large and show a clear advantage of the intervention. In this case, continuing the trial is unethical because the control group will be denied the clearly superior alternative. On the other hand, early discontinuation may be advisable if the incidence of primary outcome in the control is far too low and the revised sample size is deemed to be unfeasible. Upward revision of the required sample size may result from the interim analysis as was the case in the WOMAN trial [13](#). When event rates are lower than anticipated or variability is larger than expected, methods for sample size re-estimation are available without un-blinding. The observed incidence in the placebo arm of the study is often lower than anticipated. In the WOMAN trial, the anticipated rate of death was 3.0% in the placebo arm, but the observed incidence was 1.9%. An independent data monitoring committee usually oversees the interim analysis.

Ethical considerations

Rigorous ethical principles must be applied to all RCTs involving experiments on animals, humans or human biological material [20](#). Evaluation of risk and benefit to the participants and society, obtaining ethical approval, and informed consent are crucial. Before planning and conducting an RCT, it must be considered and evaluated whether it is ethical to use randomization to allocate participants to an intervention group. Where there is previous

evidence showing superiority of an intervention over that of doing nothing, an RCT using a placebo (or doing nothing) is unethical. For example, before the development of anti-retroviral therapy, untreated human immunodeficiency virus infection was associated with near certain death. Today it is a treatable condition although newer, more effective treatments continue to be invented. Therefore, RCTs in this field comparing one drug or treatment regimen against another would be ethical, but the use of placebo would not.

Safety concerns

As is true of medicine, participants must not be harmed by the experiment. Therefore, predefined, serious, adverse events need to be reported to the sponsor and to the trial monitoring committee. Any complications or side effects of drugs should be reported to regulatory bodies. Unacceptably high frequency of adverse events in the intervention group may lead to early discontinuation of the trial, usually recommended by the data monitoring committee.

How should an RCT be reported?

There are guidelines for reporting on RCTs (<http://www.consort-statement.org>), which should be followed. Many RCTs report baseline characteristics of the two (intervention and control) groups. If the allocation was random, any differences between the baseline characteristics of the two groups must be by chance. Therefore, a comparison with statistical testing and reporting p -values is superfluous.

The type of comparison made between intervention and control groups is important, especially when efficacy of an intervention is being evaluated. Therefore, whether it is a superiority, noninferiority or equivalence trial should be reported.

It is not uncommon that some participants do not receive the intervention allocated by the randomization process. The gold standard of reporting is 'intention-to-treat' analysis. Outcomes of all participants randomized to the intervention arm should be reported in that group even if some of the participants may not have received the intervention. The analysis by intention-to-treat, and according to the intervention that they actually received (per-protocol analysis) can rarely lead to differing results. Reporting per-protocol analysis

rather than intention-to-treat analysis often results in overestimation of the effect of intervention. Grouping participants according to the actual treatment received can introduce a bias, and is discouraged. The best strategy to guard against such a possibility is to maintain protocol violations to a minimum. While reporting the primary outcome, it is becoming increasingly customary to report the effect size (and its 95% confidence interval) rather than just the p -value, as this provides meaningful information about the magnitude of change. It is best to always use two-sided p values [21](#). It is also a good practice to give the actual p -values rather than $p < 0.05$ or “significant,” and $p > 0.05$ or “not significant” [21](#).

How should RCTs be interpreted?

An RCT is an experiment. If the difference in the primary outcome is significant at the customary level of $p < 0.05$, chances are that the observed difference is real. The magnitude of the observed difference is also important. The magnitude may be small, but still statistically significant. In this case, the clinical significance is most often limited. The magnitude may be large, but still statistically nonsignificant. In this case, the study remains underpowered. Appropriate sample size calculations before embarking on the study should prevent this situation. In many cases, the observed difference is small, and is not statistically significant. This constitutes a negative trial. It does not mean that the trial has failed. When a trial is too small to detect modest treatment effects, it is appropriate to describe the findings as inconclusive rather than negative. It is a matter of great joy for the investigators when the anticipated difference is found between the two groups and is statistically significant. As great care has been observed in the conduct of the trial, the results are likely to be reproducible. However, the scientific community was recently shaken by reports that a troubling proportion of peer-reviewed preclinical studies are not reproducible. New initiatives have been proposed to increase confidence in the published studies [22](#). Generalizability should also be taken into account when interpreting results. When applying the evidence gathered from an RCT to a clinical situation, the question to ask is “is my patient so different compared with the participants of

the RCT so as to make the results of the RCT inapplicable?” The evidence is valid as long as the answer to this question is “no.” Interpretation of any trial should depend not only on the primary outcome, but on the totality of the evidence (i.e. the primary, secondary and safety outcomes).

Conclusion

The RCT is the most rigorous and robust research method for determining whether a cause–effect relation exists between an intervention and an outcome. Therefore, it is important to perform RCTs to generate evidence in basic, translational and clinical research and improve the management of our patients. However, an RCT should be conducted only if it is ethically feasible, economically viable and clinically worthwhile.

Funding

No specific funding.

- 1Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med*. 2016; **21**: 125– 7. [CrossrefPubMedGoogle Scholar](#)
- 2Milne I. Who was James Lind, and what exactly did he achieve. *J R Soc Med*. 2012; **105**: 503– 8. [CrossrefPubMedWeb of Science®Google Scholar](#)
- 3 Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *Br Med J*. 1948; **2**: 769– 82. [CrossrefPubMedGoogle Scholar](#)
- 4Hill AB. The clinical trial. *N Engl J Med*. 1952; **247**: 113– 9. [CrossrefCASPubMedWeb of Science®Google Scholar](#)
- 5Howards PP. An overview of confounding part 1: the concept and how to address it. *Acta Obstet Gynecol Scand*. 2018; **97** 394– 399. [Wiley Online LibraryPubMedWeb of Science®Google Scholar](#)
- 6Wendler D, Rid A. In defense of a social value requirement for clinical research. *Bioethics*. 2017; **31**: 77– 86. [Wiley Online LibraryPubMedWeb of Science®Google Scholar](#)
- 7Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ*. 2003; **327**: 1459– 61. [CrossrefPubMedWeb of Science®Google Scholar](#)

- 8Annas GJ. Beyond Nazi war crimes experiments: the voluntary consent requirement of the Nuremberg Code at 70. *Am J Public Health*. 2018; **108**: 42– 6. [CrossrefPubMedGoogle Scholar](#)
- 9Haas DM, Morgan AM, Deans SJ, Schubert FP. Ethanol for preventing preterm birth in threatened preterm labor. *Cochrane Database Syst Rev*. 2015; **11**: CD011445. [Google Scholar](#)
- 10Schröck A, Fidi C, Löw M, Baumgarten K. Low-dose ethanol for inhibition of preterm uterine activity. *Am J Perinatol*. 1989; **6**: 191– 5. [CrossrefCASPubMedWeb of Science®Google Scholar](#)
- 11 ICH. *Guideline for good clinical practice E6(R2)*. EMA/CHMP/ICH/135/1995. London: European Medicines Agency, 2017. Available online at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf (accessed December 11, 2018). [Google Scholar](#)
- 12Simonazzi G, Bisulli M, Saccone G, Moro E, Marshall A, Berghella V. Tranexamic acid for preventing postpartum blood loss after cesarean delivery: a systematic review and meta-analysis of randomized controlled trials. *Acta Obstet Gynecol Scand*. 2016; **95**: 28– 37. [Wiley Online LibraryCASPubMedWeb of Science®Google Scholar](#)
- 13 WOMAN Trail Collaborators. Effect of early tranexamic acid administration on mortality, hysterectomy, and other morbidities in women with post-partum haemorrhage (WOMAN): an international, randomised, double-blind, placebo-controlled trial. *Lancet*. 2017; **389**: 2105– 16. [CrossrefPubMedWeb of Science®Google Scholar](#)
- 14Nijjar SK, Khan KS. Threats to reliability risk erroneous conclusions: a survey of prospective registration and sample sizes of randomised trials in women's health. *BJOG*. 2017; **124**: 1057– 61. [Wiley Online LibraryCASPubMedWeb of Science®Google Scholar](#)
- 15Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *JAMA*. 2017; **317**: 2491– 2. [CrossrefPubMedWeb of Science®Google Scholar](#)
- 16Jozwiak M, Oude Rengerink K, Benthem M, van Beek E, Dijksterhuis MG, de Graaf IM, et al. Foley catheter versus vaginal prostaglandin E2 gel for induction of labour at term (PROBAAT trial): an open-label, randomised controlled trial. PROBAAT Study Group. *Lancet*. 2011; **378**: 2095– 103. [CrossrefCASPubMedWeb of Science®Google Scholar](#)
- 17Ialongo C. Understanding the effect size and its measures. *Biochem Med (Zagreb)*. 2016; **26**: 150– 63. [CrossrefPubMedWeb of Science®Google Scholar](#)

- 18Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen.* 2012; **141**: 2–18. [CrossrefPubMedWeb of Science®Google Scholar](#)
- 19Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA.* 1994; **22**: 122–4. [Google Scholar](#)
- 20 WMA Declaration of Helsinki. Ethical principles for medical research involving human subjects. Available online at: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/> (accessed January 11, 2018). [Google Scholar](#)
- 21Pocock SJ, McMurray JJ, Collier TJ. Making Sense of statistics in clinical trial reports: part 1 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol.* 2015; **66**: 2536– 49. [CrossrefPubMedWeb of Science®Google Scholar](#)
- 22McNutt M. Reproducibility. *Science.* 2014; **343**: 229. [CrossrefCASPubMedWeb of Science®Google Scholar](#)