CHEST

Randomized Controlled Trials

Check for updates

Emily C. Zabor, DrPH; Alexander M. Kaizer, PhD; and Brian P. Hobbs, PhD

Randomized controlled trials (RCTs) are considered the highest level of evidence to establish causal associations in clinical research. There are many RCT designs and features that can be selected to address a research hypothesis. Designs of RCTs have become increasingly diverse as new methods have been proposed to evaluate increasingly complex scientific hypotheses. This article reviews the principles and general concepts behind many common RCT designs and introduces newer designs that have been proposed, such as adaptive and cluster randomized trials. A focus on the many choices for randomization within an RCT is described, along with their potential tradeoffs. To illustrate their diversity, examples of RCTs from the literature are provided. Statistical considerations, such as power and type I error rates, are discussed with the intention of providing practical guidance about how to specify study hypotheses that address the scientific question while being statistically appropriate. Finally, the freely available Consolidated Standards of Reporting Trials guidelines and US Food and Drug Administration guidance documents are introduced, along with a set of guidelines one should consider when planning an RCT or reviewing RCTs submitted for publication in peer-reviewed academic journals.

CHEST 2020; 158(1S):S79-S87

KEY WORDS: biostatistics; clinical trials; randomized controlled trials; study design

General Overview of Study Design

Clinical studies are conducted among human participants to generate new knowledge through describing the impact of interventions devised to improve the diagnosis, prevention, and treatment of human disorders. There are many types of designs for clinical studies, but they all aim to obtain objective data to evaluate interventions with respect to an associated outcome in a target population. The two main types of clinical study are: (1) clinical trials, in which participants are assigned to receive a certain intervention according to a prespecified research plan and are then followed up prospectively to observe the outcome of interest; and (2) observational studies, in which the study investigators do not assign the exposure or intervention that participants receive. The quality of evidence generated by any study is determined by its experimental design. In all clinical studies, bias may be introduced due to misclassification of interventions or outcomes and missing data. In nonrandomized studies, bias may also be introduced through selection of the included participants and confounding due to

ABBREVIATION: RCT = randomized controlled trial

AFFILIATIONS: From the Department of Quantitative Health Sciences & Taussig Cancer Institute (Drs Zabor and Hobbs), Cleveland Clinic, Cleveland, OH; and the Department of Biostatistics and Informatics (Dr Kaizer), Colorado School of Public Health, University of Colorado-Anschutz Medical Campus, Aurora, CO.

FUNDING/SUPPORT: B. P. H. was supported in part by the Case Comprehensive Cancer Center [P30 CA043703].

CORRESPONDENCE TO: Emily C. Zabor, DrPH, Department of Quantitative Health Sciences & Taussig Cancer Institute, Cleveland Clinic, 9500 Euclid Ave, CA-60, Cleveland, OH 44195; e-mail: zabore2@ccf.org

Copyright © 2020 American College of Chest Physicians. Published by Elsevier Inc. All rights reserved.

DOI: https://doi.org/10.1016/j.chest.2020.03.013

differences in prognostic characteristics. These sources of bias may make it difficult or impossible to measure precisely the impact of the intervention under study.

The strongest evidence for causality between the exposure or intervention under study and the outcome observed comes from prospective experimental designs that randomly assign interventions to trial participants. Randomized controlled trials (RCTs) have traditionally been viewed as the gold standard of clinical trial design, residing at the top of the hierarchy of levels of evidence in clinical study; this is because the process of randomization can minimize differences in characteristics of the groups that may influence the outcome, thus providing the most definitive evidence regarding the impact of the exposure or intervention on the outcome.[1,2] In an RCT, one or more treatments are compared vs a control group, and patients are assigned to treatment or control by chance, such as by rolling a die or flipping a coin. Each group in an RCT is called an "arm," so that, for example, a two-arm study may compare an experimental treatment vs a control group, and these would then be referred to as the "treatment arm" and the "control arm," respectively.

Description of Subtypes of Study Design

Active Control

In an RCT, the control arm can take a variety of forms. If there is a well-established treatment for the disease under consideration, this standard-of-care treatment could then be used as the control arm, against which the novel experimental treatment is compared. In an active control trial, the goal may be to show that the experimental treatment is superior to the standard-of-care treatment (ie, superiority study), to show that the experimental treatment is similar to the standard-of-care treatment (ie, equivalence study), or simply to show that the experimental treatment is not much less effective than the standard-of-care treatment (ie, noninferiority study). If there is already a known treatment for the condition under study that will be used as the active control arm, then it is very important, for ethical reasons, to ensure that there is sound scientific rationale that the experimental treatment will be at least as effective.[3]

Placebo Control

In the absence of an effective treatment for a disease, the control arm may consist of a group of patients receiving no treatment or receiving sham treatment, known as a placebo group. In a placebo group, inactive medication is given to patients in a way that is indistinguishable from the active treatment. For example, if patients assigned to the experimental treatment receive a small white capsule that they take three times a day, then the placebo group would receive a small white capsule with similar appearance, although containing no active ingredients, to take three times a day. In this way, effects of taking medication in and of itself, known as the placebo effect, are controlled between the two groups to reduce bias. Placebo-controlled trials seek to show that the experimental treatment is superior to the placebo. If no effective treatment is available for the condition being studied, then there are generally minimal ethical problems with a placebo-controlled trial. However, it is generally inappropriate and unethical to use a placebo if a treatment that improves outcomes is available for the condition under study.[3]

Multiple Arm

There are two main types of multiple-arm trials. The first includes multiple dose levels or regimens of the experimental treatment all compared vs a single-control arm. In these so-called dose-response studies, it is ideal to include a zero-dose, or placebo, arm to avoid a situation in which all doses show similar activity and to establish whether any of the doses was superior to no treatment.[3] The second involves a single treatment arm with multiple control arms (eg, both an active control and a placebo control arm).

Cluster Randomized

In a cluster randomized trial, groups of subjects are randomized as opposed to individual subjects. There are several reasons for randomizing clusters as opposed to individuals, including administrative convenience, ethical considerations, and ease of application at the cluster level. This trial design is more common in health services and policy research as opposed to studies of drug interventions. For example, it may be of interest to randomize hospitals in a study of a new educational initiative for physicians both for ease of implementation of the intervention at the hospital level, as well as to avoid within-hospital contamination across individual physicians receiving different interventions. In a cluster randomized trial, it is important to carefully consider the impact of both the number of clusters and the cluster size on the power of the study.[4] A pragmatic trial design known as the stepped wedge cluster randomized trial has been gaining in popularity. In this trial design, aimed at eliminating logistical constraints, each cluster

TABLE 1] Examples of Randomized Controlled Trials Published in *CHEST*

Reference	Participants	Treatment Arm	Control Arm	Primary Outcome
Pépin et al[11]	Patients with OSA initiating CPAP	Multimodal telemonitoring	Usual care	Systolic home blood pressure change from baseline to 6 months
Tanner et al[12]	Patients undergoing diagnostic bronchoscopic evaluation	Thin bronchoscope and radial endobronchial ultrasound	Standard bronchoscopy and fluoroscopy	Diagnostic yield
Furian et al[13]	Patients with mild to moderate COPD living below 800 m	Dexamethasone 8 mg	Placebo	Incidence of acute mountain sickness or other altitude-related adverse health effect
Semler et al[14]	Patients undergoing endotracheal intubation in the ICU	Ramped position	Sniffing position	Lowest arterial oxygen saturation from start to 2 min after intubation
Alansari et al[15]	Patients with bronchiolitis	100 mg/kg IV magnesium sulfate	Placebo	Time to medical readiness for discharge

undergoes a period, or "step," with no intervention followed by a step with exposure to the intervention.[5]

Adaptive Design

In traditional RCTs, most trial elements, such as randomization allocations and the number of study arms, are fixed throughout the trial. However, adaptive designs, in which accumulating information during the trial is used to modify some aspect of the trial, are becoming increasingly popular. For example, accumulating information may inform the randomization assignment of the next patient to enroll, which represents a form of adaptive randomization. These designs may allow more patients to be accrued to the arm that is showing more promise, thus reducing ethical concerns about continuing enrollment on fixed randomization designs in the face of possibly increasing evidence that one of the treatments under study is superior and allowing more patients in the course of the trial to be given more effective treatments or doses.[6,7] We note that to maintain the established scientific and ethical standards of randomized comparative trials with the acquisition of evidence that is both prospective and objective, it is essential to prespecify potential avenues for adaptation as well as establish corresponding statistical criteria in advance of implementing the trial.

Platform

Platform trials describe multiple-arm studies with the potential to include control and experimental arms that can be opened or closed for enrollment throughout the course of the trial based on decision rules regarding efficacy.[8,9] In this way, ineffective treatments can be discontinued before many patients have been treated with them, and newly emerging treatments that show promise can be added at any point. These designs allow investigation into more experimental treatments in a shorter period of time. In addition, compared with a series of stand-alone, concurrent two-arm designs, platform designs allow more patients to be assigned to experimental treatment arms as opposed to control arms.[10]

Use Cases of Study Design

The medical literature contains a multitude of examples of RCTs across many disease and intervention types. Some examples of recent two-arm randomized controlled trials published in *CHEST* are presented in Table 1[11-15] to demonstrate various applications of RCTs, although these examples are not exhaustive.

Benefits of Study Design

The primary benefit of RCTs comes from the randomization itself, which greatly reduces confounding from both known and unknown sources. In nonrandomized studies, it may be possible to control for known confounders, but it is much more difficult to control for unknown or unmeasured confounders, although some methods that attempt to do so are available. With randomization, causal conclusions regarding the exposure or intervention and outcome can be made. Additional benefits stem from the controlled and prospective nature of an RCT. The dosage, timing, frequency, and duration of treatment can be controlled, and blinding may be possible. Blinding refers to a

treatment assignment being unknown. In a single-blind study, patients do not know which treatment they are receiving. Blinding the patient prevents outcomes from being influenced by knowledge of treatment assignment. This is particularly important if any outcomes are self-reported. In a double-blind study, neither the patient nor the provider knows the treatment assignment. This additionally ensures that any care given by the provider or provider-assessed outcomes are not biased by knowledge of treatment assignment. Blinding is typically not possible in other study designs.

Downsides of Study Design

There are also disadvantages to RCTs. Because RCTs are highly controlled, the inclusion and exclusion criteria may lead to a homogeneous patient population, thus limiting the generalizability of the results to the broader population of patients with the condition being studied. RCTs also tend to study treatments in idealized environments, which may not be perfectly in-line with real-world usage of the treatment. Due to the complexities of design and conduct, RCTs are expensive and can take a long time to complete. RCTs are also not always feasible, for example, if a disease is very rare or if there are special considerations surrounding a disease that make randomized allocation either impractical or unethical.

Study Subject Considerations

An important consideration in the design of an RCT is the subject inclusion and exclusion criteria. These criteria will affect a variety of aspects of the conduct and interpretability of the study results and are primarily meant to ensure patient safety.[16] If eligibility criteria are too strict, it could be difficult to enroll the planned number of patients because of a potentially cumbersome screening process, as many screened patients may prove to be ineligible. In addition, very strict eligibility criteria could result in a study population that is not reflective of the broader target population, thus limiting the generalizability of the study result and the ability to establish the effectiveness of the treatment. These considerations need to be balanced with the fact that more strict exclusion criteria may be necessary to establish an intervention's efficacy. It is especially important to ensure there are medically valid reasons for excluding any commonly underrepresented groups, such as women, racial and ethnic minority groups, pregnant and breastfeeding women, and children.[17]

Equally important as the factors an investigator must consider when establishing patient eligibility criteria are the factors that potential study subjects consider when deciding whether to participate in a given trial. Most patients choose to participate in clinical trials with the hope of receiving a novel therapy from which they may benefit, yet the chance of receiving placebo or standard of care is also likely (often 50% or 33%). For these patients, the effort to participate in the trial is an altruistic act often driven by a desire to further scientific knowledge that may benefit future patients, if not themselves. With this in mind, investigators should additionally consider the burden placed on patients' time and energy throughout the course of the trial, and weigh that against the scientific importance of additional follow-up visits and invasive or time-consuming testing procedures.

Statistical Considerations

End Point Definition

There are several statistical considerations that an investigator must recognize when designing a RCT. The first is a clear and specific definition of the study end point. The end point needs to be an event or outcome that can be measured objectively so that the experimental group can be compared with the control group. For example, a study may wish to compare overall survival, rates of myocardial infarction, or improvements in quality of life. When selecting an end point for an RCT, the investigator must consider how the end point will be measured so that this will be standardized across all patients enrolled in the trial. The timing of assessment of the end point is also important to consider. For example, if interest is in overall survival, patients may need to be followed up for a long time before enough patients have died to determine whether there is a difference between the study arms. Alternatively, if interest is in rates of myocardial infarction, a time frame for occurrence of the event could be defined, such as myocardial infarction within 1 year of treatment start. It is common to have a primary end point in a study that is used as the basis for determining the needed sample size and ultimately making a decision regarding the efficacy of the experimental treatment, and then to include secondary end points as well; these secondary end points are more

exploratory in nature and would not be used to make decisions about whether the experimental treatment is better than the control.

Sometimes in a clinical study, it may be of interest to determine whether the experimental treatment is efficacious with respect to more than one end point. In such cases, it is possible to have co-primary end points or to use composite end points. Co-primary end points require that all primary end points indicate that the experimental treatment is superior to control to conclude efficacy, and they require special statistical considerations (discussed later in more detail). Composite end points are a single end point composed of more than one measure. For example, disease-free survival is a composite end point defined as recurrence of disease or death. Use of a composite end point can increase the number of events observed in a study; this use must be considered carefully, however, to ensure that the true outcome of interest will be captured by the composite end point and that the individual components of the composite end point align.

One area of caution in selecting an end point for an RCT relates to the use of surrogate end points. Surrogate end points are used to represent an end point that is either difficult to measure or takes too long to occur. For example, interest may be in death from heart disease but a surrogate end point of change in cholesterol level from baseline may be used. To confidently use a surrogate end point, an investigator must be certain that the effect of the intervention on the surrogate predicts the effect of the intervention on the true clinical end point of interest, which is often difficult or impossible to establish.

Effect Size

Once the end point is specifically defined, the investigators must then establish the meaningful difference between groups that they seek to detect. This is a clinical decision that has statistical implications for the design with respect to the number of patients that will be needed in the study. For example, in a study of an experimental treatment for lung cancer, in which overall survival is the primary end point, we know that 1-year overall survival on the standard-of-care treatment is 70%. Interest is in improving overall survival to 80% in patients on the experimental treatment. These rates can now be used to define the study hypotheses and determine the sample size required to conduct the study.

Power, Error Rates, and Sample Size

The traditional statistical approach considers two possible outcomes of the trial. Consider a two-arm RCT in which p_t represents the rate of a binary outcome in the treatment arm and p_c represents the rate in the control arm. The null hypothesis then represents the scenario in which the treatment has no benefit and is denoted as H_0: $p_t = p_c$ to indicate that the rates are equivalent in the treatment and control arms. The alternative hypothesis represents the scenario in which the treatment and control differ and is denoted as H_A: $p_t \neq p_c$ to indicate that the rates are not equivalent in the treatment and control arms. Note that in this example we have used what is termed a "two-sided alternative," meaning that we are looking for the two rates to not be equal, but the rate in the treatment group could be either higher or lower than the rate in the control group. A two-sided alternative such as this provides the most definitive evidence about an experimental treatment. However, the use of a two-sided alternative will require that a larger number of patients be enrolled in the trial, and at times, a "one-sided alternative" (choosing either H_A: $p_t > p_c$ or H_A: $p_t < p_c$ as the specified alternative hypothesis) could be appropriate.

Now that the end point has been defined, the effect size of interest has been established, and the null and

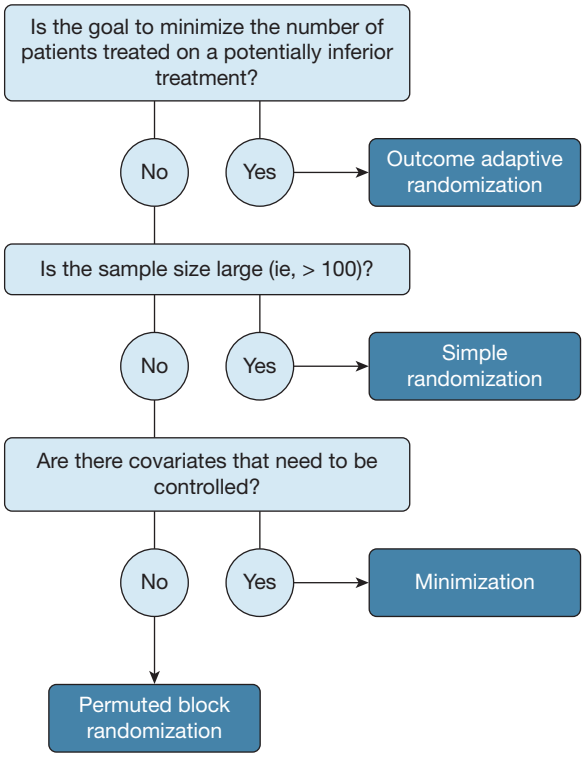

Figure 1 – *Flowchart for selecting a method of randomization.*

alternative hypotheses are fixed, the sample size needed for analysis can be calculated. Sample size is traditionally based on error rates. A type I error is rejecting a null hypothesis when it is true, and a type II error is failing to reject the null hypothesis when it is false, in which case the alternative hypothesis is assumed to be true. The type I error rate is conventionally set to 0.05, meaning that we are willing to accept a 1 in 20 chance that we will claim there is a difference between groups when in truth there is no difference. The complement of type II error (ie, 1 – type II error) is known as statistical power. Statistical power represents the probability of rejecting the null hypothesis when the alternative hypothesis is true. It is commonplace to ensure that statistical power is ≥ 0.8 in an RCT. The specific formula used to calculate the sample size will depend on many things, including the type of end point (eg, binary, continuous, time-to-event) as well as the study design, the type I and type II error rates, the variance of the end point, and allocation ratios to the various study arms. The calculation should also account for patients who drop out or who are lost to follow-up during the study. The details of all possibilities are outside the scope of this commentary, and a statistician should be consulted when designing an RCT.

For trials designed to examine multiple end points, such as the co-primary end points described previously, the possibility of committing a type I error may occur within each end point, yielding two types of false conclusions.[18] Marginal type I error rates estimate the type I error for each end point separately, whereas family-wise type I error rates consider the entire trial in violation if the null hypothesis is falsely rejected for a single end point. A family-wise type I error represents stronger control against false-positive findings for individual end points when directly compared with a marginal type I error. Marginal and family-wise errors are identical when testing a single end point. When multiple end points are being examined in a trial, control of family-wise error can be accomplished through adjustment for multiple comparisons. Common techniques include the Bonferroni procedure,[19] the Šidák procedure,[20] or the Holm's procedure.[21] There are many others, however, and consultation with a statistician is advised when designing a study that requires adjustment for multiple comparisons.

Methods for Randomization

Figure 1 depicts a flowchart for selecting from among the following methods for randomization.

Simple Randomization: In simple randomization, no restrictions are placed on the randomization sequence other than the final sample size. The randomization can be conducted with the flip of a coin or any other procedure that assigns each enrolling patient to an arm with an equal probability (eg, 1/2 for a two-arm study). However, because only the total sample size (and not the per-arm sample sizes) is being controlled in simple randomization, imbalances in the numbers assigned to the two or more arms are possible; this could lead to imbalances in subject characteristics, especially when the total sample size is small.

Permuted Block Randomization: To overcome the possible imbalances that can arise from simple randomization, the permuted-block design divides patients into blocks over time and balances the randomization to each arm within each block. If the total sample size is a multiple of the block size, balance is then guaranteed at the end of the study.[22] It is also possible to use unequal block sizes throughout the study, which would serve to further obscure any predictability of future treatment assignments.

Minimization: Although simple randomization will frequently control imbalance in prognostic factors between arms, it is still possible to have imbalances, especially if the sample size is relatively small (ie, $\leq$ 100). Often this is overcome through stratified randomization, in which simple randomization is conducted within groups based on important prognostic characteristics to ensure balance within those features. Stratified randomization can become cumbersome as the number of prognostic factors increases, and the strata must be accounted for when analyzing the resulting study data. Another approach, minimization, was introduced as an alternative to stratified randomization. With the minimization approach, assignment to a study arm is done with the intention of achieving balance between randomization groups with respect to prognostic factors of interest.[23] Minimization can greatly reduce prognostic imbalances between groups but at the cost of truly random treatment assignment, as assignment to a study arm is determined by the characteristics of patients who have already been assigned and not according to chance alone.

Outcome Adaptive Randomization: Adaptive randomization refers to randomization procedures that adjust the allocation ratio as the study progresses. In outcome adaptive randomization, the goal is to assign more patients to the more promising treatments based

on the accumulating data in the trial to minimize the expected number of treatment failures and to overcome the questionable ethics of continuing to assign patients to treatment arms for which there is evidence of poorer efficacy.[6] The most commonly used outcome adaptive randomization designs are based on the Bayesian statistical paradigm and include Bayesian adaptive randomization[24] and response adaptive randomization.[25]

Analytic Considerations

Prior to conducting an RCT, the analysis plan should be detailed. There are several ways that RCT data can be analyzed to account for lack of adherence. Consider a patient who is randomized to the experimental treatment arm but for whatever reason discontinues use of the treatment before completing the trial-specified regimen. How should this patient be incorporated into the statistical analysis at the end of the trial? One approach is "intention-to-treat," in which all patients are analyzed in the group to which they were randomized, regardless of their adherence to the treatment regimen. Intention-to-treat analysis is recommended in superiority trials to reduce bias, as the original randomization assignment is maintained. An alternative is a "per-protocol" analysis. In this approach, only patients who completed the treatment to which they were originally assigned are analyzed. A per-protocol analysis may lead to a purer estimate of the treatment effect, but the results of the analysis should be interpreted cautiously because bias can be introduced by the reasons subjects did not adhere to the planned treatment regimen. Often in RCTs, intention-to-treat will be the primary analysis approach but a per-protocol analysis will additionally be performed as a secondary analysis.

Another analytic consideration is how to accommodate the potential for treatment effect heterogeneity, which is the possibility that the treatment has different effects in different subgroups of patients. The heterogeneity of treatment effect, in its extreme, can make the overall population effect seem clinically insignificant when there are certain subpopulations that would benefit from the treatment and other subpopulations that do not benefit. To avoid potential inflation of the type I error rate, any subgroup analyses should be specified prior to the trial, rather than as a post hoc attempt at salvaging a trial with a null result. Traditionally, the approach has consisted of a priori subgroup analyses that consider subgroups "one variable at a time," with results graphically presented in forest plots. However, consensus building research has sought to identify more efficient methods to consider all relevant patient attributes simultaneously with the Predictive Approaches to Treatment Effect Heterogeneity statement.[26]

Reporting Considerations

The Consolidated Standards of Reporting Trials guidelines (http://www.consort-statement.org/) were established to guide investigators in appropriate reporting of results of RCTs. These guidelines consist of a 25-item checklist for use in putting together a report on an RCT, as well as a template for a flow diagram to include in the trial report indicating the breakdown of sample size at various stages of the study.

Available Standards

The most comprehensive guidelines regarding the design and conduct of clinical trials are published by the US Food and Drug Administration (https://www.fda.gov/regulatory-information/search-fda-guidance-documents). In addition to documents referenced earlier that provide guidance on selection of an appropriate control group, information regarding the use of placebos and blinding is also available. The International Conference on Harmonisation of Technical Requirements for Pharmaceuticals for Human Use have also created guidelines for the conduct of clinical trials that cover the selection of end points, various design options, trial monitoring, and analytic considerations.[27]

Short List of Questions to Guide the Researcher

1. **Consult with a statistician early and often.** It is never too early to involve a statistician in planning the design of an RCT.
2. **Select an intervention(s).** For treatment trials, dosage and duration of treatment should also be determined.
3. **Select an appropriate control arm and consider the practical and ethical considerations of using a placebo or standard of care.**
4. **Define the study end point as specifically as possible, including when it will be assessed.**
5. **Establish the effect size of interest.** How much improvement are you hoping to see in the experimental treatment arm?
6. **Write down the study hypotheses and determine whether a one-sided or two-sided alternative hypothesis is most appropriate.**

7. **Select acceptable type I and type II error rates and calculate the sample size needed to detect the desired effect size.** If multiple primary end points are being used, be sure to consider marginal vs family-wise error rates.
8. **Determine the feasibility of accruing the number of patients needed according to the calculated sample size.** Can this be accomplished at a single institution, or are multiple institutions needed? If multiple institutions are needed, how does this affect study implementation?
9. **Write down the analysis plan in detail before beginning the study.**

Short List of Questions to Guide the Reviewer

When reviewing a manuscript describing a randomized controlled trial, consider commenting on the following:

1. **The exposure or intervention in the treatment arm and control arm.** Was there justification for the exposure or intervention in the treatment arm? If the control arm received standard of care, was an appropriate standard of care applied? If there was a placebo control arm, was there a possibility of distinguishing the treatment and control arms by the nature of the intervention?
2. **Key features of the study methodology.** Were appropriate study end point(s) chosen? Was their measurement accurate and consistent? Was the randomization procedure appropriate? Were details of the sample size calculation, including the anticipated effect of the intervention, provided? Was drop out handled appropriately in the planning and analysis of the trial? Is the Consolidated Standards of Reporting Trials guidelines flowchart included in the report?
3. **The reported results and their interpretation.** Were the reported results in line with the planned analyses? Was the interpretation of the results made based on the planned primary end point? Was the interpretation of the results appropriate?

Acknowledgments

Financial/nonfinancial disclosures: The authors have reported to *CHEST* the following: B. P. H. is scientific advisor to Presagia; consultant for SimulStat; and receives research funds from Amgen. None declared (E. C. Z., A. M. K.).

Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

References

1. The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J*. 1979;121(9):1193-1254.
2. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. 1989;95(suppl 2):2s-4s.
3. US Department of Health and Human Services Food and Drug Administration. Guidance for Industry: Choice of Control Group and Related Issues in Clinical Trials. 2001. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e10-choice-control-group-and-related-issues-clinical-trials. Accessed May 15, 2020.
4. Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. *BMJ*. 2017;358:j3064.
5. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350:h391.
6. Hobbs BP, Berry DA, Coombes KR. Biostatistics and bioinformatics in clinical trials. In: Niederhuber JE, Armitage JO, Doroshow JH, Kastan MB, Tepper JE, eds. *Abeloff's Clinical Oncology, Sixth Edition*. Philadelphia, PA: Elsevier; 2020. 284-295.e2.
7. Hobbs BP, Lee JJ. Adaptive trial designs. In: Crowley J, Hoering A, eds. *Handbook of Statistics in Clinical Oncology*. Third Edition. Boca Raton, FL: Taylor & Francis/CRC Press; 2012:265-292.
8. Renfro LA, Sargent DJ. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Ann Oncol*. 2017;28(1):34-43.
9. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med*. 2017;377(1):62-70.
10. Hobbs BP, Chen N, Lee JJ. Controlled multi-arm platform design using predictive probability. *Stat Methods Med Res*. 2018;27(1):65-78.
11. Pépin JL, Jullian-Desayes I, Sapéne M, et al. Multimodal remote monitoring of high cardiovascular risk patients with OSA initiating CPAP: a randomized trial. *Chest*. 2019;155(4):730-739.
12. Tanner NT, Yarmus L, Chen A, et al. Standard bronchoscopy with fluoroscopy vs thin bronchoscopy and radial endobronchial ultrasound for biopsy of pulmonary lesions: a multicenter, prospective, randomized trial. *Chest*. 2018;154(5):1035-1043.
13. Furian M, Lichtblau M, Aeschbacher SS, et al. Efficacy of dexamethasone in preventing acute mountain sickness in COPD patients: randomized trial. *Chest*. 2018;154(4):788-797.
14. Semler MW, Janz DR, Russell DW, et al. A multicenter, randomized trial of ramped position vs sniffing position during endotracheal intubation of critically ill adults. *Chest*. 2017;152(4):712-722.
15. Alansari K, Sayyed R, Davidson BL, Al Jawala S, Ghadier M. IV magnesium sulfate for bronchiolitis: a randomized trial. *Chest*. 2017;152(1):113-119.
16. Kim ES, Bruinooge SS, Roberts S, et al. Broadening eligibility criteria to make clinical trials more representative: American Society of Clinical Oncology and Friends of Cancer Research Joint Research Statement. *J Clin Oncol*. 2017;35(33):3737-3744.
17. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA*. 2007;297(11):1233-1240.
18. Hobbs BP, Thall PF, Lin SH. Bayesian group sequential clinical trial design using total toxicity burden and progression-free survival. *J R Stat Soc Ser C Appl Stat*. 2016;65(2):273-297.
19. Dunn OJ. Multiple comparisons among means. *J Am Stat Assoc*. 1961;56(293):52-64.

20. Šidák Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc*. 1967;62(318):626-633.

21. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6(2):65-70.

22. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin Trials*. 1988;9(4):327-344.

23. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975;31(1):103-115.

24. Thall PF, Wathen JK. Practical Bayesian adaptive randomisation in clinical trials. *Eur J Cancer*. 2007;43(5):859-866.

25. Lee JJ, Xuemin G, Suyu L. Bayesian adaptive randomization designs for targeted agent development. *Clin Trials*. 2010;7(5):584-596.

26. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement [published online ahead of print November 12, 2019]. *Ann Intern Med*. https://doi.org/10.7326/M18-3667.

27. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, Statistical Principles for Clinical Trials. 1998. https://database.ich.org/sites/default/files/E9_Guideline.pdf. Accessed May 15, 2020.