Clinical Trials in Cardiovascular Medicine in an Era of Marginal Benefit, Bias, and Hyperbole

Joseph Loscalzo, MD, PhD

Clinical trials in the modern medical era began in the mid-20th century with the first randomized, double-blind assessment of an antibiotic, streptomycin, for the treatment of pulmonary tuberculosis. Over the course of the next 50 years, the randomized, controlled trial became the “gold standard” for testing the efficacy of new therapies and is generally required for their approval by regulatory agencies worldwide. To be sure, many effective cardiovascular therapies have been identified and approved for use following this generally accepted approach. In this editorial, I will review the many factors that can and do adversely affect the performance and interpretation of contemporary clinical trials in cardiovascular medicine, hoping to shed light on what I view as a series of growing and only partly remediable challenges to effective clinical investigation.

Clinical trials in cardiovascular medicine set the standard for randomized, controlled trial design owing to the population prevalence of the major cardiovascular diseases, the event rates in these patient populations, and the practically achievable sample sizes sufficient to minimize type II errors. In the early studies of fibrinolytic therapy for myocardial infarction (ISIS [International Study of Infarct Survival] and TIMI [Thrombolysis In Myocardial Infarction]), for example, the expected 5-week mortality rate for conventionally treated patients was 12% to 13% in Europe and the expected 5-week mortality rate for conventionally treated patients was 12% to 13% in Europe and the United States. Thus, with several thousand patients in each treatment arm, one could expect to detect confidently true reductions in absolute mortality of 3% to 4%, as was the case in these landmark trials.

Over the past three decades, mortality rates for highly prevalent cardiovascular diseases, including acute coronary syndromes, heart failure, and sudden death, have continuously improved owing to the clear benefits of therapies proved to be efficacious in double-blind, randomized, controlled trials. With these mounting, cumulative successes, however, the marginal benefit of any proposed intervention decreases. Realistic limits, both operational and financial, to the size of study samples decrease the statistical power and decreases. Realistic limits, both operational and financial, to the size of study samples decrease the statistical power and the absolute treatment effect detectable in these trials. To appreciate the implications of narrowing clinical margins for designing and interpreting clinical cardiovascular trials, we must digress to discuss a few relevant statistical principles, including the application of Bayesian analysis to trial design.

Although statistical power and significance are well-known determinants of the probability that a research observation is true, another important factor to consider is the prior (prestudy) probability that the observation is true.\(^1\)\(^2\) According to Ioannidis,\(^3\) the likelihood that a statistically significant research finding is more likely true than false is defined by the inequality \((1 - \beta) R > \alpha\), where \(\beta\) is the type II error rate, \((1 - \beta)\) is the statistical power, \(\alpha\) is the type I error rate, and \(R\) is the prestudy odds of a true relationship to no relationship [which defines the prestudy probability, \(R/(R + 1)\)]. As \(R\) increases above 1, the prestudy probability approaches 1, and satisfying the inequality requires less statistical power; however, as \(R\) decreases below 1, the prestudy probability approaches 0, and satisfying the inequality requires increasing statistical power. In all cases, \(R\) must be greater than \(\alpha\) to satisfy the inequality. Put in more practical terms, the higher the prestudy probability that a treatment is effective, the smaller the sample size required to prove its benefit. Although this statement is intuitively obvious, its implications for trial design are interesting to contemplate. Clearly, a reasonable assessment of prestudy probability can influence sample size calculations greatly (and, in the absence of preliminary data, estimates of effect size may implicitly incorporate prestudy probability). Ideally, high prestudy probabilities are preferred because they limit sample sizes required for clinical trials, whether they are estimated directly or incorporated implicitly in sample size estimates. Even in the setting of a high prestudy probability, however, a small sample size compromises the precision of the estimated treatment effect.

We should not lose sight of an interesting ethical trap that a high prestudy probability yields for clinical trialists, viz, it minimizes clinical equipoise or the belief by the trialist that all possible trial outcomes are equally likely to occur. Without clinical equipoise, subjects enrolled in the placebo or standard treatment arm of a therapeutic trial are, one can argue, unnecessarily exposed to substandard therapy. Obviously, if the prestudy probability of benefit of an intervention is very high, there is not only no ethical justification for enrolling subjects in the control arm of the trial but also no reason to perform the trial in the first place. Thus, one needs to balance a prestudy probability that is greater than 0.5 (but not too much greater) with the cost of an excessively sized trial and the importance of defining the benefit of an as yet unproved therapy for the affected population. In a trial of a highly effective standard therapy compared with the added effect of a new therapy, considerations of truly marginal...
Two other factors reduce the probability that a reported positive clinical trial observation is true: bias and the statistical effect of repeated, independent testing of the same experimental question. In its most general sense, bias is defined as the unwarranted influence of experimental design or interpretation on study outcome. Bias can cause the investigator to conclude either that a true relationship exists when one does not (standard bias) or that no relationship exists when one does (reverse bias). The influence of bias offsets the statistical power of an experimental trial by a factor that reflects the proportion of false research findings reported as true findings; the greater this ratio, the greater the impact of bias on statistical power.

In cardiovascular medicine, important clinical questions are frequently addressed by several groups. And, of course, experimental reproducibility by external validation is essential for acceptance of any experimental outcome by the scientific community. Although it is surely comforting to find that more than one group reaches the same conclusion, there is a statistical dark side to multiple independent testing of the same question: as the number of studies increases, the probability that at least one of them will claim a statistically significant difference between groups increases, and thus, the positive predictive value of the study decreases. Quantitatively, when \( n \) independent studies of identical power are applied to a common experimental question, an analysis of the probability that an observation is true rather than false leads to the expression \( R(1-\beta)^n/R(1-\beta)^n+1-(1-\alpha)^n \), which defines the positive predictive value (PPV) of the observation. As \( n \), the number of independent studies, increases, \( (1-\alpha)^n \) approaches 0, \( (1-\beta)^n \) approaches 1, and the PPV decreases, approaching \( R/(R+1) \), the prestudy probability. Stated qualitatively, increasing the number of studies on the same question reduces the PPV of any single study toward the limit of the prestudy probability itself, which suggests (counterintuitively) that the number of studies is inversely related to their individual ability to predict a positive effect.

In his recent analysis, Ioannidis applies Bayesian principles to an assessment of the influence of bias and of multiple testing on poststudy outcomes and shows a dramatic range of effects. Notwithstanding the theoretical and practical impact of this kind of analysis, it is important to realize that regulatory authorities in the United States and abroad generally insist on the strict frequentist approach rather than the Bayesian approach to trial design and interpretation. This perspective is a sensible one in that it minimizes the chance that ineffective agents would be approved and optimizes sample size for the detection of infrequent adverse effects.

Turning from this theoretical analysis to the real world of cardiovascular trials, let me now expand on the practical application of these statistical concerns and consider common causes of erroneously interpreted trial results that derive from them. These issues fall into four general categories: common errors in trial design and statistical assessment, bias and conflict of interest, shortcomings of the peer review process, and postpublication interpretation of trial results and their dissemination to practitioners and the public.

Errors in clinical trial design and statistical assessment are, unfortunately, more common that a careful student of the art should accept. Inappropriate control groups (eg, historical controls), study groups inadequately sized to minimize type II errors, lack of blinding during outcome assessment, failure to correct for repeatedly inspecting the data during the course of the trial, excessive reliance on \( \alpha=0.05 \) without regard for prestudy probability and marginal effect size, and lack of appropriate statistical correction for multiple comparisons between study groups—these all-too-common flaws pepper the clinical trial landscape, not infrequently invalidate study conclusions, and are, unfortunately, not always identified before publication.

Two elements of trial design that are increasingly used in cardiovascular clinical studies warrant special consideration: the use of composite end points and the use of a noninferiority design. Although neither of these approaches is intrinsically flawed, the interpretation of the results of trials that use them requires caution. In the case of combined end-point analysis, it is not uncommon to observe that the major difference between treatment arms is driven by a single element in the composite; yet, in announcing the results of the trial, other elements of the combined end point, either tacitly or overtly, are inappropriately viewed as affected by the treatment. In the case of noninferiority trials, one iteratively approaches an effect rate that shows the treatment to be no better than placebo; after several such iterations, however, this kind of analysis can mitigate the true benefit of a treatment. Clearly, the analysis of clinical trials that employ these design features requires careful interpretation before one decides about implications of such studies for clinical care paradigms.

Bias, of course, takes many practical forms, not the least of which is the failure of a trialist to assume the null hypothesis at the outset of the study. Regardless of type, bias can compromise the internal validity of a trial. In addition, data obtained in cell systems, animal studies, or pilot human studies, as well as investigator hyperbole, often prejudice the trialist in favor of a positive result. With the first publication purporting to show a statistically significant benefit of an intervention, the stage is set for subsequent studies to affirm this initial result, often through tacit intrusion of selection or attrition bias.

Bias can also intervene during the course of a trial as a material reflection of conflict of interest among investigators. Conflict of interest can itself take many forms, the most obvious being the successful demonstration of the benefit of an intervention in which the trialist has a personal interest (financial or academic, direct or indirect). Failure of the sponsoring pharmaceutical or device company to remain at arm’s length in maintaining the trial data set and in its analysis represents the most challenging of conflicts through which significant bias can be imposed on an otherwise well-conducted trial. Recently publicized cases highlight specific difficulties in this arena, leading to a call from the editors of major medical journals, including Circulation, for the strict application of common principles to all submitted clinical trial manuscripts to limit inappropriate influence of the sponsoring organization on data analysis and interpretation.
Flaws in the review process are, regrettably, not uncommon. Limited expertise by the reviewer, hurried reviews that fail to detect major problems, reviewer bias, and insufficient attention to detail can, on occasion, plague the review of even the very best papers by the very best journals. It is the responsibility of the editors to identify these problems and address them as they arise by seeking alternate reviewers, responding fairly to rebuttals by the authors, and continually striving to improve the reviewer pool and the review process. To be sure, peer review is an imperfect process; however, it is the best process currently available. All of us whose careers have benefited from fair and impartial reviews are, in my view, obligated to maintain and improve this process by working with editors.

In addition to reviewer bias, editors are themselves subject to publication bias. Publication bias refers to the widely recognized observation that studies with positive results have a better chance of being published than those with negative results. Although all editors are aware of this concern, and many have made a concerted effort to publish scientifically sound studies with negative results to address it, positive trial results by their very nature tend to be more appealing to editors, reviewers, and readers than negative trial results. For this reason, vigilance is essential to minimize the often insidious influence of this common editorial predisposition. Another interesting variation on this theme of publication bias involves decisions not to publish articles that are sound from the perspective of study design and statistical analysis but which, in the best estimation of the reviewers and editors, have the potential to affect adversely patient care or medical practice. I am certain that some editors would not refer to these kinds of decisions as a reflection of bias but, rather, as a reflection of the consideration of a thoughtful editor in the exercise of his/her responsibilities to the reader, and I, for one, do not disagree with this conclusion.

After publication, a host of additional events can occur that adversely affect the interpretation of a clinical trial. These postpublication problems largely fall under the category of hyperbole, and this exaggeration or overinterpretation of study results can arise from several quarters. The authors can (with or without the sponsoring institution’s complicity) publicize the outcome of a trial in a less than objective way. The media can, of course, do the same, with or without the authors’ consent. Other elements of society likely to be influenced by the outcome of a trial can also distort the message of a study to suit their own ends, which can include competing interests that are political, social, and economic in nature. Institutional drivers may also promote an exaggeration of the study’s message, especially in an environment that is increasingly competitive for national (and international) institutional recognition: high-profile studies are often used by even the finest academic institutions as a marketing tool for expanding referrals from other physicians and from patients themselves. If academic physicians are often their own worst enemies by virtue of overstating the importance and timeliness of a trial result, institutions only amplify that concern by failing to temper their interpretation of the exaggerated message.

As yet another factor that affects postpublication interpretation of a clinical trial, we should consider meta-analyses and mega-analyses. These statistical approaches, although ideally utilized to generate hypotheses that warrant further testing, have, on occasion, erroneously influenced practice recommendations. Without taking into account all available trials (and often the negative trials are not available in the published literature) of similar design and demographics, one can only interpret the results of these analyses as suggestive of future trial questions and study design. To use them to guide or justify changes in clinical practice overstates their importance and utility and can mislead the practitioner and the public.

Clearly, there are many factors that can affect the performance and interpretation of modern cardiovascular trials adversely. These factors influence studies at many different phases of their implementation and dissemination. A timeline is presented in the Figure that summarizes the major challenges to interpretation of a study and their influence during the course of trial design, performance, publication, and dissemination.

Clinical trials in cardiovascular medicine have evolved dramatically over the past three decades, and their results have led to significant improvements in patient care and outcome. Although we as a society continue to have much to gain from a well-conducted and fairly interpreted study, we also have much to lose from one that is flawed, overinterpreted, or distorted in message. A thoughtful and responsible approach to the problem is certainly warranted, and that approach should be developed, advocated, and sustained by trialists and their study subjects, as well as medical journal editors and the lay press.
At Circulation, we are intensifying our efforts to ensure that the studies we publish are presented and viewed in the most objective light. We adhere to the International Medical Journal Editors’ policy statement on the acceptance of studies that are free from unwarranted influence by sponsoring entities.9 As with most biomedical journals, we require financial conflict of interest disclosure by all authors. At the weekly associate editors’ meeting, we discuss each manuscript, inviting comments from all associate editors and three statistical consultants, even those not directly responsible for handling the submission, to optimize the breadth of perspective from which each manuscript is scrutinized. In addition, we are embarking on a special series on statistics for cardiovascular investigators, with Dr Martin Larson as Series Editor. We will use this series as another communication tool through which to offer a better understanding of the increasingly complex field of statistics as applied to clinical studies.

Lastly, we work with the lay press, both through the American Heart Association and directly, to attempt to ensure that the results of studies published in the journal are presented in the most objective and conservative light. We truly believe that only through this rigorous approach can we as editors hope (and you as readers expect) that what we publish is of the highest quality, presented carefully, with minimal hyperbole, for maximal patient benefit.

Acknowledgments
The author wishes to thank Drs Martin Larson, Elliott Antman, and Joseph Vita for helpful comments and suggestions.

References

Key Words: Editorials ■ trials
Clinical Trials in Cardiovascular Medicine in an Era of Marginal Benefit, Bias, and Hyperbole
Joseph Loscalzo

Circulation. 2005;112:3026-3029
doi: 10.1161/CIRCULATIONAHA.105.594614

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/112/20/3026

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/