

# Reporting and Interpretation of Randomized Clinical Trials

Howard Bauchner, MD; Robert M. Golub, MD; Phil B. Fontanarosa, MD, MBA

**High-quality randomized clinical trials** (RCTs) occupy the highest position on the evidence pyramid, either as stand-alone studies or as part of meta-analyses. However, the interpretation of an RCT is often complicated and involves an understanding of the factors involved in the overall study design, decisions regarding prespecified outcome measures, assumptions made in statistical analysis plans, actual conduct of the trial, event rates observed and estimates of uncertainty around the outcomes, and generalizability of the findings. In addition, the long-standing concern about the difference between statistical significance and clinical importance continues to be discussed, along with other issues, including the balance of potential benefits and risks, the cost of the intervention, and the value that individual patients and a population might achieve from the intervention. The interpretation of an RCT must consider all of these factors.

These issues and others related to the reporting and interpretation of RCTs sometimes conflate the very different roles of scientific journals (and editors), study investigators (including research methodologists), and the ultimate consumers of the trials (including other researchers, clinicians, and patients). The distinction among these roles is critical.

## The Role of Journals and Editors

Journals and editors have 2 fundamental roles with respect to RCTs. The first is to conduct editorial evaluation and peer review with the goal of reaching a conclusion about whether the study findings are likely to represent an estimate of the truth. This process entails assessment of whether the study design and the statistical methods were appropriate to answer the question posed by the researchers and, if so, whether the study execution may have introduced important problems that undermined its validity. The evaluation also includes determining whether the study adhered to the protocol, statistical analysis plan, and trial registration as specified by the investigators.

This review process is not always straightforward and may be complicated by various factors, such as early stopping of the trial (for futility or other reasons), poor adherence to study interventions, high loss to follow-up, or extensive modifications in trial procedures or prespecified statistical analytic plans. Moreover, because no study is perfect in design or execution, judgment is required as to whether there is a high likelihood of validity despite study flaws and limitations.

The second role of journals and editors is to try to ensure that the article as published represents a precise, accurate, and clear scientific report of the study as designed and executed, and that the report includes the key material an informed reader needs to critically evaluate the study. This involves presenting an article using a standard format that, in the judg-

ment of the journal, communicates the information in a consistent and accessible way, and that does not go beyond what the data in the trial support. For clinical trials, reporting guidelines, such as the Consolidated Standards of Reporting Trials (CONSORT),<sup>1</sup> and CONSORT extensions, such as for reporting noninferiority and equivalence trials<sup>2</sup> or multiple-group trials,<sup>3</sup> provide a useful framework for presentation of the main elements of trial methodology and results.

A key principle is that the published article is intended to represent a faithful scientific record of the clinical trial. The report needs to follow the study protocol and statistical plan as prespecified or formally amended, with strong justification for deviations. The report does not represent the study the editors or external peer reviews would have preferred the investigators had designed. Equally important is that the report of the trial does not represent the study other methodologists and statisticians would have designed given their preference, even if their community may be moving in directions different from when the trial was formalized. The obligation of the journal is to report, not to modify, the trial, although in some cases, it may be reasonable for reviewers and editors to request additional analyses to supplement and clarify the prespecified analytic approach.

It is also important for researchers and readers to recognize that the article is not the study. The report of the trial can only be a limited representation of the study in the same way that a photograph of a body is not the body, and even a body is only a representation of a person. While appended protocols and statistical analysis plans can provide details about the study design, it is impossible for the article to fully convey what occurred during the execution of the trial, and impractical for the article to present all the raw data collected (or even, usually, all the analyzed data). The published article represents a formalized record of the study, with limitations in what can be communicated even if article length were not a barrier. Moreover, if the article is only a summary of the study, the Abstract is but a skeleton of the article. The Abstract cannot provide sufficient detail to critically appraise the study for validity, and for studies that have any complexity the Abstract should not be used as the basis for nuanced interpretation, even if caveats are included.

Some trials have straightforward results, such as those that have very little dropout of enrolled participants, no early termination, clinically meaningful end points, unambiguous primary results, and reasonably consistent secondary end points. When this is not the case, it is responsibility of the journal editors to make sure that the challenging issues are clearly conveyed and that the nuances are objectively addressed in the Discussion section of the article.

### The Role of Investigators and Methodologists

As with all fields of science, the sciences of study design and statistics are in constant evolution. Methods that were considered standard and important at the time a trial was designed may no longer be favored by the time the trial is finished, particularly for trials that take many years to complete. As one example, for an older trial last observation carried forward may have been the acceptable and prespecified approach to handling missing data at the study inception, but this approach has generally been superseded by other techniques such as multiple imputation, when appropriate. Other methods such as Bayesian statistics have gained traction but have not become widely embraced as a new standard.<sup>4,5</sup> Many other issues in study design, such as the best approach to significance testing, are still under active discussion.

For instance, recently there has been rekindling of the debate about the use of the term “significance” and reporting of *P* values based on statistical testing when reporting findings from clinical trials (and other types of studies).<sup>6</sup> Some have advocated removing the term “significance” from descriptions of the results of clinical trials, simply providing effect sizes, generally with 95% confidence intervals, and allowing the authors (and readers and others) to use some other approach to interpret whether the observed findings are likely to represent a true effect vs a sampling error, as well as whether the effect size is important. These arguments acknowledge that the most commonly used significance threshold of .05 represents a historical tradition rather than a rationally established cut point. Others have continued to advocate for describing results of clinical trials in terms of statistical significance, in part because of the need for a starting point in discussion; decisions that are made by regulatory bodies, such as the US Food and Drug Administration, which are generally dichotomous; and the need to assist clinicians and patients in their interpretation and operationalization of clinical trial results.

Over time, the research community comes to a consensus about preferred methods—a consensus that may last only until the next set of methods are developed. As the community evolves, the newer methods become integrated into study design and then appear with increasing frequency in the studies submitted to the journals. For example, when alternatives to hypothesis testing using significance thresholds become established, they will be incorporated into the study from inception. When this process occurs, those methods become part of the scientific report.

However, at times, journals may publish post hoc analyses that use other or newer methods that may be as (and in some cases, perhaps more) informative than the preplanned design, although these reports should include proper context for the new analyses and appropriate caveats. For example, a post hoc Bayesian reanalysis of a previously published RCT that compared the effects of venovenous extracorporeal membrane oxygenation (ECMO) vs conventional mechanical ventilation on mortality among patients with severe acute respiratory distress syndrome was performed because of divergence between the clinical vs statistical significance of the trial findings and continued controversy over the benefit of ECMO.<sup>7</sup>

### The Role of Readers and Other Interested Individuals

Clinicians may have the most challenging role in understanding the results of RCTs because they have to take the (necessarily) incomplete information provided in the scientific report, decide whether it is potentially actionable, and, if so, decide whether and how the findings can be applied to individual patients. An Editorial that puts the study in context of other research, delves into the clinical implications, and highlights the limitations in study interpretation can be helpful for readers in interpreting the study.<sup>8</sup> But ultimately it is the responsibility of the clinician and others to read the article in depth, and in particular to consider complex issues raised in the Discussion, where strengths and limitations can and should be addressed at length.

### Approaches Used at JAMA for Reports of Randomized Clinical Trials

For clinical trials submitted to *JAMA*, authors are required to ensure that the study design and results are reported with fidelity to and consistency with the a priori decisions that investigators made in designing their trials, as prespecified and documented in the trial registration, study protocol, and statistical analytic plan. The editors review these documents carefully for consistency and to ensure that authors provide explanations and justification for any differences among these documents. Investigators declare in those documents the hypotheses being tested in the trial, the primary outcome the study is designed to examine, and the statistical criteria that allow investigators to evaluate the probability or implausibility of observing the results obtained. These factors form the basis for the interpretation of a trial that demonstrates the effect of an intervention, or the interpretation of a trial that fails to demonstrate an effect (ie, “null” or “neutral” results), in either case determined by a priori criteria.

For trials in which investigators use a frequentist approach (the most common approach to statistical inference in the current clinical literature) in the design and analysis, the reporting of results should reflect that approach, including interpretation based on the prespecified effect size estimate, the sample size anticipated (vs achieved), and the criteria and threshold for declaring statistical significance. For studies in which hypothesis testing involves prespecified use of *P* values, those values should be reported with the study results, along with confidence intervals for estimates of effect size. Findings should be described as “statistically significant differences” (or “no statistically significant differences”) rather than just differences, to differentiate between the results of the prespecified statistical testing and the clinical interpretation of the trial. Effect sizes likely to have clinical relevance are termed “clinically important” rather than “significant” to help make this distinction clear.

For trials in which investigators designed the study using a Bayesian approach (or other evolving designs) rather than a frequentist approach to analysis, the reporting of the results should be consistent with the prespecified analytic plan and follow accepted norms for that analytic approach.

The presentation and interpretation of the results from most clinical trials are usually straightforward. The results for the prespecified primary outcome take priority and precedence over all

### Box. Examples of Abstract Conclusion and Relevance Statements for Some Clinical Trials That Highlight Important Limitations, Caveats, or Other Important Factors in the Trial Interpretation

For most clinical trials, the conclusion can be represented with a single statement of the primary outcome:

"Among obese patients undergoing surgery under general anesthesia, an intraoperative mechanical ventilation strategy with a higher level of PEEP and alveolar recruitment maneuvers, compared with a strategy with a lower level of PEEP, did not reduce postoperative pulmonary complications."<sup>9</sup>

For some clinical trials, an additional statement in the Conclusion and Relevance section may be helpful to highlight important limitations or other issues, or to provide caveats to aid in interpretation, without overstating the importance of the study.

- For a trial that had event rates that were lower than expected or failed to achieve anticipated sample size:
 

"Among ambulatory adults with hypertension, treating to a systolic blood pressure goal of less than 120 mm Hg compared with a goal of less than 140 mm Hg did not result in a significant reduction in the risk of probable dementia. Because of early study termination and fewer than expected cases of dementia, the study may have been underpowered for this end point."<sup>10</sup>
- For a trial that encountered unexpected challenges, such as imbalance in baseline characteristics between groups, differential drop-out or crossover between groups:
 

"Among patients with AF, the strategy of catheter ablation, compared with medical therapy, did not significantly reduce the primary composite end point of death, disabling stroke, serious bleeding, or cardiac arrest. However, the estimated treatment effect of catheter ablation was affected by lower-than-expected event rates and treatment crossovers, which should be considered in interpreting the results of the trial."<sup>11</sup>
- For a trial that reported preliminary findings or results from a single-center investigation and needs to signal the importance of additional confirmatory research before results are adopted in clinical practice:
 

"In this emergency department, use of a bougie compared with an endotracheal tube + stylet resulted in significantly higher first-attempt intubation success among patients undergoing emergency endotracheal intubation. However, these findings should be considered provisional until the generalizability is assessed in other institutions and settings."<sup>12</sup>

- For a trial in which the primary outcome may meet prespecified criteria for declaring statistical significance, but the observed effect size does not meet the prespecified criteria for the minimally important clinical difference:

"Among patients undergoing THA, paracetamol plus ibuprofen significantly reduced morphine consumption compared with paracetamol alone in the first 24 hours after surgery; there was no statistically significant increase in SAEs in the pooled groups receiving ibuprofen alone vs with paracetamol alone. However, the combination did not result in a clinically important improvement over ibuprofen alone, suggesting that ibuprofen alone may be a reasonable option for early postoperative oral analgesia."<sup>13</sup>

- For a trial in which the primary outcome met prespecified criteria for declaring statistical significance, but the observed effect size was modest and the interventions were associated with important adverse events:

"Among patients with moderate to severe OA of the knee or hip and inadequate response to standard analgesics, tanezumab, compared with placebo, resulted in statistically significant improvements in scores assessing pain and physical function, and in PGA-OA, although the improvements were modest and tanezumab-treated patients had more joint safety events and total joint replacements. Further research is needed to determine the clinical importance of these efficacy and adverse event findings."<sup>14</sup>

- For a trial in which the primary outcome does not meet the prespecified criteria for declaring significance, but the interpretation should include consideration of observed effect size:

"Among patients with morbid obesity, use of laparoscopic sleeve gastrectomy compared with use of laparoscopic Roux-en-Y gastric bypass did not meet criteria for equivalence in terms of percentage excess weight loss at 5 years. Although gastric bypass compared with sleeve gastrectomy was associated with greater percentage excess weight loss at 5 years, the difference was not statistically significant, based on the prespecified equivalence margins."<sup>15</sup>

other outcomes, should be the focus of the submitted manuscript, and should be reported in detail. For instance, in a parallel-group trial, results usually include reporting of absolute event rates for the primary outcomes for the study groups, with between-group differences and confidence intervals reflecting the estimated uncertainty around the effect size, and risk ratios, as appropriate. The primary outcome forms the basis for the Discussion section and for the Conclusion of the article.

Clinical trials also include prespecified secondary analyses and outcomes, subgroup analyses, exploratory outcomes, and post hoc analyses. These findings are commonly included in reports of clinical trials but must be presented appropriately, objectively, and in context. Authors should indicate the number of secondary outcomes and how many are being reported in the manuscript. The methods and analyses on which these find-

ings are based should be detailed in the Methods section of the manuscript, and the findings should be reported in the Results section and should be clearly designated according to prespecified definitions and analytic plans; they should be presented in an organized, logical, and consistent fashion. With analyses of outcomes beyond the primary outcome, it is essential that appropriate prespecified analytic strategies are in place to account for type 1 error due to multiple comparisons; interpretation of these outcomes must address the possibility of false discovery and erroneous inferences, and acknowledge that these observations may need to be considered exploratory.

Despite the prominent position of the Abstract in a research article, the inherent space limitations necessitate that the Abstract of the report of an RCT be considered only a presentation of the key points in study design and the key study results.

For clinical trials published in *JAMA*, the Abstract includes a description of the study population, brief information about the interventions being compared, and designation of the primary and secondary outcomes. The Results section of the Abstract highlights the primary outcome, with detailed reporting of its results, including absolute rates for the outcomes. Reporting the results of secondary outcomes is determined on a trial-by-trial basis. For example, reporting findings from a trial that has a limited number of prespecified secondary outcomes (ie, 2 or 3) is usually more straightforward than reporting findings in a trial with multiple secondary outcomes (ie, 10 or 20). In the former, these results may be considered for inclusion in the Abstract. However, in the latter, reporting of only selective results without the context required to interpret these within the totality of the numerous secondary outcomes examined creates challenges and may represent an unbalanced presentation of the study findings. The results of other outcomes, such as exploratory and post hoc analyses, are not included in the Abstracts of clinical trial reports in *JAMA* but are reported in the full article.

For articles published in *JAMA*, the final section of the Abstract is entitled Conclusions and Relevance; it also serves as the Conclusion at the end of the Discussion section. The primary interpretation is a sentence that represents an objective statement of the results of the primary outcome and is usually reported in the PICO format (patient/population, intervention, comparator, outcome). It is based directly on prespecified criteria (such as threshold for declaring statistical significance or for meeting noninferiority criteria).

However, for some clinical trials the editors may conclude that a straightforward PICO statement alone is insuffi-

cient in the overall context of the study results. In these reports, the editors may then suggest that authors include a second sentence in the Conclusion and Relevance section to help ensure that important limitations and other issues are highlighted for readers, or to provide caveats to an interpretation based on significance testing, without overstating the importance of the study. Various examples from RCTs recently published in *JAMA* are included in the **Box**. The content for this additional sentence is always based on information included in the Abstract (usually related to the primary outcome) and as such is necessarily conservative; nuanced interpretation or speculation requires far more information than the Abstract can provide and is the role of the Discussion section of the article.

### Conclusion

The interpretation of the results of RCTs must be understood in the context of the importance of the question, design and conduct of the trial, generalizability, preexisting evidence, actual results and consistency of results, statistical testing, and clinical importance. To fully appreciate and understand the findings of a clinical trial requires careful reading of the entire article, including study methods, results, and the Discussion section. Precise and objective reporting, not only in the Abstract but throughout the report of the trial, is the starting point for proper interpretation, and no single approach in describing the results of RCTs will obviate the need to consider numerous factors in the final interpretation. Ultimately, clinicians, often with patients, need to determine the importance of the findings and the application in clinical care.

### ARTICLE INFORMATION

**Author Affiliations:** Editor (Bauchner); Deputy Editor (Golub); and Executive Editor (Fontanarosa).

**Corresponding Author:** Howard Bauchner, MD, *JAMA*, 330 N Wabash Ave, Chicago, IL 60611 (Howard.Bauchner@jamanetwork.org).

**Conflict of Interest Disclosures:** None reported.

### REFERENCES

- Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. EQUATOR Network website. <http://www.equator-network.org/reporting-guidelines/consort/>. 2010. Accessed July 25, 2019.
- Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012;308(24):2594-2604. doi:10.1001/jama.2012.87802
- Juszczak E, Altman DG, Hopewell S, Schulz K. Reporting of multi-arm parallel-group randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2019;321(16):1610-1620. doi:10.1001/jama.2019.3087
- Lewis RJ, Angus DC. Time for clinicians to embrace their inner Bayesian? reanalysis of results of a clinical trial of extracorporeal membrane oxygenation. *JAMA*. 2018;320(21):2208-2210. doi:10.1001/jama.2018.16916
- McGlothlin AE, Viele K. Bayesian hierarchical models. *JAMA*. 2018;320(22):2365-2366. doi:10.1001/jama.2018.17977
- Ioannidis JPA. The importance of predefined rules and prespecified statistical analyses: do not abandon significance. *JAMA*. 2019;321(21):2067-2068. doi:10.1001/jama.2019.4582
- Goligher EC, Tomlinson G, Hajage D, et al. Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome and posterior probability of mortality benefit in a post hoc Bayesian analysis of a randomized clinical trial. *JAMA*. 2018;320(21):2251-2259. doi:10.1001/jama.2018.14276
- Fontanarosa PB. Editorial matters: guidelines for writing effective editorials. *JAMA*. 2014;311(21):2179-2180. doi:10.1001/jama.2014.6535
- Bluth T, Serpa Neto A, Schultze MJ, Pelosi P, Gama de Abreu M; Writing Committee for the PROBESE Collaborative Group of the PROtective VEntilation Network (PROVENet) for the Clinical Trial Network of the European Society of Anaesthesiology. Effect of intraoperative high positive end-expiratory pressure (PEEP) with recruitment maneuvers vs low PEEP on postoperative pulmonary complications in obese patients: a randomized clinical trial. *JAMA*. 2019;321(23):2292-2305. doi:10.1001/jama.2019.7505
- Williamson JD, Pajewski NM, Auchus AP, et al; SPRINT MIND Investigators for the SPRINT Research Group. Effect of intensive vs standard blood pressure control on probable dementia: a randomized clinical trial. *JAMA*. 2019;321(6):553-561. doi:10.1001/jama.2018.21442
- Packer DL, Mark DB, Robb RA, et al; CABANA Investigators. Effect of catheter ablation vs antiarrhythmic drug therapy on mortality, stroke, bleeding, and cardiac arrest among patients with atrial fibrillation: the CABANA randomized clinical trial. *JAMA*. 2019;321(13):1261-1274. doi:10.1001/jama.2019.0693
- Driver BE, Prekker ME, Klein LR, et al. Effect of use of a bougie vs endotracheal tube and stylet on first-attempt intubation success among patients with difficult airways undergoing emergency intubation: a randomized clinical trial. *JAMA*. 2018;319(21):2179-2189. doi:10.1001/jama.2018.6496
- Thybo KH, Hågi-Pedersen D, Dahl JB, et al. Effect of combination of paracetamol (acetaminophen) and ibuprofen vs either alone on patient-controlled morphine consumption in the first 24 hours after total hip arthroplasty: the PANSALID randomized clinical trial. *JAMA*. 2019;321(6):562-571. doi:10.1001/jama.2018.22039
- Schnitzer TJ, Easton R, Pang S, et al. Effect of tanezumab on joint pain, physical function, and patient global assessment of osteoarthritis among patients with osteoarthritis of the hip or knee: a randomized clinical trial. *JAMA*. 2019;322(1):37-48. doi:10.1001/jama.2019.8044
- Salminen P, Helmiö M, Ovaska J, et al. Effect of laparoscopic sleeve gastrectomy vs laparoscopic Roux-en-Y gastric bypass on weight loss at 5 years among patients with morbid obesity: the SLEEVEPASS randomized clinical trial. *JAMA*. 2018;319(3):241-254. doi:10.1001/jama.2017.20313