

Teddy D. Warner

Evaluating the quality of published studies and their outcomes is much more complex than is typically imagined. Biomedical science has been slow to develop rigorous *uniform standards* for designing, conducting, analyzing, and reporting studies [1–5]. This lack of uniformity makes it difficult or even sometimes impossible for readers to properly assess the validity of empirical findings in the biomedical literature. For example, randomized controlled trials (RCTs) designed to evaluate interventions are often quite inadequate [6]. Beyond the fact that the studies may have been poorly conducted, the results may also be poorly reported. Inadequate reporting of specific randomization processes in studies is associated with highly biased estimates of treatment effects [7]. Thus, without complete and clear study reports, readers, reviewers, and editors cannot judge the validity and usefulness of health research outcomes [6].

Because published research and its reported outcomes may be flawed in various ways [8–15], scientists, practitioners, and other readers should not rely on published findings as credible and valid *simply because they are published*, even in high-level journals. Currently there is very little empirical evidence to support the value of

editorial peer review in ensuring the validity of published studies or of the outcomes they report [16]. Most biomedical reviewers and editors are not formally trained in how to critique and analyze studies, manuscripts, and articles, and thus they often fail to detect serious study flaws. Even if flaws are identified, it is often very difficult for individuals to determine the extent that the flaws should erode the credibility of the research data and their interpretation—certainly, there is no algorithm for translating specific study flaws that are detected into degrees of validity and credibility. Thus, to some real degree, each consumer of a published study is responsible for carefully assessing each study.

Errors in statistical procedures, both simple and complex, compromise the value and interpretation of results [10, 17–22]. Some of the most common of these errors include the following:

1. Focusing on reporting simple statistical significance without indication of the size of observed effects or their practical importance.
2. Use of inappropriate statistical models.
3. Analyzing clustered data with models that do not account for the clustering effect, thus overestimating the size and significance of effects of the primary variables in the model.
4. Conduct of exploratory analyses (i.e., not hypothesis-driven) not clearly described as such.
5. Inappropriate handling of missing data.

---

T.D. Warner (✉)  
Department of Family and Community Medicine,  
University of New Mexico School of Medicine,  
MSC 09 5040, Albuquerque, NM, USA

6. Inferences of causation from nonexperimental data without properly framing the limitations of such inferences.
7. Categorizing continuous data or variables without justification, thus greatly reducing measurement precision and statistical power.
8. Using analysis of covariance to statistically adjust for baseline differences in groups as if that equates the groups at study outset.
9. Interpreting studies showing non-significance in statistical tests, especially with relatively small or biased or unrepresentative samples, as “negative” results (i.e., concluding that no effect actually exists), when such results are properly interpreted only as “inconclusive”.
10. Not reporting study results in practical or clinically meaningful units (e.g., total cohort mortality rate, effort to yield, number needed to treat, minimum clinically important difference).

Most biomedical researchers are not trained to understand or deal with these (and many, many more) important statistical and design issues, despite their immense good intentions and strong abilities. To avoid some of the most common errors in reporting of research, guidelines have been developed, and these represent a valuable resource for academic faculty.

---

## Guidelines for Reporting Empirical Studies

Because of the recognized problems in reporting biomedical research, in 1979 the International Committee of Medical Journal Editors (ICMJE) first published *reporting* guidelines for authors (Table 28.1). These initial guidelines were limited only to formatting issues, but over time the ICJME has provided broader reporting guidelines (*Uniform Requirements for Manuscripts Submitted to Biomedical Journals*; see <http://www.icmje.org>). It is important to note that most ICJME recommendations do not focus on actual standards for the proper design, conduct, analysis, and interpretation of health-related research, only the reporting of the methods that are used. Thus, they are not directly useful in helping to assess the validity of study outcomes that do fully

and clearly report, although conscientious application of the ICJME guidelines definitely helps to reduce the confound between poor reporting and ability to assess study validity.

Many other types of guidelines have also been published in recent years by various organizations (e.g., Cochrane), which do contribute to readers’ ability to properly evaluate how a study was designed, conducted, analyzed, and reported. The prime example is the *CONSORT (Consolidated Standards of Reporting Trials) Statement*, which was originated for readers, researchers, reviewers, and editors almost 20 years ago. CONSORT efforts include a range of initiatives developed to alleviate the problems arising from the inadequate reporting of randomized controlled trials (RCTs). The *CONSORT 2010 Statement* includes a 25-item checklist focused on reporting how the trial was designed, analyzed, and interpreted, plus a flow diagram that shows the movement of all participants through the trial. The CONSORT Statement is an evidence-based, minimum set of recommendations for reporting RCTs to serve as a standard way for researchers to prepare reports of trial outcomes with complete and transparent reporting, enabling readers to assess study validity. The CONSORT Statement evolves with periodic changes as new evidence emerges regarding design, conduct, analysis, and reporting of studies. The CONSORT website (<http://www.consort-statement.org/>) contains the current version of the CONSORT Statement and information on various extensions and explanations of the statement.

The CONSORT Statement is endorsed by over 300 biomedical journals and many leading editorial organizations. CONSORT is part of a broader effort to improve the reporting of health research and to improve the quality of research used in decision-making in healthcare. No practitioner, researcher, reviewer, editor, or professional consumer of the medical literature should attempt to evaluate research outcomes without thorough knowledge of the CONSORT Statement and its related documents. Researchers who follow these guidelines maximize the ability of readers, reviewers, and editors to evaluate the validity of study findings. Evidence from the last decade suggests that use of the CONSORT Statement checklist improves the quality of reporting [23].

**Table 28.1** Major resources for evaluating and reporting studies and study outcomes*Uniform Requirements for Manuscripts Submitted to Biomedical Journals* (<http://www.icmje.org>)

These requirements were developed and first published in 1979 by a small group of medical journal editors, which expanded and evolved into the International Committee of Medical Journal Editors (ICMJE), which now meets annually. The ICMJE has gradually broadened its concerns to include ethical principles related to publication in biomedical journals, and it has produced multiple editions of the Uniform Requirements. Issues have also arisen that go beyond manuscript preparation, resulting in development of a number of Separate Statements on editorial policy. The entire Uniform Requirements document was revised in 1997, and sections were updated in 1999, 2000, and 2001. In 2003, the committee revised and reorganized the entire document and incorporated the Separate Statements. The ICMJE prepared the current revision in 2010. The over 600 journals that agree to use the Uniform Requirements are encouraged to state in their Instructions to Authors that their requirements are in accordance with the Uniform Requirements and to cite the 2010 version. The ICMJE is a small working group of editors of general medical journals and is not an open-membership organization. Open-membership organizations for editors and others in biomedical publication include the World Association of Medical Editors ([www.WAME.org](http://www.WAME.org)), the Council of Science Editors ([www.councilscienceeditors.org/](http://www.councilscienceeditors.org/)), and the European Association of Science Editors ([www.ease.org.uk](http://www.ease.org.uk))

*Cochrane Collaboration* (<http://www.cochrane.org/>)

The Cochrane Collaboration is an international network of more than 28,000 dedicated people from over 100 countries established in 1993. It works to help healthcare providers, policy-makers, patients, patient advocates, and caregivers make well-informed decisions about healthcare, based on the best available research evidence, by preparing, updating, and promoting the accessibility of *Cochrane Reviews*—over 4,600 published so far online in *The Cochrane Library*. The Cochrane Collaboration vision is that healthcare decision-making worldwide should be informed by high-quality, timely research evidence. Its work is internationally recognized as the benchmark for high-quality information about the effectiveness of healthcare. As such its reviews and standards for conducting systematic reviews are essential resources for healthcare providers, consumers, and researchers

*Cochrane Handbook for Systematic Reviews of Interventions* is the official document that describes in detail the process of preparing and maintaining Cochrane systematic reviews on the effects of healthcare interventions. It is available in various formats from the Cochrane website, which has links to many resources related to evaluation of health studies. The Handbook is a detailed resource that also provides considerable information about evaluating clinical studies

*CONSORT (Consolidated Standards of Reporting Trials) Statement* (<http://www.consort-statement.org/>)

The CONSORT Statement and associated documents strive to alleviate the problems that arise from inadequate reporting of randomized controlled trials (RCTs). The CONSORT Statement is an evidence-based set of recommendations for reporting RCTs. It provides a standard way to prepare reports of trial findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and interpretation by practitioners, policy-makers, consumers, and researchers. The CONSORT Statement includes a 25-item checklist and a flow diagram along with some brief descriptive text. The checklist items focus on reporting how the trial was designed, analyzed, and interpreted, and the flow diagram displays the progress of all participants through the trial

If all researchers followed CONSORT and the other published guidelines (see “Additional Resources”), the quality of reporting of studies will likely increase substantially, which in turn will enhance scientific progress.

## Additional Research Reporting Guidelines

During the last decade or so, over 80 reporting guidelines have been developed, covering a broad range of specific study designs and data. Most guidelines were created idiosyncratically because little literature informs guideline developers

about how to develop them, and thus these guidelines themselves may be flawed or incomplete.

To help improve the quality of reporting (and thus evaluating) of health research and its outcomes, the Enhancing the Quality and Transparency Of health Research (EQUATOR) Network was established in 2008 (<http://www.equator-network.org/>). EQUATOR is intended to improve the quality of scientific publications by promoting transparent and accurate reporting through achievement of five major goals:

1. To build a comprehensive *web-based resource center* to develop and maintain up-to-date information, tools, and other materials related to reporting health research, including online

resources for editors and peer reviewers related to teaching scientific writing and reporting.

2. To set up a *network of reporting guideline developers* and to maintain mutual collaboration among them, including providing developers scientific support for guideline development and information about how to best develop reporting guidelines.
3. To *promote reporting guidelines* and their use by developing online training courses for editors, peer reviewers, and researchers in their use, and to promote activities to raise the awareness of the importance of using reporting guidelines.
4. To conduct regular *assessment of how journals implement reporting guidelines*—recent data indicate substantial need for improvement in reporting and using of reporting guidelines.
5. To conduct an *annual audit of reporting quality* across the health literature because most journals do not have an objective means for judging the quality of their published health research, thus providing data on the influence of reporting guidelines on published literature (adapted from <http://www.equator-network.org/>).

Sponsors and researchers such as those engaged in the EQUATOR endeavor see the use of reporting guidelines as an important method for helping to improve the quality of health-related research overall.

Not all studies can be evaluated with the same set of standards (hence, the dozens of reporting guidelines developed or being developed). There are many ways to classify and categorize empirical research, and for this brief chapter, I suggest that biomedical studies be categorized into one of four main types: (1) experimental and quasi-experimental studies; (2) observational studies (i.e., nonexperimental studies, of which there are many subtypes); (3) qualitative studies (of which there are also many subtypes); and (4) literature reviews, which can further be categorized as narrative reviews, systematic reviews, or meta-analyses.

Some general methodological principles apply to all research types (e.g., clear and complete description of the study design, objectives, hypothesis (if any), and main procedures; reliable

measurement of outcome variables; minimization or control of confounding variables), but many issues are unique to a particular type or subtype of research (e.g., randomization for experiments). We briefly list many of the general questions a research study evaluator should ask in the “words to the wise” section at the end of the chapter. Different types of research may require many additional specific questions to enable the full evaluation of a research report.

It has been common to label randomized controlled trials (RCTs) as the “gold standard of research” because RCTs provide stronger direct evidence of cause–effect relationships (i.e., *efficacy*

### Key Concepts

- *Outcomes* are the dependent variables or the effects on the dependent variable in a research study or the results from a study. This is the simple sense in which this term is used in this chapter. However, many definitions of outcomes can be found, and “outcomes research” has evolved to be an area of research itself, which applies to research that is concerned with the effectiveness of public health interventions and health services, that is, the outcomes of these services. Outcomes research may also refer to effectiveness of healthcare delivery, with measures such as cost-effectiveness, health status, and disease burden.
- *Internal validity* refers to the degree to which results of a study can be properly attributed to the variation in the independent or predictor variables rather than to flaws in the research design. In other words, internal validity is the extent to which one can properly draw conclusions about the causal effects of one variable on another variable or in nonexperimental research on the relationship between two or more variables. Internal validity refers to the absence of the effects of confounding or extraneous variables on the relationship between two other variables.

(continued)

(continued)

- *External validity* is a synonym for generalizability, which refers to the degree that results or outcomes from a study can properly be applied to individuals, situations, or settings beyond those studied directly in a research project. A study can have high internal validity but low external validity, but not vice versa.
- The *CONSORT Statement* is a document with an extensive objective checklist of criteria intended to improve the clear and accurate reporting of a randomized controlled trial (RCT), thus enabling readers to understand the design, conduct, analysis, and interpretation of the trial and to evaluate the validity of the trial outcomes and results.

of interventions) than other types of studies. Some research areas have been more amenable to the use of RCTs (e.g., pharmacological trials) than others, and calls for greater use of RCTs have been offered in various areas of clinical research over the years (e.g., surgical treatments, non-pharmacological psychiatric treatments). However, all the other various research designs complement the evidence from RCTs and are often necessary under the many circumstances when the RCTs are ethically inappropriate or highly impractical or even impossible or when the research question is not about the efficacy of an intervention. On the other hand, in conducting RCTs, many efforts are made to maximize treatment compliance that are extraordinarily rigorous and therefore do not fit well with everyday medical practice situations. Consequently, the outcomes of RCTs are not readily translatable into practice. Indeed, the National Institutes of Health in recent years has emphasized the importance of conducting research to determine the degree to which RCTs and other highly controlled studies actually show findings that produce (translate) into meaningful effects in naturally working healthcare systems [24].

Thus, the failure to use randomization or experimental methods in a study is not a fatal

flaw—indeed, many situations and conditions require research evidence other than RCTs. For example, early in the course of studying some phenomena, basic observational or qualitative studies are often required to form some background for designing more complex studies. Later, cohort studies add to what earlier case studies or case series contributed to the knowledge base. Qualitative studies may, in fact, contribute considerably to understanding reasons behind clinician or patient actions that could not easily be revealed in a controlled quantitative study.

In general, a “good study” is one that is designed to answer a properly framed research question and that can be conducted within the limits of the situation and available resources. Recognition of the place of different types of research has important implications for research methodology, for the quality of care in clinical practice, and for research funding policy. Every type of study design has problems in particular applications and if designed, conducted, or analyzed improperly, and thus all studies should be evaluated by the specifically focused criteria. Recognition of just how data from various study types can contribute to the evolving knowledge in an area is important. There is no true single gold standard, and each study should be judged on its strengths, weaknesses, and ability to advance understanding in a field given the current state of knowledge.

---

## Research Evidence Hierarchies

Over the past 30 years or so, various hierarchies of evidence have been proposed and widely used to grade the quality of health research. Use of such hierarchies themselves may be overly reductionist and yield anomalous measures of research quality [25]. Perhaps the major problem with research evaluation hierarchies is that they tend to collapse multiple dimensions of study quality (e.g., design, conduct, sample size, measurement reliability and validity, blinding success, follow-up losses, analysis methods, question relevance, effect sizes detected) into a *single grade or score*. Some study characteristics are more important for some clinical problems, for some outcomes, and for some

study objectives than others. Thus, a summary of the published main dimensions of evidence may be superior and more useful than a graded hierarchy with single overall study quality scores. Such a summary should be accompanied with an evaluation of why specific dimensions of study quality are important in the context being assessed [25]. A study could have high scores on many or most of multiple dimensions but a very low score on a single dimension, which alone may call the validity of the outcomes into question. Thus, average or summative scores should be used only with great caution, if at all, to evaluate studies.

---

## Evidenced-Based Medicine

Biomedical research, evidence-based medicine, systematic reviews, and practice guidelines are part of contemporary medical science and medical practice. *Evidence-based medicine* (EBM) appears to motivate the search for answers to many questions related to the efficacy and effectiveness of healthcare as well as costs of and access to care. Valid scientific evidence is essential in medicine for questions about quality care, healthcare policy-making, and various medical–legal issues. Thus, EBM brings together relevant trustworthy information through acquisition of systematic valid empirical data, the valid analysis and interpretation of such data, and the translation of research findings into clinical practice, health systems management, and healthcare policy. EBM, systematic reviews, meta-analysis, and practice guidelines evolve through sound research methodology that enables valid understanding of the empirical data (outcomes) that can then be effectively applied in clinical settings. EBM is defined as a conscientious, explicit, and judicious use of the current best empirical evidence in making decisions about the care of individual patients or groups of patients.

Evidence-based practice includes recognition of the patient’s problem, construction of an objective clinical question, search of empirical literature to retrieve the best available evidence to answer the question, critical appraisal of all available evidence, and integration of the evidence with all aspects and contexts of the clinical circumstances. *Systematic literature reviews*

provide the application of scientific strategies that limit bias by the systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific topic. Systematic reviews are similar to meta-analyses but are very different from traditional narrative reviews.

*Clinical practice guidelines* are systematically developed statements that are intended to assist physicians and patients in making the best healthcare decisions, given the available empirical evidence. Evidence-based clinical practice guidelines are designed to improve the quality of patient care, patient access to care, treatment appropriateness, efficiency and effectiveness with minimal cost. Well-developed clinical practice guidelines consider the available empirical evidence with multiple dimensions: validity, reliability, clinical applicability, clinical flexibility, clarity for practice, careful means of documentation, all gathered through systematic valid empirical studies that may use various designs. Thus systematic reviews assess research outcomes, and clinical practice guidelines apply scientific outcomes to clinical care practices.

---

## Conclusion

Many guidelines have been developed to help academic faculty in reporting study findings and understanding the adequacy of the study design, conduct, analysis, and interpretation. Dedicated efforts to apply these guidelines will bring benefit to individual health and society at large.

### Words to the Wise

As you evaluate research studies and their outcomes, answer the following questions:

- Are you familiar with the accepted standards for proper design, conduct, analysis, and reporting for the various types of studies (e.g., RCTs, cohort studies, other observational studies, systematic reviews, meta-analyses, qualitative studies) that should be applied to determine the validity and credibility of reported outcomes?

(continued)

(continued)

- Are you familiar with many of the common basic flaws in study design and statistical analysis of biomedical studies reported in the literature?
- Do you understand how various study types (e.g., experiments, quasi-experiments, cohort studies of various kinds, various direct observational studies, epidemiological studies, clinical case reports, qualitative studies) provide valuable evidence?

#### Ask Your Mentor or Colleagues

- How can access to literature be expanded through Internet searches and web resources?
- What journals, websites, and listservs are essential reading?
- What implications do recent published empirical studies have for practice or research?
- What are some important questions that could be answered by research that you are excited about and currently are prepared to conduct?

## References

- Schulz KF, Altman DG, Moher D, The CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. *PLoS Med.* 2010;7(3):e1000251.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147:573–7.
- Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess.* 2003;7:iii-173.
- Daly J, Willis K, Small R, Green J, Welch N, Kealy M, Huges E. A hierarchy of evidence for assessing qualitative health research. *J Clin Epidemiol.* 2007;60:43–9.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Goetzsche PC, et al. The PRISMA Statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 2009;6(7):e1000100.
- Moher D, Simeria I, Schulz KF, Hoey J, Altman DG. Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *Biomed Cent Med.* 2008;6:13.
- Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol.* 2007;36:847–57.
- Altman DG. The scandal of poor medical research: we need less research, better research, and research done for the right reasons. *Br Med J.* 1994;308:283–4.
- Altman DG. Poor-quality medical research: what can journals do? *J Am Med Assoc.* 2002;287(21):2765–7.
- Fernandez-Taylor S, Hyun JK, Reeder RN, Harris AHS. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *Biomed Cent Res Notes.* 2011;4:304.
- Ioannidis JP. Why most discovered published research findings are false. *PLoS Med.* 2005;2:e124.
- Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology.* 2008;19:640–8.
- Ioannidis JP, Karassa F. The need to consider the wider agenda in systematic reviews and meta-analyses. *Br Med J.* 2010;341:762–5.
- Kaul S, Diamond GA. Trial and error. How to avoid commonly encountered limitations of published clinical trials. *J Am Coll Cardiol.* 2010;55:415–27.
- Stone GW, Pocock SJ. Randomized trials, statistics, and clinical inference. *J Am Coll Cardiol.* 2010;55:428–31.
- Jefferson T, Rudin M, Folsie SB, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. The Cochrane Library, The Cochrane Collaboration, 2008, Issue 2. Hoboken, NJ: John Wiley & Sons.
- Harris AHS, Reeder R, Hyun JK. Common statistical and research design problems in manuscripts submitted to high impact psychiatry journals: what editors and reviewers want authors to know. *J Psychiatr Res.* 2009;43:1231–4.
- Lang T. Twenty statistical errors even YOU can find in biomedical research articles. *Croat Med J.* 2004;45(4):361–70.
- Malek MH, Berger DE, Coburn JW. On the inappropriateness of stepwise regression analysis for model building and testing. *Eur J Appl Physiol.* 2007;101:263–4.
- Stasak AM, Aaman Q, Pfeiffer KP, Gobel G, Ulmer H. Statistical errors in medical research – review of common pitfalls. *Swiss Med Wkly.* 2007;137:44–9.
- Schatz P, Jay KA, McComb J, McLaughlin JR. Misuse of statistical tests in Archives of Clinical Neuropsychology publications. *Arch Clin Neuropsychol.* 2005;20:1053–9.
- Ziliak ST, McCloskey DN. The cult of statistical significance: how the standard error costs us jobs, justice, and lives. Ann Arbor, MI: University of Michigan Press; 2008.

23. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust.* 2006;185(5):263–7. Review.
  24. Woolf SH. The meaning of translational research and why it matters. *J Am Med Assoc.* 2008;299(2):211–3.
  25. Glasziou P, Vandenbroucke J, Chalmers I. Assessing the quality of research. *Br Med J.* 2004;328:39–41.
- accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med.* 2003;138:40–4.
- Cohen DJ, Crabtree BF. Evaluative criteria for qualitative research in health care: controversies and recommendations. *Ann Fam Med.* 2008;6(4):331–9.
- Long T, Johnson M. Rigor, reliability and validity in qualitative research. *Clin Eff Nurs.* 2000;4:30–7.
- Malterud K. Qualitative research: standards, challenges, and guidelines. *Lancet.* 2001;358:483–8.
- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, et al. CONSORT 2010 Explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869.
- Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group preferred reporting items for systematic reviews and meta-analyses. *PLoS Med.* 2009;6(7):e1000097.

## Additional Resources

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and