

Review Criteria

ABSTRACT

Following the common IMRaD format for scientific research reports, the authors present review criteria and discuss background information and issues related to the review criteria for each section of a research report.

Introduction. The authors discuss the criteria reviewers should be aware of for establishing the context for the research study: prior literature to introduce and describe the problem statement, the conceptual framework (theory) underlying the problem, the relevance of the research questions, and the justification of their research design and methods.

Method. The authors discuss a variety of methods used to advance knowledge and practice in the health professions, including quantitative research on educational interventions, qualitative observational studies, test and measurement development projects, case reports, expository essays, and quantitative and qualitative research synthesis. As background information for reviewers, the authors discuss how investigators use these and other methods in concert with data-collection instruments, samples of research participants, and data-analysis procedures to address educational, policy, and clinical questions. The authors explain the key role that research methods play in scholarship and the role of the reviewer in judging their quality, details, and richness.

Results. The author describes issues related to reporting statistical analyses in the results, particularly data that do not have many of the properties that were anticipated

when the data analysis was planned. Further, the author discusses the presentation of the body of evidence collected within the study, offering information for reviewers on evaluating the selection and organization of data, the balance between descriptive and inferential statistics, narrative presentation, contextualization of qualitative data, and the use of tables and figures.

Discussion. The authors provide information to enable reviewers to evaluate whether the interpretation of the evidence is adequately discussed and appears reliable, valid, and trustworthy. Further, they discuss how reviewers can weigh interpretations, given the strengths and limitations of the study, and can judge the generalizability and practical significance of conclusions drawn by investigators.

Title, authors, and abstract. The author discusses a reviewer's responsibility in judging the title, authors, and abstract of a manuscript submitted for publication. While this triad orients the reader at the beginning of the review process, only after the manuscript is analyzed thoroughly can these elements be effectively evaluated.

Other. The authors discuss the reviewer's role in evaluating the clarity and effectiveness of a study's written presentation and issues of scientific conduct (plagiarism, proper attribution of ideas and materials, prior publication, conflict of interest, and institutional review board approval).

Acad. Med. 2001;76:922–951.

MANUSCRIPT INTRODUCTION

Problem Statement, Conceptual Framework, and Research Question

William C. McGaghie, Georges Bordage, and Judy A. Shea*

REVIEW CRITERIA

- The introduction builds a logical case and context for the problem statement.
- The problem statement is clear and well articulated.
- The conceptual (theoretical) framework is explicit and justified.
- The research question (research hypothesis where applicable) is clear, concise, and complete.
- The variables being investigated are clearly identified and presented.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Introduction

A scholarly manuscript starts with an Introduction that tells a story. The Introduction orients the reader to the topic of the report, moving from broad concepts to more specific ideas.¹ The Introduction should convince the reader, and all the more the reviewer, that the author has thought the topic through and has developed a tight, “researchable” problem. The Introduction should move logically from the known to the unknown. The actual components of an Introduction (including its length, complexity, and organization) will vary with the type of study being reported, the traditions of the research community or discipline in which it is based, and the style and tradition of the journal receiving the manuscript. It is helpful for the reviewer to evaluate the Introduction by thinking about its overall purpose and its individual components: problem statement, conceptual framework, and research question. Two related articles, “Reference to the Literature” and “Relevance,” follow the present article.

Problem Statement

The Introduction to a research manuscript articulates a *problem statement*. This essential element conveys the issues and context that gave rise to the study. Two examples of problem

statements are: “With the national trend toward more patient care in outpatient settings, the numbers of patients on inpatient wards have declined in many hospitals, contributing to the inadequacy of inpatient wards as the primary setting for teaching students,”² and “The process of professional socialization, regardless of the philosophical approach of the educational program, can be stressful . . . few studies have explored the unique stressors associated with PBL in professional education.”³ These statements help readers anticipate the goals of each study. In the case of the second example, the Introduction ended with the following statement: “The purpose of this qualitative study was to identify stressors perceived by physiotherapy students during their initial unit of study in a problem-based program.” In laying out the issues and context, the Introduction should not contain broad generalizations or sweeping claims that will not be backed up in the paper’s literature review. (See the next article.)

Conceptual Framework

Most research reports cast the problem statement within the context of a *conceptual* or *theoretical* framework.⁴ A description of this framework contributes to a research report in at least two ways because it (1) identifies research variables, and (2) clarifies relationships among the variables. Linked to the problem statement, the conceptual framework “sets the stage” for presentation of the specific research question that drives the investigation being reported. For example, the conceptual framework and research question would be different for a formative evaluation study than for a summative study, even though their variables might be similar.

*Lloyd Lewis, PhD, emeritus professor of the Medical College of Georgia, participated in early meetings of the Task Force and contributed to the earliest draft of this section.

Scholars argue that a conceptual or theoretical framework always underlies a research study, even if the framework is not articulated.⁵ This may seem incongruous, because many research problems originate from practical educational or clinical activities. Questions often arise such as “I wonder why such an event did not [or did] happen?” For example, why didn’t the residents’ test-interpretation skills improve after they were given feedback? There are also occasions when a study is undertaken simply to report or describe an event, e.g., pass rates for women versus men on high-stakes examinations such as the United States Medical Licensing Examination (USMLE) Step 1. Nevertheless, it is usually possible to construct at least a brief theoretical rationale for the study. The rationale in the USMLE example may be, for instance, about gender equity and bias and why these are important issues. Frameworks are usually more elaborate and detailed when the topics that are being studied have long scholarly histories (e.g., cognition, psychometrics) where active researchers traditionally embed their empirical work in well-established theories.

Research Question

A more precise and detailed expression of the problem statement cast as a specific *research question* is usually stated at the end of the Introduction. To illustrate, a recent research report states, “The research addressed three questions. First, do students’ pulmonary physiology concept structures change from random patterns before instruction to coherent, interpretable structures after a focused block of instruction? Second, can an MDS [multidimensional scaling] solution account for a meaningful proportion of variance in medical and veterinary students’ concept structures? Third, do individual differences in the ways in which medical and veterinary students intellectually organize the pulmonary physiology concepts as captured by MDS correlate with course examination achievement?”⁶

Variables

In experimental research, the logic revealed in the Introduction might result in explicitly stated hypotheses that would include specification of dependent and independent variables.⁷ By contrast, much of the research in medical education is not experimental. In such cases it is more typical to state general research questions. For example, “In this [book] section, the meaning of medical competence in the worlds of practicing clinicians is considered through the lens of an ethnographic story. The story is about the evolution of relationships among obstetrical providers and transformations in obstetrical practice in one rural town in California, which I will call ‘Coast Community,’ over the course of a decade.”⁸

For some journals, the main study variables (e.g., medical competence) will be defined in the Introduction. Other journals will place this in the Methods section. Whether specific hypotheses or more general research questions are stated, the reviewer (reader) should be able to anticipate what will be revealed in the Methods.

SUMMARY

The purpose of the Introduction is to construct a logical “story” that will educate the reader about the study that follows. The order of the components may vary, with the problem statement sometimes coming after the conceptual framework, while in other reports the problem statement may appear in the first paragraph to orient the reader about what to expect. However, in all cases the Introduction will engage, educate, and encourage the reader to finish the manuscript.

REFERENCES

1. Zeiger M. *Essentials of Writing Biomedical Research Papers*. 2nd Ed. London, U.K.: McGraw-Hill, 1999.
2. Fincher RME, Case SM, Ripkey DR, Swanson DB. Comparison of ambulatory knowledge of third-year students who learned in ambulatory settings with that of students who learned in inpatient settings. *Acad Med*. 1997;72(10 suppl):S130–S132.
3. Soloman P, Finch E. A qualitative study identifying stressors associated with adapting to problem-based learning. *Teach Learn Med*. 1998;10:58–64.
4. Chalmers AF. *What is This Thing Called Science?* St. Lucia, Qld., Australia: University of Queensland Press, 1982.
5. Hammond KR. Introduction to Brunswikian theory and methods. In: Hammond KR, Wascoe NE (eds). *New Directions for Methodology of Social and Behavioral Sciences*, No. 3: Realizations of Brunswikian Representative Design. San Francisco, CA: Jossey-Bass, 1980.
6. McGaghie WC, McCrimmon DR, Thompson JA, Ravitch MM, Mitchell G. Medical and veterinary students’ structural knowledge of pulmonary physiology concepts. *Acad Med*. 2000;75:362–8.
7. Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education*. 4th ed. New York: McGraw-Hill, 2000.
8. DelVecchio Good M-J. *American Medicine: The Quest for Competence*. Berkeley, CA: University of California Press, 1995.

RESOURCES

- American Psychological Association. *Publication Manual of the American Psychological Association*. 4th ed. Washington, DC: APA, 1994:11–2.
- Creswell JW. *Research Design: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications, 1994:1–16.
- Day RA. *How to Write and Publish a Scientific Paper*. 5th ed. Phoenix, AZ: Oryx Press, 1998:33–35.
- Erlandson DA, Harris EL, Skipper BL, Allen SD. *Doing Naturalistic Inquiry: A Guide to Methods*. Newbury Park, CA: Sage Publications, 1993:42–65.
- Glesne C, Peshkin A. *Becoming Qualitative Researchers: An Introduction*. White Plains, NY: Longman Publishing Group, 1992:13–37.

Reference to the Literature and Documentation

Sonia J. Crandall, Addeane S. Caelleigh, and Ann Steinecke

REVIEW CRITERIA

- The literature review is up-to-date.
- The number of references is appropriate and their selection is judicious.
- The review of the literature is well integrated.
- The references are mainly primary sources.
- Ideas are acknowledged appropriately (scholarly attribution) and accurately.
- The literature is analyzed and critically appraised.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Research questions come from observing phenomena or reading the literature. Regardless of what inspired the research, however, study investigators must thoroughly review existing literature to adequately understand the scope of the issues relevant to their questions. Although systematic reviews of the literature conducted in the social and biomedical sciences, such as those produced by the Cochrane Collaboration (for clinical issues) and the Campbell Collaboration (for areas of social science) may be quite different in terms of the types of evidence provided and the natures of the outcomes, their goals are the same, that is, to present the best evidence to inform research, practice, and policy. These reviews are usually carried out by large teams, which follow strict protocols common to the whole collaboration. Individual researchers also conduct thorough reviews, albeit usually less structured and in-depth. They achieve three key research aims through a thorough analysis of the literature: refinement of their research questions, defense of their research design, and ultimately support of their interpretations of outcomes and conclusions. Thus, in the research report, the reviewer should find a clear demonstration of the literature's contribution to the study and its context.¹

Before discussing the specifics of each of the three aims, it is important to offer some distinctions regarding the research continuum. Where researchers fit along the quantitative–qualitative continuum influences how they use literature within a study, although there are no rigid rules about how to use it. Typically, at the quantitative end of the spectrum, researchers review the bulk of the literature primarily at the beginning of the study in order to establish the theoretical or conceptual framework for the research question or problem. They also use the literature to validate the

use of specific methods, tools, and (statistical) analyses, adding citations in the appropriate sections of the manuscript. At the qualitative end of the spectrum, the researchers weave the relevant literature into all phases of the study and use it to guide the evolution of their thinking as data are gathered, transcribed, excerpted, analyzed, and placed before the reader.² They also use the literature to reframe the problem as the study evolves. Although the distinction is not crystal-clear, the difference between the ends of the continuum might be viewed as the difference between testing theory-driven hypotheses (quantitative) and generating theory-building hypotheses (qualitative).

Researchers all along this continuum use the literature to inform their early development of research interests, problems, and questions and later in the conduct of their research and the interpretation of their findings. A review of relevant literature sets the stage for a study. It provides a logically organized world view of the researcher's question, or of the situation the researcher has observed—what knowledge exists relevant to the research question, how the question or problem has been previously studied (types of designs and methodologic concerns), and the concepts and variables that have been shown to be associated with the problem (question).³ The researcher evaluates previous work “in terms of its relevance to the research question of interest,⁴ and synthesizes what is known, noting relationships that have been well studied and identifying areas for elaboration, questions that remain unanswered, or gaps in understanding.^{1,3,5,6} The researcher documents the history and present status of the study's question or problem. The literature reviewed should not only be current, but also reflect the contributions of salient published and unpublished research, which may be quite dated but play a significant role in the evolution of the research. Regardless of perspective (qualitative, quanti-

tative, or mixed method), the researcher must frame the problem or research questions as precisely as possible from a chronologic and developmental perspective, given the confines of the literature.² For example, when presenting the tenets of adult learning as a basis for program evaluation an author would be remiss if he or she omitted the foundational writings of Knowles,⁷ Houle,⁸ and perhaps Lindeman⁹ from the discussion.

Equally important to using the literature to identify current knowledge is using it to defend and support the study and to inform the design and methods.¹⁰ The researcher interprets and weighs the evidence, presents valid points making connections between the literature and the study design, reasons logically for specific methods, and describes in detail the variables or concepts that will be scrutinized. Through the literature, the researcher provides a map guiding the reader to the conclusion that the current study is important and necessary and the design is appropriate to answer the questions.⁶

Once they have the study outcomes, researchers offer explanations, challenge assumptions, and make recommendations considering the literature used initially to frame the research problem. Authors may apply some of the most salient literature at the end of the manuscript to support their conclusions (fully or partially), refute current knowledge, revise a hypothesis, or reframe the problem.⁵ The authors use literature to bring the reader back to the theory tested (quantitative) or the theory generated (qualitative).

Reviewers must consider the pertinence of the literature and documentation with regard to the three key research aims stated earlier. They should also consider the types of resources cited and the balance of the perspectives discussed within the literature reviewed. When considering the types of resources cited, reviewers should determine whether the references are predominantly general sources (textbooks),⁴ primary sources (research articles written by those who conducted the research),⁴ or secondary sources (articles where a researcher describes the work of others).⁴ References should be predominantly primary sources, whether published or unpublished. Secondary sources are acceptable, and desirable, if primary sources are unavailable or if they provide a review (meta-analysis, for example) of what is known about the research problem. Researchers may use general resources as a basis for describing, for example, a theoretical or methodologic principle, or a statistical procedure.

Researchers may have difficulty finding all of the pertinent literature because it may not be published (dissertations), and not all published literature is indexed in electronic databases. Manual searching is still necessary. Reviewers are cautioned to look for references that appear inclusive of the whole body of existing literature. For example, some relevant articles are not indexed in Medline, but are indexed in

ERIC. Reviewers can tell whether multiple databases were searched for relevant literature by the breadth of disciplines represented by the citations. Thus, it is important that the researcher describe how he or she found the previous work used to study his or her problem.¹¹

A caveat for reviewers is to be wary of researchers who have not carried out a thorough review of the literature. They may report that there is a paucity of research in their area when in fact plenty exists. At times, authors must be pushed. At the very minimum, reviewers should comment on whether the researchers described to the reviewers' satisfaction how they found study-related literature and the criteria used to select the sources that were discussed. Reviewers must decide whether this process was satisfactorily described. If only published reports found in electronic databases are discussed, then the viewpoint presented "may be biased toward well-known research" that presents only statistically significant outcomes.¹

When considering the perspectives presented by the author, reviewers should pay attention to whether the discussion presents all views that exist in the literature base, that is, conflicting, consensus, or controversial opinions.^{5,12} The thoroughness of the discussion also depends upon the author's explanation of how literature was located and chosen for inclusion. For example, Bland and colleagues¹³ have provided an excellent example of how the process of location and selection was accomplished.

The mechanics of citing references are covered in "Presentation and Documentation" later in this chapter.

REFERENCES

1. Haller KB. Conducting a literature review. *MCN: Am Maternal Child Nurs.* 1988;13:148.
2. Haller EJ, Kleine PF. Teacher empowerment and qualitative research. In: Haller EJ, Kleine PF (eds). *Using Educational Research: A School Administrator's Guide.* New York: Addison Wesley Longman, 2000: 193-237.
3. Rodgers J, Smith T, Chick N, Crisp J. Publishing workshops number 4. Preparing a manuscript: reviewing literature. *Nursing Praxis in New Zealand.* 1997;12:38-42.
4. Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education.* 4th ed. Boston, MA: McGraw-Hill Higher Education, 2000.
5. Martin PA. Writing a useful literature review for a quantitative research project. *Appl Nurs Res.* 1997;10:159-62.
6. Bartz C. It all starts with an idea. *Alzheimer Dis and Assoc Dis.* 1999; 13:S106-S110.
7. Knowles MS. *The Modern Practice of Adult Education: From Pedagogy to Andragogy.* Chicago, IL: Association Press, 1980.
8. Houle CO. *The Inquiring Mind.* Madison, WI: University of Wisconsin Press, 1961.
9. Lindeman EC. *The Meaning of Adult Education.* Norman, OK: University of Oklahoma Research Center for Continuing Professional and Higher Education, 1989 [Orig. pub. 1926].

10. Glesne C, Peshkin A. *Becoming Qualitative Researchers: An Introduction*. White Plains, NY: Longman Publishing Group, 1992.
11. Smith AJ, Goodman NW. The hypertensive response to intubation: do researchers acknowledge previous work? *Can J Anaesth*. 1997;44:9–13.
12. Bruette V, Fitzig C. The literature review. *J NY State Nurs Assoc*. 1993; 24:14–5.
13. Bland CJ, Meurer LN, Maldonado G. Determinants of primary care specialty choice: a non-statistical meta-analysis of the literature. *Acad Med*. 1995;70:620–41.

RESOURCES

- Bartz C. It all starts with an idea. *Alzheimer Dis and Assoc Dis*. 1999;13: S106–S110.
- Best Evidence in Medical Education (BEME). (<http://www.mailbase.ac.uk/lists/beme>). Accessed 3/30/01.
- Bland CJ, Meurer LN, Maldonado G. A systematic approach to conducting a non-statistical meta-analysis of research literature. *Acad Med*. 1995;70: 642–53.
- The Campbell Collaboration. (<http://campbell.gse.upenn.edu>). Accessed 3/30/01.

- The Cochrane Collaboration. (<http://www.cochrane.org>). Accessed 3/30/01.
- Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med*. 1997;126:376–80.
- Cooper HM. *Synthesizing Research. A Guide for Literature Reviews*. 3rd ed. Thousand Oaks, CA: Sage, 1998.
- Fraenkel JR, Wallen NE. Reviewing the literature. In: *How to Design and Evaluate Research in Education*. 4th ed. Boston, MA: McGraw-Hill Higher Education, 2000;70–101.
- Gall JP, Gall MD, Borg WR. *Applying Educational Research: A Practical Guide*. 4th ed. White Plains, NY: Longman Publishing Group, 1998: chapters 2, 3, 4.
- Mulrow CD. Rationale for systematic reviews. *BMJ*. 1994;309:597–9.

(Although the following Web sites are learning resources for evidence-based research and practice, the information is applicable across research disciplines.)

- Middlesex University. *Teaching/Learning Resources for Evidence Based Practice*. (<http://www.mdx.ac.uk/www/rctsh/ebp/main.htm>). Accessed 3/30/01.
- Centres for Health Evidence. *Users' Guides to Evidence-Based Practice*. (http://www.cche.net/principles/content_all.asp). Accessed 3/30/01.

Relevance

Louis Pangaro and William C. McGaghie

REVIEW CRITERIA

- The study is relevant to the mission of the journal or its audience.
- The study addresses important problems or issues; the study is worth doing.
- The study adds to the literature already available on the subject.
- The study has generalizability because of the selection of subjects, setting, and educational intervention or materials.

ISSUES AND EXAMPLES RELATED TO CRITERIA

An important consideration for editors in deciding whether to publish an article is its relevance to the community (or usually, communities) the journal serves. Relevance has several connotations and all are judged with reference to a specific group of professionals and to the tasks of that group. Indeed, one thing is often spoken of as being “relevant to” something else, and that something is the necessary context that establishes relevance.

First, editors and reviewers must gauge the applicability of the manuscript to problems within the journal's focus; the more common or important the problem addressed by an article is to those involved in it, the more relevant it is. The essential issue is whether a rigorous answer to this study's

question will affect what readers will do in their daily work, for example, or what researchers will do in their next study, or even what policymakers may decide. This can be true even if a study is “negative,” that is, does not confirm the hypothesis at hand. For studies without hypotheses (for instance, a systematic review of prior research or a meta-analysis), the same question applies: Does this review achieve a synthesis that will directly affect what readers do?

Second, a manuscript, especially one involving qualitative research, may be pertinent to the community by virtue of its contribution to theory building, generation of new hypotheses, or development of methods. In this sense, the manuscript introduces, refines, or critiques issues that, for example, underlie the teaching and practice of medicine, such as cognitive psychology, ethics, and epistemology. Thus

a study may be quite relevant even though its immediate, practical application is not worked out.

Third, each manuscript must be judged with respect to its appropriateness to the mission of the specific journal. Reviewers should consider these three elements of relevance irrespective of the merit or quality of an article.

The relevance of an article is often most immediately apparent in the first paragraphs of a manuscript, especially in how the research question or problem posed by the paper is framed. As discussed earlier in "Problem Statement, Conceptual Framework, and Research Question," an effective article explicitly states the issue to be addressed, in the form of either a question to be answered or a controversy to be resolved. A conceptual or theoretical framework underlies a research question, and a manuscript is stronger when this framework is made explicit. An explicit presentation of the conceptual framework helps the reviewer and makes the study's importance or relevance more clear.

The relevance of a research manuscript may be gauged by its purpose or the intention of the study, and a vocabulary drawn from clinical research is quite applicable here. Feinstein classifies research according to its "architecture," the effort to create and evaluate research structures that have both "the reproducible documentation of science and the elegant design of art."¹ Descriptive research provides collections of data that characterize a problem or provide information; no comparisons are inherent in the study design, and the observations may be used for policy decisions or to prepare future, more rigorous studies. Many papers in social science journals, including those in health professions education, derive their relevance from such an approach. In cause-effect research, specific comparisons (for instance, to the subjects' baseline status or to a separate control group) are made to reach conclusions about the efficacy or impact of an intervention (for instance, a new public health campaign or an innovative curriculum). The relevance of such research architecture derives from its power to establish the causality, or at least the strong effects, from innovations. In research that deals with process issues, as defined by Feinstein, the products of a new process or the performance of a particular procedure (for instance, a new tool for the assessment of clinical competence) are studied as an indication of the quality or value of the process or procedure. In this case relevance is not from a cause-and-effect relationship but from a new measurement tool that could be applied to a wide variety of educational settings.^{1,p.15-16}

The relevance of a topic is related to, but is not the same as, the feasibility of answering a research question. Feasibility is related to study design and deals with whether and how we can get an answer. Relevance more directly addresses whether the question is significant enough to be worth asking.² The relevance of a manuscript is more complex than

that of the topic per se, and the relevance includes the importance of the topic as well as whether the execution of the study or of the discussion is powerful enough to affect what others in the field think or do.

Relevance is, at times, a dichotomous, or "yes-no," decision; but often it is a matter of degree, as illustrated by the criteria. In this more common circumstance, relevance is a summary conclusion rather than a simple observation. It is a judgment supported by the applicability of the principles, methods, instruments, and findings that together determine the weight of the relevance. Given a limited amount of space in each issue of a journal, editors have to choose among competing manuscripts, and relevance is one way of summarizing the importance of a manuscript's subject, thesis, and conclusions to the journal's readership.

Certain characteristics or strengths can establish a manuscript's relevance: Would a large part of the journal's community—or parts of several of its overlapping communities—consider the paper worth reading? Is it important that this paper be published even though the journal can publish only a fixed percentage of the manuscripts it receives each year? As part of their recommendation to the editor (see Chapter 3), reviewers are asked to rate how important a manuscript is: extremely, very, moderately, slightly, or not important. Issues that may influence reviewers and editors to judge a paper to be relevant include:

1. Irrespective of a paper's methods or study design, the topic at hand would be considered common and/or serious by the readership. As stated before, relevance is a summary judgment and not infallible. One study of *clinical* research papers showed that readers did not always agree with reviewers on the relevance of studies to their own practice.³ Editors of medical education research journals, for example, must carefully choose to include the perspective of educational practitioners in their judgment of relevance, and try to reflect the concerns of these readers.

2. Irrespective of immediate and practical application, the author(s) provides important insights for understanding theory, or the paper suggests innovations that have the potential to advance the field. In this respect, a journal leads its readership and does not simply reflect it. The field of clinical medicine is filled with examples of great innovations, such as the initial description of radioimmunoassay or the citric acid cycle by Krebs, that were initially rejected for publication.⁴ To use medical education as the example again, specific evaluation methods, such as using actors to simulate patients, gradually pervaded undergraduate medical education but initially might have seemed unfeasible.⁵

3. The methods or conclusions described in the paper are applicable in a wide variety of settings.

In summary, relevance is a necessary but not sufficient

criterion for the selection of articles to publish in journals. The rigorous study of a trivial problem, or one already well studied, would not earn pages in a journal that must deal with competing submissions. Reviewers and editors must decide whether the question asked is worth answering at all, whether its solution will contribute, immediately or in the longer term, to the work of medical education and, finally, whether the manuscript at hand will be applicable to the journal's readership.

REFERENCES

1. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia, PA: W. B. Saunders, 1985;4.
2. Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Ed-*

ucation. 4th ed. New York: McGraw-Hill Higher Education, 2000:30-7.

3. Justice AC, Berlin JA, Fletcher SW, Fletcher RH, Goodman SN. Do readers and peer reviewers agree on manuscript quality? *JAMA*. 1994; 272:117-9.
4. Horrobin DF. The philosophical basis of peer review and the suppression of innovation. *JAMA*. 1990;263:1438-41.
5. Barrows HS. Simulated patients in medical teaching. *Can Med Assoc J*. 1968;98:674-6.

RESOURCES

Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia, PA: W. B. Saunders, 1985.

Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education*. 4th ed. New York: McGraw-Hill Higher Education, 2000.

Fincher RM (ed). *Guidebook for Clerkship Directors*. Washington, DC: Association of American Medical Colleges, 2000.

METHOD

Research Design

William C. McGaghie, Georges Bordage, Sonia Crandall, and Louis Pangaro

REVIEW CRITERIA

- The research design is defined and clearly described, and is sufficiently detailed to permit the study to be replicated.
- The design is appropriate (optimal) for the research question.
- The design has internal validity; potential confounding variables or biases are addressed.
- The design has external validity, including subjects, settings, and conditions.
- The design allows for unexpected outcomes or events to occur.
- The design and conduct of the study are plausible.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Research design has three key purposes: (1) to provide answers to research questions, and (2) to provide a road map for conducting a study using a planned and deliberate approach that (3) controls or explains quantitative variation or organizes qualitative observations.¹ The design helps the investigator focus on the research question(s) and plan an orderly approach to the collection, analysis, and interpretation of data that address the question.

Research designs have features that range on a continuum from controlled laboratory investigations to observational studies. The continuum is seamless, not sharply segmented,

going from structured and formal to evolving and flexible. A simplistic distinction between quantitative and qualitative inquiry does not work because research excellence in many areas of inquiry often involves the best of both. The basic issues are: (1) Given a research question, what are the best research design options? (2) Once a design is selected and implemented, how is its use justified in terms of its strengths and limits in a specific research context?

Reviewers should take into account key features of research design when evaluating research manuscripts. The key features vary in different sciences, of course, and reviewers, as experts, will know the ones for their fields. Here the example is from the various social sciences that conduct re-

search into human behavior, including medical education research. The key features for such studies are stated below as a series of five general questions addressing the following topics: appropriateness of the design, internal validity, external validity, unexpected outcomes, and plausibility.

Is the research design appropriate (or as optimal as possible) for the research question? The matter of congruence, or “fit,” is at issue because most research in medical education is descriptive, comparative, or correlational, or addresses new developments (e.g., creation of measurement scales, manipulation of scoring rules, and empirical demonstrations such as concept mapping^{2,3}).

Scholars have presented many different ways of classifying or categorizing research designs. For example, Fraenkel and Wallen⁴ have recently identified seven general research methods in education: experimental, correlational, causal-comparative, survey, content analysis, qualitative, and historical. Their classification illustrates some of the overlap (sometimes confusion) that can exist among design, data-collection strategies, and data analysis. One could use an experimental design and then collect data using an open-ended survey and analyze the written answers using a content analysis. Each method or design category can be subdivided further. Rigorous attention to design details encourages an investigator to focus the research method on the research question, which brings precision and clarity to a study. To illustrate, Fraenkel and Wallen⁴ break down experimental research into four subcategories: weak experimental designs, true experimental designs, quasi-experimental designs, and factorial designs. Medical education research reports should clearly articulate the link between research question and research design and should embed that description in citations to the methodologic literature to demonstrate awareness of fine points.

Does the research have internal validity (i.e., integrity) to address the question rigorously? This calls for attention to a potentially long list of sources of bias or confounding variables, including selection bias, attrition of subjects or participants, intervention bias, strength of interventions (if any), measurement bias, reactive effects, study management, and many more.

Does the research have external validity? Are its results generalizable to subjects, settings, and conditions beyond the research situation? This is frequently (but not exclusively) a matter of sampling subjects, settings, and conditions as deliberate features of the research design.

Does the research design permit unexpected outcomes or events to occur? Are allowances made for expression of surprise results the investigator did not consider or could not anticipate? Any research design too rigid to accommodate the un-

expected may not properly reflect real-world conditions or may stifle the expression of the true phenomenon studied.

Is the research design implausible, given the research question, the intellectual context of the study, and the practical circumstances where the study is conducted? Common flaws in research design include failure to randomize correctly in a controlled trial, small sample sizes resulting in low statistical power, brief or weak experimental interventions, and missing or inappropriate comparison (control) groups. Signals of research implausibility include an author's failure to describe the research design in detail, failure to acknowledge context effects on research procedures and outcomes, and the presence of features of a study that appear unbelievable, e.g., perfect response rates, flawless data. Often there are tradeoffs in research between theory and pragmatics, precision and richness, elegance and application. Is the research design attentive to such compromises?

Kenneth Hammond explains the bridge between design and conceptual framework, or theory:

Every method, however, implies a methodology, expressed or not; every methodology implies a theory, expressed or not. If one chooses not to examine the methodological base of [one's] work, then one chooses not to examine the theoretical context of that work, and thus becomes an unwitting technician at the mercy of implicit theories.¹

REFERENCES

1. Hammond KR. Introduction to Brunswikian theory and methods. In: Hammond KR, Wascoe NE (eds). *New Directions for Methodology of Social and Behavioral Sciences*, No. 3: Realizations of Brunswik's Representative Design. San Francisco, CA: Jossey-Bass, 1980:2.
2. McGaghie WC, McCrimmon DR, Mitchell G, Thompson JA, Ravitch MM. Quantitative concept mapping in pulmonary physiology: comparison of student and faculty knowledge structures. *Am J Physiol: Adv Physiol Educ.* 2000;23:72-81.
3. West DC, Pomeroy JR, Park JK, Gerstenberger EA, Sandoval J. Critical thinking in graduate medical education: a role for concept mapping assessment? *JAMA.* 2000;284:1105-10.
4. Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education*. 4th ed. New York: McGraw-Hill, 2000.

RESOURCES

- Campbell DT, Stanley JC. *Experimental and Quasi-experimental Designs for Research*. Boston, MA: Houghton Mifflin, 1981.
- Cook TD, Campbell DT. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally, 1979.
- Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*. 3rd ed. Baltimore, MD: Williams & Wilkins, 1996.
- Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston, MA: Little, Brown, 1987.
- Kazdin AE (ed). *Methodological Issues and Strategies in Clinical Research*. Washington, DC: American Psychological Association, 1992.
- Patton MQ. *Qualitative Evaluation and Research Methods*. 2nd ed. Newbury Park, CA: Sage, 1990.

Instrumentation, Data Collection, and Quality Control

Judy A. Shea, William C. McGaghie, and Louis Pangaro

REVIEW CRITERIA

- The development and content of the instrument are sufficiently described or referenced, and are sufficiently detailed to permit the study to be replicated.
- The measurement instrument is appropriate given the study's variables; the scoring method is clearly defined.
- The psychometric properties and procedures are clearly presented and appropriate.
- The data set is sufficiently described or referenced.
- Observers or raters were sufficiently trained.
- Data quality control is described and adequate.

ISSUES AND EXAMPLES RELATED TO CRITERIA

Instrumentation refers to the selection or development and the later use of tools to make observations about variables in a research study. The observations are collected, recorded, and used as primary data.

In the social and behavioral sciences—covering health outcomes, medical education, and patient education research, for example—these instruments are usually “paper-and-pencil” tools. In contrast, the biological sciences and physical sciences usually rely on tools such as microscopes, CAT scans, and many other laboratory technologies. Yet the goals and process in developing and using instruments are the same across the sciences, and therefore each field has appropriate criteria within the overall standards of scientific research. Throughout this section, the focus and examples are from the social sciences and in particular from health professions research, although the general principles of the criteria apply across the sciences.

Instrumentation builds on the study design and problem statement and assumes that both are appropriately specified. In considering the quality of instrumentation and data collection, the reviewer should focus on the rigor with which data collection is executed. Reviewers are looking for or evaluating four aspects of the execution: (1) selecting or developing the instrument, (2) creating scores from the data captured by the instrument, (3) using the instrument appropriately, and (4) a sense that the methods employed met at least minimum quality standards.

Selection and Development

Describing the instrumentation starts with specifying in what way(s) the variables will be captured or measured. The reviewer needs to know what was studied and how the data were collected. There are many means an author can choose. A broad definition is used here that includes, but is not limited to, a wide variety of tools such as tests and examinations, attitude measures, checklists, surveys, abstraction forms, interview schedules, and rating forms. Indeed, scholars recommend that investigators use multiple measures to address the same research construct, a process called triangulation.¹ Instrumentation is often relatively direct because existing and well-known tools are used to capture a variable of interest (e.g., Medical College Admission Test [MCAT] for medical school “readiness” or “aptitude”; National Board of Medical Examiners [NBME] subject examinations for “acquisition of medical knowledge”; Association of American Medical Colleges [AAMC] Graduation Questionnaire for “curricular experiences”). But sometimes the process is less straightforward. For example, if clinical competence of medical students after a required core clerkship is the variable of interest, it may be measured from a variety of perspectives. One approach is to use direct observations of students performing a clinical task, perhaps with standardized patients. Another approach is to use a written test to ask them what they would do in hypothetical situations. Another option is to collect ratings made by clerkship directors at the end of the clerkship that attest to students' clinical skills. Other

alternatives are peer- and self-ratings of competence. Or patient satisfaction data could be collected. Choosing among several possible measures of a variable is a key decision when planning a research study.

Often a suitable measurement instrument is not available, and instruments must be developed. Typically, when new instruments are used for research, more detail about their development is expected than when existing measures are employed. Reviewers do not have to be experts in instrument development, but they need to be able to assess that the authors did the right things. Numerous publications describe the methods that should be followed in developing academic achievement tests,^{2,3} rating and attitude scales,^{4,6} checklists,⁷ and surveys.⁸ There is no single best approach to instrument development, but the process should be described rigorously and in detail, and reviewers should look for citations provided for readers to access this information.

Instrument development starts with specifying the content domain, conducting a thorough review of past work to see what exists, and, if necessary, beginning to create a new instrument. If an existing instrument is used, the reviewer needs to know and learn from the manuscript the rationale and original sources. When new items are developed, the content can be drawn from many sources such as potential subjects, other instruments, the literature, and experts. What the reviewer needs to see is that the process followed was more rigorous than a single investigator (or two) simply putting thoughts on paper. The reviewers should make sure that the items were critically reviewed for their clarity and meaning, and that the instrument was pilot tested and revised, as necessary. For some instruments, such as a data abstraction form, pilot testing might mean as little as trying out the form on a sample of hospital charts. More stringent testing is needed for instruments that are administered to individuals.

Creating Scores

For any given instrument, the reviewer needs to be able to discern how scores or classifications are derived from the instrument. For example, how were questionnaire responses summed or dichotomized such that respondents were grouped into those who “agreed” and “disagreed” or those who were judged to be “competent” and “not competent”? If a manuscript is *about* an instrument, as opposed to the more typical case, when authors *use* an instrument to assess some question, investigators might present methods for formal scale development and evaluation, often focusing on subscale definition, reliability estimation, reproducibility, and homogeneity.⁹ Large development projects for instruments designed to measure individual differences on a variable of interest will also need to pay attention to validity issues, sensitivity, and stability of scores.¹⁰ Other types of instruments do not lend themselves well to aggregated

scores. Nevertheless, reviewers need to be clear about how investigators operationalized research variables and judged the technical properties (i.e., reliability and validity) of research data.

Decisions made about cut-scores and classifications also need to be conveyed to readers. For example, in a study on the perceived frequency of feedback from preceptors and residents to students, the definition of “feedback” needs to be reported and justified. For example, is it a report of *any* feedback in a certain amount of time, or is it feedback at a higher frequency, maybe more than twice a day? Investigators make many decisions in the course of conducting a study. Not all need to be reported in a paper but enough should be present to allow readers to understand the operationalization of the variables of interest.

This discussion of score creation applies equally when the source of data is an existing data set, such as the AAMC Faculty Roster or the AMA Master File. These types of data raise yet more issues about justification of analytic decisions. A focus of these manuscripts should be how data were selected, cleaned, and manipulated. For example, if the AMA Master File is being used for a study on primary care providers, how exactly was the sample defined? Was it by training, board certification, or self-reports of how respondents spent their professional time? Does it include research and administrative as well as clinical time? Does it include both family medicine and internal medicine physicians? When researchers do secondary data analyses, they lose intimate knowledge of the database and yet must provide information. The reviewer must look for evidence of sound decisions about sample definition and treatment of missing data that preceded the definition of scores.

Use of the Instrument

Designing an instrument and selecting and scoring it are only two parts of instrumentation. The third and complementary part involves the steps taken to ensure that the instrument is used properly. For many self-administered forms, the important information may concern incentives and processes used to gather complete data (e.g., contact of non-responders, location of missing charts). For instruments that may be more reactive to the person using the forms (e.g., rating forms, interviews), it is necessary to summarize coherently the actions that were taken to minimize differences related to the instrument user. This typically involves discussions of rater or interviewer training and computation of inter- or intra-rater reliability coefficients.⁵

General Quality Control

In addition to reviewing the details about the actual instruments used in the study, reviewers need to gain a sense that

a study was conducted soundly.¹¹ In most cases, it is impossible and unnecessary to report internal methods that were put in place for monitoring data collection and quality. This level of detail might be expected for a proposal application, but it does not fit in most manuscripts. Still, depending on the methods of the study under review, the reviewer must assess a variety of issues such as unbiased recruitment and retention of subjects, appropriate training of data collectors, and sensible and sequential definitions of analytic variables. The source of any funding must also be reported.

These are generic concerns for any study. It would be too unwieldy to consider here all possible elements, but the reviewer needs to be convinced that the methods are sound—sloppiness or incompleteness in reporting (or worse) should raise a red flag. In the end the reviewer must be convinced that appropriate rigor was used in selecting, developing, and using measurement tools for the study. Without being an expert in measurement, the reviewer can look for relevant details about instrument selection and subsequent score development. Optimally the reviewer would be left confident and clear about the procedures that the author followed in developing and implementing data collection tools.

REFERENCES

1. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959;56:81–105.

2. Linn RL, Gronlund NE. *Measurement and Assessment in Teaching.* 7th ed. Englewood Cliffs, NJ: Prentice-Hall, 1995.
3. Millman J, Green J. The specification and development of tests of achievement and ability. In: Linn RL (ed). *Educational Measurement.* 3rd ed. New York: McMillan, 1989:335–66.
4. Medical Outcomes Trust. Instrument review criteria. *Med Outcomes Trust Bull.* 1995;2:1–IV.
5. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* 2nd ed. Oxford, U.K.: Oxford University Press, 1995.
6. DeVellis RF. *Scale Development: Theory and Applications.* Applied Social Research Methods Series, Vol. 26. Newbury Park, CA: Sage, 1991.
7. McGaghie WC, Renner BR, Kowlowitz V, et al. Development and evaluation of musculoskeletal performance measures for an objective structured clinical examination. *Teach Learn Med.* 1994;6:59–63.
8. Woodward CA. Questionnaire construction and question writing for research in medical education. *Med Educ.* 1998;22:347–63.
9. Kerlinger FN. *Foundations of Behavioral Research.* 3rd ed. New York: Holt, Rinehart and Winston, 1986.
10. Nunnally JC. *Psychometric Theory.* New York: McGraw-Hill, 1978.
11. McGaghie WC. Conducting a research study. In: McGaghie WC, Frey JJ (eds). *Handbook for the Academic Physician.* New York: Springer Verlag, 1986:217–33.

RESOURCES

- Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education.* 3rd ed. New York: McGraw-Hill, 1996.
- Linn RL, Gronlund NE. *Measurement and Assessment in Teaching.* 8th ed. Englewood Cliffs, NJ: Merrill, 2000.

Population and Sample

William C. McGaghie and Sonia Crandall*

REVIEW CRITERIA

- The population is defined clearly, for both subjects (participants) and stimulus (intervention), and is sufficiently described to permit the study to be replicated.
- The sampling procedures are sufficiently described.
- Subject samples are appropriate to the research question.
- Stimulus samples are appropriate to the research question.
- Selection bias is addressed.

ISSUES AND EXAMPLES RELATED TO CRITERIA

Investigators in health outcomes, public health, medical education, clinical practice, and many other domains of scholarship and science are expected to describe the research population(s), sampling procedures, and research sample(s) for the empirical studies they undertake. These descriptions must be clear and complete to allow reviewers and research consumers to decide whether the research results are valid internally and can be generalized externally to other research samples, settings, and conditions. Given necessary and sufficient information, reviewers and consumers can judge whether an investigator's population, sampling methods, and research sample are appropriate to the research question.

Sampling from populations has become a key issue in 20th and 21st century applied research. Sampling from populations addresses research efficiency and accuracy. To illustrate, the Gallup Organization achieves highly accurate (± 3 percentage points) estimates about opinions of the U.S. population (280 million) using samples of approximately 1,200 individuals.¹

Sampling from research populations goes in at least two dimensions: from subjects or participants (e.g., North American medical students), and from stimuli or conditions (e.g., clinical problems or cases). Some investigators employ a third approach—matrix sampling—to address research subjects and stimuli simultaneously.² In all cases, however, reviewers should find that the subject and stimulus populations and the sampling procedures are defined and described clearly.

Given a population of interest (e.g., North American medical students), how does an investigator define a population subset (sample) for the practical matter of conducting a research study? Textbooks provide detailed, scholarly descriptions of purist sampling procedures^{3,4} Other scholars, however, offer practical guides. For example, Fraenkel and Wallen⁵ identify five sampling methods that a researcher may use to draw a representative subset from a population of interest. The five sampling methods are: random, simple, systematic, stratified random, and cluster.

Experienced reviewers know that most research in medical education involves convenience samples of students, residents, curricula, community practitioners, or other units of analysis. Generalizing the results of studies done on convenience samples of research participants or other units is risky unless there is a close match between research subjects and the target population where research results are applied. In some areas, such as clinical studies, the match is crucial, and there are many excellent guides (for example, see Fletcher, Fletcher and Wagner⁶). Sometimes research is deliberately done on "significant"⁷ or specifically selected samples, such as Nobel Laureates or astronauts and cosmonauts,⁸ where descriptions of particular subjects, not generalization to a subject population, is the scholarly goal.

Once a research sample is identified and drawn, its members may be assigned to study conditions (e.g., treatment and control groups in the case of intervention research). By contrast, measurements are obtained uniformly from a research sample for single-group observational studies looking at statistical correlations among variables. Qualitative observational studies of intact groups such as the surgery residents described in *Forgive and Remember*⁹ and the internal medicine residents in *Getting Rid of Patients*¹⁰ follow a similar approach but use words, not numbers, to describe their research samples.

*Lloyd Lewis, PhD, emeritus professor of the Medical College of Georgia, participated in early meetings of the Task Force and contributed to the earliest draft of this section.

Systematic sampling of subjects or other units of analysis from a population of interest allows an investigator to generalize research results beyond the information obtained from the sample values. The same logic holds for the stimuli or independent variables involved in a research enterprise (e.g., clinical cases and their features in problem-solving research). Careful attention to stimulus sampling is the cornerstone of representative research.¹¹⁻¹³

An example may make the issue clearer. (The specifics here are from medical education and are directly applicable to health professions education and generally applicable to wide areas of social sciences.) Medical learners and practitioners are expected to solve clinical problems of varied degrees of complexity as one indicator of their clinical competence. However, neither the population of eligible problems nor clear-cut rules for sampling clinical problems from the parent population have been made plain. Thus the problems, often expressed as cases, used to evaluate medical personnel are chosen haphazardly. This probably contributes to the frequently cited finding of case specificity (i.e., non-generalizability) of performance in research on medical problem solving.¹⁴ An alternative hypothesis is that case specificity has more to do with how the cases are selected or designed than with the problem-solving skill of physicians in training or practice.

Recent work on construction of examinations of academic achievement in general^{15,16} and medical licensure examinations in particular¹⁷ is giving direct attention to stimulus sampling and representative design. Conceptual work in the field of facet theory and design¹⁸ also holds promise as an organizing framework for research that takes stimulus sampling seriously.

Research protocols that make provisions for systematic, simultaneous sampling of subjects and stimuli use matrix sampling.² Matrix sampling is especially useful when an investigator aims to judge the effects of an overall program on a broad spectrum of participants.

Isolating and ruling out sources of bias is a persistent problem when identifying research samples. Subject-selection bias is more likely to occur when investigators fail to specify and use explicit inclusion and exclusion criteria; when there is differential attrition (drop out) of subjects from study conditions; or when samples are insufficient (too small) to give

a valid estimate of population parameters and have low statistical power. Reviewers must be attentive to these potential flaws. Research reports should also describe use of incentives, compensation for participation, and whether the research participants are volunteers.

REFERENCES

1. Gallup Opinion Index. Characteristics of the Sample. Princeton, NJ: Gallup Organization, 1999.
2. Sirotnik KA. Introduction to matrix sampling for the practitioner. In: Popham WJ (ed). *Evaluation in Education: Current Applications*. Berkeley, CA: McCutchan, 1974.
3. Henry GT. Practical sampling. In: *Applied Social Research Methods Series*, Vol. 21. Newbury Park, CA: Sage, 1990.
4. Patton MQ. *Qualitative Evaluation and Research Methods*. 2nd ed. Newbury Park, CA: Sage, 1990.
5. Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education*. 4th ed. Boston, MA: McGraw-Hill, 2000.
6. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*. 3rd ed. Baltimore, MD: Williams & Wilkins, 1996.
7. Simonton DK. Significant samples: the psychological study of eminent individuals. *Psychol Meth*. 1999;4:425-51.
8. Santy PA. *Choosing the Right Stuff: The Psychological Selection of Astronauts and Cosmonauts*. Westport, CT: Praeger, 1994.
9. Bosk CL. *Forgive and Remember: Managing Medical Failure*. Chicago, IL: University of Chicago Press, 1979.
10. Mizrahi T. *Getting Rid of Patients: Contradictions in the Socialization of Physicians*. New Brunswick, NJ: Rutgers University Press, 1986.
11. Brunswik E. *Systematic and Representative Design of Psychological Experiments*. Berkeley, CA: University of California Press, 1947.
12. Hammond KR. *Human Judgment and Social Policy*. New York: Oxford University Press, 1996.
13. Maher BA. Stimulus sampling in clinical research: representative design revisited. *J Consult Clin Psychol*. 1978;46:643-7.
14. van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med*. 1990;2:58-76.
15. Linn RL, Gronlund NE. *Measurement and Assessment in Teaching*. 7th ed. Englewood Cliffs, NJ: Prentice-Hall, 1995.
16. Millman J, Green J. The Specification and Development of Tests of Achievement and Ability. In: Linn RL (ed). *Educational Measurement*. 3rd ed. New York: Macmillan, 1989.
17. LaDuca A. Validation of professional licensure examinations: professions theory, test design, and construct validity. *Eval Health Prof*. 1994; 17:178-97.
18. Shye S, Elizur D, Hoffman M. Introduction to Facet Theory: Content Design and Intrinsic Data Analysis in Behavioral Research. *Applied Social Methods Series Vol. 35*. Thousand Oaks, CA: Sage, 1994.

Data Analysis and Statistics

*William C. McGaghie and Sonia Crandall**

REVIEW CRITERIA

- Data-analysis procedures are sufficiently described, and are sufficiently detailed to permit the study to be replicated.
- Data-analysis procedures conform to the research design; hypotheses, models, or theory drives the data analyses.
- The assumptions underlying the use of statistics are fulfilled by the data, such as measurement properties of the data and normality of distributions.
- Statistical tests are appropriate (optimal).
- If statistical analysis involves multiple tests or comparisons, proper adjustment of significance level for chance outcomes was applied.
- Power issues are considered in statistical studies with small sample sizes.
- In qualitative research that relies on words instead of numbers, basic requirements of data reliability, validity, trustworthiness, and absence of bias were fulfilled.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Data analysis along the “seamless web” of quantitative and qualitative research (see “Research Design,” earlier in this chapter) must be performed and reported according to scholarly conventions. The conventions apply to statistical treatment of data expressed as numbers and to qualitative data expressed as observational records, field notes, interview reports, abstracts from hospital charts, and other archival records. Data analysis must “get it right” to ensure that the research progression of design, methods (including data analysis), results, and conclusions and interpretation is orderly and integrated. Amplification of the seven data-analysis and statistical review criteria in this section underscores this assertion. The next article, entitled “Reporting of Statistical Analyses,” extends these ideas.

Quantitative

Statistical, or quantitative, analysis of research data is *not* the keystone of science. It does, however, appear in a large proportion of the research papers submitted to medical education journals. Reviewers expect a clear and complete de-

scription of research samples and data-analysis procedures in such papers.

Statistical analysis methods such as *t*-tests or analysis of variance (ANOVA) used to assess group differences, correlation coefficients used to assess associations among measured variables within intact groups, or indexes of effect such as odds ratios and relative risk in disease studies flow directly from the investigator’s research design. (Riegelman and Hirsch¹ give specific examples.) Designs focused on differences between experimental and control groups should use statistics that feature group contrasts. Designs focused on within-group associations should report results as statistical correlations in one or more of their many forms. Other data-analytic methods include meta-analysis,² i.e., quantitative integration of research data from independent investigations of the same research problem; procedures used to reduce large, complex data sets into more simplified structures, as in factor analysis or cluster analysis; and techniques to demonstrate data properties empirically, as in reliability analyses of achievement-test or attitude-scale data, multidimensional scaling, and other procedures. However, in all cases research design dictates statistical analysis of research data. Statistical analyses, when they are used, must be driven by the hypotheses, models, or theories that form the foundation of the study being judged.

Statistical analysis of research data often rests on assumptions about data measurement properties and the normality

*Lloyd Lewis, PhD, emeritus professor of the Medical College of Georgia, participated in early meetings of the Task Force and contributed to the earliest draft of this section.

of data distributions, and many other features. These assumptions must be satisfied to make the data analysis legitimate. By contrast, nonparametric, or “distribution-free,” statistical methods can be used to evaluate group differences or the correlations among variables when research measurements are in the form of categories (female–male, working–retired) or ranks (tumor stages, degrees of edema). Reviewers need to look for signs that the statistical analysis methods were based on sound assumptions about characteristics of the data and research design.

A reviewer must be satisfied that statistical tests presented in a research manuscript have been used and reported properly. Signs of flawed data analysis include inappropriate or suboptimal analysis (e.g., wrong statistics) and failure to specify post hoc analyses before collecting data.

Statistical analysis of data sets that is done without attention to an explicit research design or an a priori hypothesis can quickly become an exercise in “data dredging.” The availability of powerful computers, user-friendly statistical software, and large institutional data sets increases the likelihood of such mindless data analyses. Being able to perform hundreds of statistical tests in seconds is not a proxy for thoughtful attention to research design and focused data analysis. Reviewers should also be aware that, for example, in the context of only 20 statistical comparisons, one of the tests will be likely to achieve “significance” solely by chance. Multiple statistical tests or comparisons call for adjustment of significance levels (*p*-values) using the Bonferroni or a similar procedure to ensure accurate data interpretation.³

Research studies that involve small numbers of participants often lack enough statistical power to demonstrate significant results.⁴ This shortfall can occur even when a larger study would show a significant effect for an experimental intervention or for a correlation among measured variables. Whenever a reviewer encounters a “negative” study, the power question needs to be posed and ruled out as the reason for a nonsignificant result.

Qualitative

Analysis of qualitative data, which involves manipulation of words and symbols rather than numbers, is also governed by rules and rigor. Qualitative investigators are expected to use established, conventional approaches to ensure data quality and accurate analysis. Qualitative flaws include (but are not limited to) inattention to data triangulation (i.e., cross-checking information sources); insufficient description (lack of “thick description”) of research observations; failure to use recursive (repetitive) data analysis and interpretation; lack of independent data verification by colleagues (peer debriefing); lack of independent data verification by stakeholders (member checking); and absence of the a priori expres-

sion of the investigator’s personal orientation (e.g., homeopathy) in the written report.

Qualitative data analysis has a deep and longstanding research legacy in medical education and medical care. Well-known and influential examples are *Boys in White*, the classic study of student culture in medical school, published by Howard Becker and colleagues⁵; psychiatrist Robert Coles’ five-volume study, *Children of Crisis*⁶; the classic participant observation study by clinicians of patient culture on psychiatric wards published in *Science*⁷; and Terry Mizrahi’s observational study of the culture of residents on the wards, *Getting Rid of Patients*.⁸ Reviewers should be informed about the scholarly contribution of qualitative research in medical education. Prominent resources on qualitative research^{9–13} provide research insights and methodologic details that would be useful for the review of a complex or unusual study.

REFERENCES

1. Riegelman RK, Hirsch RP. Studying a Study and Testing a Test: How to Read the Medical Literature. 2nd ed. Boston, MA: Little, Brown, 1989.
2. Wolf FM. Meta-Analysis: Quantitative Methods for Research Synthesis. Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 59. Beverly Hills, CA: Sage, 1986.
3. Dawson B, Trapp RG. Basic and Clinical Biostatistics. 3rd ed. New York: Lange Medical Books/McGraw-Hill, 2001.
4. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Rev. ed. New York: Academic Press, 1977.
5. Becker HS, Geer B, Hughes EC, Strauss A. Boys in White: Student Culture in Medical School. Chicago, IL: University of Chicago Press, 1961.
6. Coles R. Children of Crisis: A Study of Courage and Fear. Vols. 1–5. Boston, MA: Little, Brown, 1967–1977.
7. Rosenhan DL. On being sane in insane places. *Science*. 1973;179:250–8.
8. Mizrahi T. Getting Rid of Patients: Contradictions in the Socialization of Physicians. New Brunswick, NJ: Rutgers University Press, 1986.
9. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, IL: Aldine, 1967.
10. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook. 2nd ed. Thousand Oaks, CA: Sage, 1994.
11. Harris IB. Qualitative methods. In: Norman GR, van der Vleuten CPM, Newble D (eds). International Handbook for Research in Medical Education. Dordrecht, The Netherlands, Kluwer, 2001.
12. Gicomini MK, Cook DJ. Users’ guide to the medical literature. XXIII. qualitative research in health care. A. Are the results of the study valid? *JAMA*. 2000;284:357–62.
13. Gicomini MK, Cook DJ. Users’ guide to the medical literature. XXIII. qualitative research in health care. B. What are the results and how do they help me care for my patients? *JAMA*. 2000;284:478–82.

RESOURCES

Goetz JP, LeCompte MD. Ethnography and Qualitative Design in Educational Research. Orlando, FL: Academic Press, 1984.

Guba EG, Lincoln YS. *Effective Evaluation*. San Francisco, CA: Jossey-Bass, 1981.

Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley & Sons, 1981.

Pagano M, Gauvreau K. *Principles of Biostatistics*. Belmont, CA: Duxbury

Press, 1993.

Patton MQ. *Qualitative Evaluation and Research Methods*. 2nd ed. Newbury Park, CA: Sage, 1990.

Winer BJ. *Statistical Principles in Experimental Design*. 2nd ed. New York: McGraw-Hill, 1971.

RESULTS

Reporting of Statistical Analyses

Glenn Regehr

REVIEW CRITERIA

- The assumptions underlying the use of statistics are considered, given the data collected.
- The statistics are reported correctly and appropriately.
- The number of analyses is appropriate.
- Measures of functional significance, such as effect size or proportion of variance accounted for, accompany hypothesis-testing analyses.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Even if the planned statistical analyses as reported in the Method section are plausible and appropriate, it is sometimes the case that the implementation of the statistical analysis as reported in the Results section is not. Several issues may have arisen in performing the analyses that render them inappropriate as reported in the Results section. Perhaps the most obvious is the fact that the data may not have many of the properties that were anticipated when the data analysis was planned. For example, although a correlation between two variables was planned, the data from one or the other (or both) of the variables may demonstrate a restriction of range that invalidates the use of a correlation. When a strong restriction of range exists, the correlation is bound to be low, not because the two variables are unrelated, but because the range of variation in the particular data set does not allow for the expression of the relationship in the correlation. Similarly, it may be the case that a *t*-test was planned to compare the means of two groups, but on review of the data, there is a bimodal distribution that raises doubts about the use of a mean and standard deviation to describe the data set. If so, the use of a *t*-test to evaluate the differences between the two groups becomes inappropriate. The reviewer should be alert to these potential problems and en-

sure, to the extent possible, that the data as collected continue to be amenable to the statistics that were originally intended. Often this is difficult because the data necessary to make this assessment are not presented. It is often necessary simply to assume, for example, that the sample distributions were roughly normal, since the only descriptive statistics presented are the mean and standard deviation. When the opportunity does present itself, however, the reviewer should evaluate the extent to which the data collected for the particular study satisfy the assumptions of the statistical tests that are presented in the Results section.

Another concern that reviewers should be alert to is the possibility that while appropriate analyses have been selected and performed, they have been performed poorly or inappropriately. Often enough data are presented to determine that the results of the analysis are implausible given the descriptive statistics, that "the numbers just don't add up." Alternatively, it may be the case that data and analyses are insufficiently reported for the reviewer to determine the accuracy or legitimacy of the analyses. Either of these situations is a problem and should be addressed in the review.

A third potential concern in the reporting of statistics is the presence in the Results section of analyses that were not anticipated in the Method section. In practice, the results of an analysis or a review of the data often lead to other

obvious questions, which in turn lead to other obvious analyses that may not have been anticipated. This type of expansion of analyses is not necessarily inappropriate, but the reviewer must determine whether it has been done with control and reflection. If the reviewer perceives an uncontrolled proliferation of analyses or if the new analyses appear without proper introduction or explanation, then a concern should be raised. It may appear to the reviewer that the author has fallen into a trap of chasing an incidental finding too far, or has enacted an unreflective or unsystematic set of analysis to “look for anything that is significant.” Either of these possibilities implies the use of inferential statistics for purposes beyond strict hypothesis testing and therefore stretches the statistics beyond their intended use.

On a similar note, reviewers should be mindful that as the number of statistical tests increases, the likelihood that at least one of the analyses will be “statistically significant” by chance alone also increases. When analyses proliferate it is important for the reviewer to determine whether the significance levels (*p*-values) have been appropriately adjusted to reflect the need to be more conservative.

Finally, it is important to note that statistical significance does not necessarily imply practical significance. Tests of statistical significance tell an investigator the probability that chance alone is responsible for study outcomes. But inferential statistical tests, whether significant or not, do not reveal the strength of association among research variables or the effect size. Strength of association is gauged by indexes of the proportion of variance in the dependent variable that is “explained” or “accounted for” by the independent variables in an analysis. Common indexes of explained variation are η^2 (η^2) in ANOVA and R^2 (coefficient of determination) in correlational analyses. Reviewers must be alert to the fact that statistically significant research results tell only part of the story. If a result is statistically significant, but the

independent variable accounts for only a very small proportion of the variance in the dependent variable, the result may not be sufficiently interesting to warrant extensive attention in the Discussion section. If none of the independent variables accounts for a reasonable proportion of the variance, then the study may not warrant publication.

RESOURCES

- Begg C, Cho M, Eastwood S., et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996; 276:637–9.
- Cohen J. The earth is round ($p < .05$). *Am Psychol*. 1994;49:997–1003.
- Dawson B, Trapp RG. *Basic and Clinical Biostatistics*. 3rd ed. New York: Lange Medical Books/McGraw–Hill, 2001.
- Hays WL. *Statistics*. New York: Holt, Rinehart and Winston, 1988.
- Hopkins KD, Glass GV. *Statistical Methods in Education and Psychology*. Boston, MA: Allyn & Bacon, 1995.
- Howell DC. *Statistical Methods for Psychology*. 4th ed. Belmont, CA: Wadsworth, 1997.
- Lang TA, Secic M. *How to Report Statistics in Medicine*. Philadelphia, PA: College of Physicians, 1997.
- Meehl PE. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *J Consult Clin Psychol*. 1978;46: 806–34.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet*. 1999. 354:1896–900.
- Norman GR, Striner DL. *Biostatistics: The Bare Essentials*. St. Louis, MO: Mosby, 1994 [out of print].
- Norusis MJ. *SPSS 9.0 Guide to Data Analysis*. Upper Saddle River, NJ: Prentice–Hall, 1999.
- Rennie D. CONSORT revised—improving the reporting of randomized trials. *JAMA*. 2001;285:2006–7.
- Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283: 2008–12.

Presentation of Results

Glenn Regehr

REVIEW CRITERIA

- Results are organized in a way that is easy to understand.
- Results are presented effectively; the results are contextualized.
- Results are complete.
- The amount of data presented is sufficient and appropriate.
- Tables, graphs, or figures are used judiciously and agree with the text.

ISSUES AND EXAMPLES RELATED TO CRITERIA

The Results section of a research paper lays out the body of evidence collected within the context of the study to support the conclusions and generalizations that are presented in the Discussion section. To be effective in supporting conclusions, the study results and their relation to the research questions and discussion points must be clear to the reader. Unless this relationship is clear, the reader cannot effectively judge the quality of the evidence or the extent to which it supports the claims in the Discussion section. Several devices can maximize this presentation, and reviewers need to be aware of these techniques so that they can effectively express their concerns about the Results section and provide useful feedback to the authors.

Organization of the Data and Analyses

The organization of the data and analyses is critical to the coherence of the Results section. The data and analyses should be presented in an orderly fashion, and the logic inherent in that order should be made explicit. There are several possible ways to organize the data, and the choice of organization ought to be strategic, reflecting the needs of the audience and the nature of the findings being presented. The reviewer should be alert to the organization being adopted and determine whether this particular organization is effective in conveying the results coherently.

One very helpful type of organization is to use a parallel structure across the entire research paper, that is, to make the organization of the results consistent with the organization of the other sections of the paper. Thus, the organization of the results section would mirror the organization of the research questions that were established in the Intro-

duction, it would be foreshadowed by the descriptions provided in the Method section, and it would anticipate the organization of points to be elaborated in the Discussion. If there are several research questions, hypotheses, or important findings, the Results section may be best presented as a series of subsections, with each subsection presenting the results that are relevant to a given question, hypothesis, or set of findings. This type of organization clarifies the point of each set of results or analyses and thus makes it relatively easy to determine how the results or analyses speak to the research questions. In doing so, this organization also provides an easy method for determining whether each of the research questions has been addressed appropriately and completely, and it provides a structure for identifying post hoc or additional analyses and serendipitous findings that might not have been initially anticipated.

However, there are other ways to organize a Results section that also maintain clarity and coherence and may better represent the data and analyses. Many of these methods are used in the context of qualitative research, but may also be relevant to quantitative/experimental/hypothesis-testing research designs. Similar to the description above, the results may be grouped according to themes arising in response to articulated research objectives (although, because themes often overlap, care must be taken to focus the reader on the theme under consideration while simultaneously identifying and explaining its relationship to the others). Alternately, the data may be organized according to the method of collection (interviews, observations, documents) or to critical phases in the data-analysis process (e.g., primary node coding and axial coding).

Regardless of the choice of organization, if it does not clearly establish the relevance of the data presented and the analyses performed, then the point of the presentation has not been properly established and the Results section has

failed in its purpose. If the results are not coherent, the reviewer must consider whether the problem lies in a poor execution of the analyses or in a poor organization of the Results section. If the first, the paper is probably not acceptable. If the second, the reviewer might merely want to suggest an organizational structure that would convey the results effectively.

Selection of Qualitative Data for Presentation

Qualitative research produces great amounts of raw material. And while the analysis process is designed to order and explain this raw material, at the point of presenting results the author still possesses an overwhelming set of possible excerpts to provide in a Results section. Selecting which data to present in a Results section is, therefore, critical. The logic that informs this selection process should be transparent and related explicitly to the research questions and objectives. Further, the author should make clear any implicit relationships among the results presented in terms of trends, contrasting cases, voices from a variety of perspectives on an issue, etc. Attention should be paid to ensuring that the selection process does not distort the overall gist of the entire data set. Further, narrative excerpts should be only as long as required to represent a theme or point of view, with care taken that the excerpts are not minimized to the point of distorting their meaning or diluting their character. This is a fine line, but its balance is essential to the efficient yet accurate presentation of findings about complex social phenomena.

The Balance of Descriptive and Inferential Statistics for Quantitative Data

In quantitative/hypothesis-testing papers, a rough parallel to the qualitative issue of selecting data for presentation is the balance of descriptive and inferential statistics. One common shortcoming in quantitative/hypothesis-testing papers is that the Results section focuses very heavily on inferential statistics with little attention paid to proper presentation of descriptive statistics. It is often forgotten that the inferential statistics are presented only to aid in the reasonable interpretation of the descriptive statistics. If the data (or pattern of data) to which the inferential statistics are being applied are not clear, then the point of the inferential statistics has not been properly established and the Results section has failed in its purpose. Again, however, this is a fine balance. Excessive presentation of descriptive statistics that do not speak to the research objectives may also make the Results section unwieldy and uninterpretable.

The Use of Narration for Quantitative Data

The Results section is not the place to elaborate on the implications of data collected, how the data fit into the larger theory that is being proposed, or how they relate to other literature. That is the role of the Discussion section. This being said, however, it is also true that the Results section of a quantitative/hypothesis-testing study should not be merely a string of numbers and Greek letters. Rather, the results should include a narrative description of the data, the point of the analysis, and the implications of the analysis for the data. The balance between a proper and complete description of the results and an extrapolation of the implications of the results for the research questions is a fine line. The distinction is important, however. Thus, it is reasonable—in fact, expected—that a Results section include a statement such as “Based on the pattern of data, the statistically significant two-way interaction in the analysis of variance implies that the treatment group improved on our test of knowledge more than the control group.” It is not appropriate for the Results section to include a statement such as “The ANOVA demonstrates that the treatment is effective” or, even more extreme, “the ANOVA demonstrates that we should be using our particular educational treatment rather than the other.” The first statement is a narrative description of the data interpreted in the context of the statistical analysis. The second statement is an extrapolation of the results to the research question and belongs in the Discussion. The third is an extreme over-interpretation of the results, a highly speculative value judgment about the importance of the outcome variables used in the study relative to the huge number of other variables and factors that must be weighed in any decision to adopt a new educational method (and, at least in the form presented above, should not appear anywhere in the paper). It is the reviewer’s responsibility to determine whether the authors have found the appropriate balance of description. If not, areas of concern (too little description or too much interpretation) should be identified in feedback to the authors.

Contextualization of Qualitative Data

Again, there is a parallel issue regarding the narrative presentation of data in qualitative studies. In the process of selecting material from a set of qualitative data (for example, when carving out relevant narrative excerpts from analyzed focus group transcripts), it is important that data not become “disconnected” and void of their original meaning(s). Narrative results, like numeric data, cannot stand on their own. They require descriptions of their origins in the data set, the nature of the analysis conducted, and the implications of the

analysis for the understandings achieved. A good qualitative Results section provides a framework for the selected data to ensure that their original contexts are sufficiently apparent that the reader can judge whether the ensuing interpretation is faithful to and reflects those contexts.

The Use of Tables and Figures

Tables and figures present tradeoffs because they often are the best way to convey complex data, yet they are also generally expensive of a journal's space. This is true for print (that is, paper) journals; but the situation is often different with electronic journals or editions. Most papers are still published in print journals, however. Thus, the reviewer must evaluate whether the tables and figures presented are the most efficient or most elucidating method of presenting the data and whether they are used appropriately sparingly. If it would be easy to present the data in the text without losing the structure or pattern of interest, this should be the preferred method of presentation. If tables or figures are used, every effort should be made to combine data into only a few. In addition, if data are presented in tables or figures, they should not be repeated in their entirety in the text. Rather, the text should be used to describe the table or figure, highlighting the key elements in the data as they pertain to the

relevant research question, hypothesis, or analysis. It is also worth noting that, although somewhat mundane, an important responsibility of the reviewer is to determine whether the data in the tables, the figures, and the text are consistent. If the numbers or descriptions in the text do not match those in the tables or figures, serious concern must be raised about the quality control used in the data analysis and interpretation.

The author gratefully acknowledges the extensive input and feedback for this chapter provided by Dr. Lorelei Lingard.

RESOURCES

- American Psychological Association. *Publication Manual*. 4th ed. Washington, DC: American Psychological Association, 1994.
- Harris IB. Qualitative methods. In: Norman GR, van der Vleuten CPM, Newble D (eds). *International Handbook for Research in Medical Education*. Amsterdam, The Netherlands: Kluwer, 2001.
- Henry GT. *Graphing Data: Techniques for Display and Analysis*. Applied Social Research Methods Series Vol. 36. Thousand Oaks, CA: Sage, 1995.
- Regehr G. The experimental tradition. In: Norman GR, van der Vleuten CPM, Newble D (eds). *International Handbook for Research in Medical Education*. Amsterdam, The Netherlands: Kluwer, 2001.
- Tufte ER. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983 (1998 printing).

DISCUSSION AND CONCLUSION

Discussion and Conclusion: Interpretation

Sonia J. Crandall and William C. McGaghie

REVIEW CRITERIA

- The conclusions are clearly stated; key points stand out.
- The conclusions follow from the design, methods, and results; justification of conclusions is well articulated.
- Interpretations of the results are appropriate; the conclusions are accurate (not misleading).
- The study limitations are discussed.
- Alternative interpretations for the findings are considered.
- Statistical differences are distinguished from meaningful differences.
- Personal perspectives or values related to interpretations are discussed.
- Practical significance or theoretical implications are discussed; guidance for future studies is offered.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Research follows a logical process. It starts with a problem statement and moves through design, methods, and results. Researchers' interpretations and conclusions emerge from these four interconnected stages. Flaws in logic can arise at any of these stages and, if they occur, the author's interpretations of the results will be of little consequence. Flaws in logic can also occur at the interpretation stage. The researcher may have a well-designed study but obscure the true meaning of the data by misreading the findings.¹

Reviewers need to have a clear picture of the meaning of research results. They should be satisfied that the evidence is discussed adequately and appears reliable, valid, and trustworthy. They should be convinced that interpretations are justified given the strengths and limitations of the study. In addition, given the architecture, operations, and limitations of the study, reviewers should judge the generalizability and practical significance of its conclusions.

The organization of the Discussion section should match the structure of the Results section in order to present a coherent interpretation of data and methods. Reviewers need to determine how the discussion and conclusions relate to the original problem and research questions. Most important, the conclusions must be clearly stated and justified, illustrating key points. Broadly, important aspects to consider include whether the conclusions are reasonable based on the description of the results; on how the study results relate to other research outcomes in the field, including consensus, conflicting, and unexpected findings; on how the study outcomes expand the knowledge base in the field and inform future research; and on whether limitations in the design, procedures, and analyses of the study are described. Failure to discuss the limitations of the study should be considered a serious flaw.

On a more detailed level, reviewers must evaluate whether the authors distinguish between (1) inferences drawn from the results, which are based on data-analysis procedures and (2) extrapolations to the conceptual framework used to design the study. This is the difference between formal hypothesis testing and theoretical discussion.

Quantitative Approaches

From the quantitative perspective, when interpreting hypothesis-testing aspects of a study, authors should discuss the meaning of both statistically significant and non-significant results. A statistically significant result, given its *p*-value and confidence interval, may have no implications for practice.² Authors should explain whether each hypothesis is confirmed or refuted and whether each agrees or conflicts with

previous research. Results or analyses should not be discussed unless they are presented in the Results section.

Data may be misrepresented or misinterpreted, but more often errors come from over-interpreting the data from a theoretical perspective. For example, a reviewer may see a statement such as "The sizeable correlation between test scores and 'depth of processing' measures clearly demonstrates that the curriculum should be altered to encourage students to process information more deeply." The curricular implication may be true but it is not supported by data. Although the data show that encouraging an increased depth of processing improves test scores, this outcome does not demonstrate the need to change curriculum. The intent to change the curriculum is a value statement based on a judgment about the utility of high test scores and their implications for professional performance. Curricular change is not implied directly from the connection between test scores and professional performance.

The language used in the Discussion needs to be clear and precise. For example, in research based on a correlation design, the Discussion needs to state whether the correlations derive from data collected concurrently or over a span of time.³ Correlations over time suggest a predictive relationship among variables, which may or may not reflect the investigator's intentions. The language used to discuss such an outcome must be unambiguous.

Qualitative Approaches

Qualitative researchers must convince the reviewer that their data are trustworthy. To describe the trustworthiness of the collected data, the author may use criteria such as credibility (internal validity) and transferability (external validity) and explain how each was addressed.⁴ (See Giacomini and Cook, for example, for a thorough explanation of assessing validity in qualitative health care research.⁵) Credibility may be determined through data triangulation, member checking, and peer debriefing.^{4,6} Triangulation compares multiple data sources, such as a content analysis of curriculum documents, transcribed interviews with students and the faculty, patient satisfaction questionnaires, and observations of standardized patient examinations. Member checking is a process of "testing" interpretations and conclusions with the individuals from whom the data were collected (interviews).⁴ Peer debriefing is an "external check on the inquiry process" using disinterested peers who parallel the analytic procedures of the researcher to confirm or expand interpretations and conclusions.⁴ Transferability implies that research findings can be used in other educational contexts (generalizability).^{6,7} The researcher cannot, however, establish external validity in the same way as in quantitative research.⁴ The

reviewer must judge whether the conclusions transfer to other contexts.

Biases

Both qualitative and quantitative data are subject to bias. When judging qualitative research, reviewers should carefully consider the meaning and impact of the author's personal perspectives and values. These potential biases should be clearly explained because of their likely influence on the analysis and presentation of outcomes. Those biases include the influence of the researcher on the study setting, the selective presentation and interpretation of results, and the thoroughness and integrity of the interpretations. Peshkin's work is a good example of announcing one's subjectivity and its potential influence on the research process.⁸ He and other qualitative researchers acknowledge their responsibility to explain how their values may affect research outcomes. Reviewers of qualitative research need to be convinced that the influence of subjectivity has been addressed.⁶

REFERENCES

1. Day RA. *How to Write and Publish a Scientific Paper*. 5th ed. Phoenix, AZ: Oryx Press, 1998.
2. Rosenfeld RM. The seven habits of highly effective data users [editorial]. *Otolaryngol Head Neck Surg*. 1998;118:144–58.
3. Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education*. 4th ed. Boston, MA: McGraw–Hill Higher Education, 2000.
4. Lincoln YS, Guba EG. Chapter 11. *Naturalistic Inquiry*. Newbury Park, CA: Sage, 1985.
5. Giacomini MK, Cook DJ. Users' guide to the medical literature. XXIII. Qualitative research in health care. A. Are the results of the study valid? *JAMA*. 2000;284:357–62.
6. Grbich C. *Qualitative Research in Health*. London, U.K.: Sage, 1999.
7. Erlandson DA, Harris EL, Skipper BL, Allen SD. *Doing Naturalistic Inquiry: A Guide to Methods*. Newbury Park, CA: Sage, 1993.

8. Peshkin A. *The Color of Strangers, the Color of Friends*. Chicago, IL: University of Chicago Press, 1991.

RESOURCES

- Day RA. *How to Write and Publish A Scientific Paper*. 5th ed. Phoenix, AZ: Oryx Press, 1998 [chapter 10].
- Erlandson DA, Harris EL, Skipper BL, Allen SD. *Doing Naturalistic Inquiry: A Guide to Methods*. Newbury Park, CA: Sage, 1993.
- Fraenkel JR, Wallen NE. *How to Design and Evaluate Research in Education*. 4th ed. Boston, MA: McGraw–Hill Higher Education, 2000 [chapters 19, 20].
- Gehlbach SH. *Interpreting the Medical Literature*. 3rd ed. New York: McGraw–Hill, 1992.
- Guiding Principles for Mathematics and Science Education Research Methods: Report of a Workshop. Draft. Workshop on Education Research Methods, Division of Research, Evaluation and Communication, National Science Foundation, November 19–20, 1998, Ballston, VA. Symposium presented at the meeting of the American Education Research Association, April 21, 1999, Montreal, Quebec, Canada. (<http://bear.berkeley.edu/publications/report11.html>). Accessed 5/1/01.
- Huth EJ. *Writing and Publishing in Medicine*. 3rd ed. Baltimore, MD: Williams & Wilkins, 1999.
- Lincoln YS, Guba EG. *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications, 1985 [chapter 11].
- Miller WL, Crabtree BF. Clinical research. In: Denzin NK, Lincoln YS (eds). *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, 1994:340–53.
- Patton MQ. *Qualitative Evaluation and Research Methods*. 2nd ed. Newbury Park, CA: Sage, 1990.
- Peshkin A. The goodness of qualitative research. *Educ Res*. 1993;22:23–9.
- Riegelman RK, Hirsch RP. *Studying a Study and Testing a Test: How to Read the Health Science Literature*. 3rd ed. Boston, MA: Little, Brown, 1996.
- Teaching/Learning Resources for Evidence Based Practice. Middlesex University, London, U.K. (<http://www.mdx.ac.uk/www/rctsh/ebp/main.htm>). Accessed 5/1/01.
- Users' Guides to Evidence-Based Practice. Centres for Health Evidence [Canada]. (http://www.cche.net/principles/content_all.asp). Accessed 5/1/01.

TITLE, AUTHORS, AND ABSTRACT

Title, Authors, and Abstract

Georges Bordage and William C. McGaghie

REVIEW CRITERIA

- The title is clear and informative.
- The title is representative of the content and breadth of the study (not misleading).
- The title captures the importance of the study and the attention of the reader.
- The number of authors appears to be appropriate given the study.
- The abstract is complete (thorough); essential details are presented.
- The results in the abstract are presented in sufficient and specific detail.
- The conclusions in the abstract are justified by the information in the abstract and the text.
- There are no inconsistencies in detail between the abstract and the text.
- All of the information in the abstract is present in the text.
- The abstract overall is congruent with the text; the abstract gives the same impression as the text.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

When a manuscript arrives, the reviewer immediately sees the title and the abstract, and in some instances—depending on the policy of the journal—the name of the authors. This triad of title, authors, and abstract is both the beginning and the end of the review process. It orients the reviewer, but it can be fully judged only after the manuscript is analyzed thoroughly.

Title

The title can be viewed as the shortest possible abstract. Consequently, it needs to be clear and concise while accurately reflecting the content and breadth of the study. As one of the first “outside” readers of the manuscript, the reviewer can judge if the title is too general or misleading, whether it lends appropriate importance to the study, and if it grabs the reader’s attention.

The title of an article must have appeal because it prompts the reader’s decision to study the report. A clear and informative title orients the readers and reviewers to relevant information. Huth¹ describes two key qualities of titles, “indicative” and “informative.” The indicative aspect of the

title tells the reader about the nature of the study, while the informative aspect presents the message derived from the study results. To illustrate, consider the following title: “A Survey of Academic Advancement in Divisions of General Internal Medicine.” This title tells the readers what was done (i.e., it is indicative) but fails to convey a message (i.e., it is not informative). A more informative title would read “A Survey of Academic Advancement in Divisions of General Internal Medicine: Slower Rate and More Barriers for Women.” The subtitle now conveys the message while still being concise.

Authorship

Reviewers are not responsible for setting criteria for authorship. This is a responsibility of editors and their editorial boards. When authors are revealed to the reviewer, however, the reviewer can help detect possible “authorship inflation” (too many authors) or “ghost authors” (too few true authors).

The *Uniform Requirements for Manuscripts Submitted to Biomedical Journals*² covers a broad range of issues and contains perhaps the most influential single definition of authorship, which is that

Each author should have participated sufficiently in the work to take public responsibility for the content. Authorship credit should be based only on substantial contributions to (a) conception and design, or analysis and interpretation of data; and to (b) drafting the article or revising it critically for important intellectual content; and on (c) final approval of the version to be published. Conditions (a), (b), and (c) must all be met.

Furthermore, "Any part of an article critical to its main conclusions must be the responsibility of at least one author," that is, a manuscript should not contain any statement or content for which none of the authors can take responsibility. More than 500 biomedical journals have voluntarily allied themselves with the *Uniform Requirements* standards, although not all of them accept this strict definition of authorship. Instead, they use different numbers of authors and/or combinations of the conditions for their definitions. Also, different research communities have different traditions of authorship, some of which run counter to the *Uniform Requirements* definition.

The number of authors per manuscript has increased steadily over the years, both in medical education and in clinical research. Dimitroff and Davis report that the number of articles with four or more authors in medical education is increasing faster than the number of papers with fewer authors.³ Comparing numbers in 1975 with those in 1998, Drenth found that the mean number of authors of original articles in the *British Medical Journal* steadily increased from 3.21 (SD = 1.89) to 4.46 (SD = 2.04), a 1.4-fold jump.⁴ While having more authors is likely to be an indication of the increased number of people involved in research activities, it could also signal inflation in the number of authors to build team members' curricula vitae for promotion. From an editorial standpoint, this is "unauthorized" authorship.

More and more journals are publishing their specific criteria for authorship to help authors decide who should be included in the list of authors. Some journals also require each author to complete and sign a statement of authorship indicating their significant contributions to the manuscript. For example, the *Annals of Internal Medicine* offers a list of contribution codes that range from conception and design of the study to obtaining funds or collecting and assembling data, as well as a space for "other contributions." The contribution codes and signed statement are a sound reminder and acknowledgement for authors and a means for editors to judge eligibility of authorship.

Huth argues that certain conditions alone do not justify authorship. These conditions include acquiring funds, collecting data, administering the project, or proofreading or editing manuscript drafts for style and presentation, not ideas.^{5,6} Under these conditions, doing data processing without statistical conceptualization is insufficient to qualify for

authorship. Such contributions can be recognized in a footnote or in an acknowledgement. Other limited or indirect contributions include providing subjects, participating in a pilot study, or providing materials or research space.⁷ Finally, some so-called "contributions" are honorary, such as crediting department chairpersons, division chiefs, laboratory directors, or senior faculty members for *pro forma* involvement in creative work.⁸

Conversely, no person involved significantly in the study should be omitted as an author. Flanagan et al.⁸ found that 11% of articles in three large-circulation general medicine journals in 1996 had "ghost authors," individuals who were not named as authors but who had contributed substantially to the work. A reviewer may suspect ghost authorship when reviewing a single-authored manuscript reporting a complex study.

When authors' names are revealed on a manuscript, reviewers should indicate to the editor any suspicion about there being too many or too few authors.

Abstracts

Medical journals began to include abstracts with articles in the late 1960s. Twenty years later an ad hoc working group proposed "more informative abstracts" (MIAs) based on published criteria for the critical appraisal of the medical literature.⁹ The goals of the MIAs were threefold: "(1) assist readers to select appropriate articles more quickly, (2) allow more precise computerized literature searches, and (3) facilitate peer review before publication." The group proposed a 250-word, seven-part abstract written in point form (versus narrative). The original seven parts were soon increased to eight^{10,11}: *objective* (the exact question(s) addressed by the article), *design* (the basic design of the study), *setting* (the location and level of clinical care [or education]), *patients or participants* (the manner of selection and numbers of patients or participants who entered and completed the study), *interventions* (the exact treatment or intervention, if any), *main outcome measures* (the primary study outcome measure), *results* (key findings), and *conclusions* (key conclusions including direct clinical [or educational] applications).

The working group's proposal was published in the *Annals of Internal Medicine* and was called by *Annals* editor Edward Huth the "structured abstract."¹² Most of the world's leading clinical journals followed suit. Journal editors anticipated that giving reviewers a clear summary of salient features of a manuscript as they begin their review would facilitate the review process. The structured abstract provides the reviewer with an immediate and overall sense of the reported study right from the start of the review process. The "big picture" offered by the structured abstract helps reviewers frame their analysis.

The notion of MIAs, or structured abstracts, was soon extended to include review articles.¹³ The proposed format of the structured abstract for review articles contained six parts: *purpose* (the primary objective of the review article), *data identification* (a succinct summary of data sources), *study selection* (the number of studies selected for review and how they were chosen), *data extraction* (the type of guidelines used for abstracting data and how they were applied), *results of data synthesis* (the methods of data analysis and key results), and *conclusions* (key conclusions, including potential applications and research needs).

While there is evidence that MIAs do provide more information,^{14,15} some investigators found that substantial amounts of information expected in the abstract was still missing even when that information was present in the text.¹⁶ A study by Pitkin and Branagan showed that specific instructions to authors about three types of common defects in abstracts—inconsistencies between abstract and text, information present in the abstract but not in the text, and conclusions not justified by the information in the abstract—were ineffective in lowering the rate of defects.¹⁷ Thus reviewers must be especially attentive to such defects.

REFERENCES

1. Huth EJ. Types of titles. In: *Writing and Publishing in Medicine*. 3rd ed. Baltimore, MD: Williams & Wilkins, 1999:131–2.
2. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. 5th ed. *JAMA*. 1997;277:927–34. (<http://jama.ama-assn.org/info/auinst>). Accessed 5/23/01.
3. Dimitroff A, Davis WK. Content analysis of research in undergraduate education. *Acad Med*. 1996;71:60–7.
4. Drenth JPH. Multiple authorship. The contribution of senior authors. *JAMA*. 1998;280:219–21.
5. Huth EJ. Chapter 4. Preparing to write: materials and tools. appendix A, guidelines on authorship, and appendix B, the “uniform requirements” document: an abridged version. In: *Writing and Publishing in Medicine*, 3rd ed. Baltimore, MD: Williams & Wilkins, 1999:41–4, 293–6, 297–9.
6. Huth EJ. Guidelines on authorship of medical papers. *Ann Intern Med*. 1986;104:269–74.
7. Hoen WP, Walvoort HC, Overbeke JPM. What are the factors determining authorship and the order of the authors’ names? *JAMA*. 1998;280:217–8.
8. Flanagan A, Carey LA, Fontanarosa PB, et al. Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *JAMA*. 1998;280:222–4.

9. Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. *Ann Intern Med*. 1987;106:598–604.
10. Altman DG, Gardner MJ. More informative abstracts (letter). *Ann Intern Med*. 1987;107:790–1.
11. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med*. 1990;113:69–76.
12. Huth EJ. Structured abstracts for papers reporting clinical trials. *Ann Intern Med*. 1987;106:626–7.
13. Mulrow CD, Thacker SB, Pugh JA. A proposal for more informative abstracts of review articles. *Ann Intern Med*. 1988;108:613–5.
14. Comans ML, Overbeke AJ. The structured summary: a tool for reader and author. *Ned Tijdschr Geneesk*. 1990;134:2338–43.
15. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association*. *Can Med Assoc J*. 1994;150:1611–4.
16. Fromm P, Fromm J. Deficiencies in structured medical abstracts. *J Clin Epidemiol*. 1993;46:591–4.
17. Pitkin RM, Branagan MA. Can the accuracy of abstracts be improved by providing specific instructions? A randomized controlled trial. *JAMA*. 1998;280:267–9.

RESOURCES

- American College of Physicians. Resources for Authors—Information for authors: *Annals of Internal Medicine*. Available from: MS Internet Explorer via the Internet (<http://www.acponline.org/journals/resource/info4aut.htm>). Accessed 9/27/00.
- Fye WB. Medical authorship: traditions, trends, and tribulations. *Ann Intern Med*. 1990;113:317–25.
- Godlee F. Definition of authorship may be changed. *BMJ*. 1996;312:1501–2.
- Huth EJ. *Writing and Publishing in Medicine*. 3rd ed. Baltimore, MD: Williams & Wilkins, 1999.
- Lundberg GD, Glass RM. What does authorship mean in a peer-reviewed medical journal? [editorial]. *JAMA*. 1996;276:75.
- National Research Press. Part 4: Responsibilities. In: *Publication Policy*. (http://www.monographs.nrc.ca/cgi-bin/cisti/journals/rp/rp2_cust_e/pubpolicy). Accessed 6/5/01.
- Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA*. 1999;281:1110–1.
- Rennie D, Yank V, Emanuel L. When authorship fails. A proposal to make contributors accountable. *JAMA*. 1997;278:579–85.
- Shapiro DW, Wenger NS, Shapiro MF. The contributions of authors to multi-authored biomedical research papers. *JAMA*. 1994;271:438–42.
- Slone RM. Coauthors’ contributions to major papers published in the *AJR*: frequency of undeserved coauthorship. *Am J Roentgenol*. 1996;167:571–9.
- Smith J. Gift authorship: a poisoned chalice? Not usually, but it devalues the coinage of scientific publication. *BMJ*. 1994;309:1456–7.

OTHER

Presentation and Documentation

Gary Penn, Ann Steinecke, and Judy A. Shea

REVIEW CRITERIA

- The text is well written and easy to follow.
- The vocabulary is appropriate.
- The content is complete and fully congruent.
- The manuscript is well organized.
- The data reported are accurate (e.g., numbers add up) and appropriate; tables and figures are used effectively and agree with the text.
- Reference citations are complete and accurate.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Presentation refers to the clarity and effectiveness with which authors communicate their ideas. In addition to evaluating how well the researchers have constructed their study, collected their data, and interpreted important patterns in the information, reviewers need to evaluate whether the authors have successfully communicated all of these elements. Ensuring that ideas are properly presented, then, is the reviewer's final consideration when assessing papers for publication.

Clear, effective communication takes different forms. Straight prose is the most common; carefully chosen words, sentences, and paragraphs convey as much or as little detail as necessary. The writing should not be complicated by inappropriate vocabulary such as excessive jargon; inaccurately used words; undefined acronyms; or new, controversial, or evolving vocabulary. Special terms should be defined, and the vocabulary chosen for the study and presentation should be used consistently. Clarity is also a function of a manuscript's organization. In addition to following a required format, such as IMRaD, a manuscript's internal organization (sentences and paragraphs) should follow a logical progression that supports the topic. All information contained in the text should be clearly related to the topic.

In addition to assessing the clarity of the prose, reviewers should be prepared to evaluate graphic representations of information—tables, lists, and figures. When well done, they present complex information efficiently, and they reveal

ideas that would take too many words to tell. Tables, lists, and figures should not simply repeat information that is given in the text; nor should they introduce data that are not accounted for in the Method section or contradict information given in the text.

Whatever form the presentation of information takes, the reviewer should be able to grasp the substance of the communication without having to work any harder than necessary. Of course, some ideas are quite complex and require both intricate explanation and great effort to comprehend, but too often simple ideas are dressed up in complicated language without good reason. The reviewer needs to consider how well the author has matched the level of communication to the complexity of the substance in his or her presentation.

Poor presentation may, in fact, directly reflect poor content. When the description of the method of a study is incomprehensible to the reviewer, it may hint at the researcher's own confusion about the elements of his or her study. Jargon-filled conclusions may reflect a researcher's inability to apply his or her data to the real world. This is not always true, however; some excellent researchers are simply unable to transfer their thoughts to paper without assistance. Sorting these latter authors from the former is a daunting task, but the reviewer should combine a consideration of the presentation of the study with his or her evaluation of the methodologic and interpretive elements of the paper.

The reviewer's evaluation of the presentation of the manuscript should also extend to the presentation of references.

Proper documentation ensures that the source of material cited in the manuscript is accurately and fully acknowledged. Further, accurate documentation allows readers to quickly retrieve the referenced material. And finally, proper documentation allows for citation analysis, a count of the times a published article is cited in subsequent articles. Journals describe their documentation formats in their instructions to authors, and the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals* details suggested formats. Reviewers should not concern themselves with the specific details of a reference list's format; instead, they should look to see whether the documentation appears to provide complete and up-to-date information about all the material cited in the text (e.g., author's name, title, journal, date, volume, page number). Technologic advances in the presentation of information have meant the creation of citation formats for a wide variety of media, so reviewers can expect there to be documentation for any type of material presented in the text.

The extent to which a reviewer must judge presentation depends on the journal. Some journals (e.g., *Academic Medicine*) employ editors who work closely with authors to clearly shape text and tables; reviewers, then, can concentrate on the substance of the study. Other journals publish articles pretty much as authors have submitted them; in those cases, the reviewers' burden is greater. Reviewers may not be expected to edit the papers, but their comments can help authors revise any presentation problems before final acceptance.

Because ideas are necessarily communicated through words and pictures, presentation and substance often seem to overlap. As much as possible, the substantive aspects of the criteria for this section are covered in other sections of this guide.

RESOURCES

- Becker HS, Richards P. *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article*. Chicago, IL: University of Chicago Press, 1986.
- Browner WS. *Publishing and Presenting Clinical Research*. Baltimore, MD: Lippincott, Williams & Wilkins, 1999.
- Day RA. *How to Write and Publish a Scientific Paper*. 4th ed. Phoenix, AZ: Oryx Press, 1994.
- Day RA. *Scientific English: A Guide for Scientists and Other Professionals*. Phoenix, AZ: Oryx Press, 1992.
- Fishbein M. *Medical Writing: The Technic and the Art*. 4th ed. Springfield, IL: Charles C Thomas, 1972.
- Hall GM. *How to Write a Paper*. London, U.K.: BMJ Publishing Group, 1994.
- Howard VA, Barton JH. *Thinking on Paper: Refine, Express, and Actually Generate Ideas by Understanding the Processes of the Mind*. New York: William Morrow and Company, 1986.
- International Committee of Medical Journal Editors. *Uniform Requirements for Manuscripts Submitted to Biomedical Journals*. *Ann Intern Med*. 1997;126:36-47; <www.acponline.org/journals/annals/01janr97/unifreq> (updated May 1999).
- Kirkman J. *Good Style: Writing for Science and Technology*. London, U.K.: E & FN Spon, 1997.
- Matkin RE, Riggat TF. *Persist and Publish: Helpful Hints for Academic Writing and Publishing*. Niwot, CO: University Press of Colorado, 1991.
- Morgan P. *An Insider's Guide for Medical Authors and Editors*. Philadelphia, PA: ISI Press, 1986.
- Sheen AP. *Breathing Life into Medical Writing: A Handbook*. St. Louis, MO: C. V. Mosby, 1982.
- Tornquist EM. *From Proposal to Publication: An Informal Guide to Writing about Nursing Research*. Menlo Park, CA: Addison-Wesley, 1986.
- Tufte ER. *Envisioning Information*. Cheshire, CT: Graphics Press, 1990.
- Tufte ER. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.
- Tufte ER. *Visual Explanations*. Cheshire, CT: Graphics Press, 1997.
- Zeiger M. *Essentials of Writing Biomedical Research Papers*. 2nd ed. New York: McGraw-Hill, 1999.

Scientific Conduct

Louis Pangaro and William C. McGaghie

REVIEW CRITERIA

- There are no instances of plagiarism.
- Ideas and materials of others are correctly attributed.
- Prior publication by the author(s) of substantial portions of the data or study is appropriately acknowledged.
- There is no apparent conflict of interest.
- There is an explicit statement of approval by an institutional review board (IRB) for studies directly involving human subjects or data about them.

ISSUES AND EXAMPLES RELATED TO THE CRITERIA

Reviewers provide an essential service to editors, journals, and society by identifying issues of ethical conduct that are implicit in manuscripts.¹ Concerns for reviewers to consider include issues of “authorship” (defining who is responsible for the material in the manuscript—see “Title, Authors, and Abstract” earlier in this chapter), plagiarism (attributing others’ words or ideas to oneself), lack of correct attribution of ideas and insights (even if not attributing them to oneself), falsifying data, misrepresenting publication status,² and deliberate, inappropriate omission of important prior research. Because authors are prone to honest omissions in their reviews of prior literature, or in their awareness of others’ work, reviewers may also be useful by pointing out missing citations and attributions. It is not unusual for authors to cite their own work in a manuscript’s list of references, and it is the reviewer’s responsibility to determine the extent and appropriateness of these citations (see “Reference to the Literature and Documentation”) earlier. Multiple publication of substantially the same studies and data is a more vexing issue. Reviewers cannot usually tell whether parts of the study under review have already been published or detect when part or all of the study is also “in press” with another journal. Some reviewers try to do a “search” on the topic of a manuscript, and, when authorship is not masked, of the authors themselves. This may detect prior or duplicate publication and also aid in a general review of citations.

Finally, reviewers should be alert to authors’ suppression of negative results. A negative study, one with conclusions that do not ultimately confirm the study’s hypothesis (or that reject the “null hypothesis”), may be quite valuable if the research question was important and the study design was

rigorous. Such a study merits, and perhaps even requires, publication, and reviewers should not quickly dismiss such a paper without full consideration of the study’s relevance and its methods.³ Yet authors may not have the confidence to include results that do not support the hypothesis. Reviewers should be alert to this fear about negative results and read carefully to detect the omission of data that would be expected. (It is important to note that nowhere in this document of guidance for reviewers is there a criterion that labels a “negative study” as flawed because it lacks a “positive” conclusion.)

Reviewers should be alert to several possible kinds of conflict of interest. The most familiar is a material gain for the author from specific outcomes of a study. In their scrutiny of methods (as covered in all articles in the “Method” section of this chapter), reviewers safeguard the integrity of research, but financial interest in an educational project may not be apparent. Reviewers should look for an explicit statement concerning financial interest when any marketable product (such as a CD-ROM or software program) either is used or is the subject of investigation. Such an “interest” does not preclude publication, but the reviewer should expect a clear statement that there is no commercial interest or of how such a conflict of interest has been handled.

Recently, regulations for the protection of human subjects have been interpreted as applying to areas of research at universities and academic medical centers that they have not been applied to before.⁴ For instance, studying a new educational experience with a “clinical research” model that uses an appropriate control group might reveal that one of the two groups had had a less valuable educational experience. Hence, informed consent and other protections would be the expected standard for participation, as approved by

an IRB.⁵ In qualitative research, structured qualitative interviews could place a subject at risk if unpopular opinions could be attributed to the individual. Here again, an ethical and legal responsibility must be met by the researchers. We should anticipate that medical education research journals (and perhaps health professions journals also) will require statements about IRB approval in all research papers.

In summary, manuscripts should meet standards of ethical behavior, both in the process of publication and in the conduct of research. Any field that involves human subjects—particularly fields in the health professions—should meet the ethical standards for such research, including the new requirements for education research. Therefore, reviewers fulfill an essential function in maintaining the integrity of academic publications.

REFERENCES

1. Caellegh A. Role of the journal editor in sustaining integrity in research. *Acad Med.* 1993;68(9 suppl):S23–S29.
2. LaFollette MC. *Stealing Into Print: Fraud, Plagiarism, and Misconduct in Scientific Publishing.* Berkeley, CA: University of California Press, 1992.
3. Chalmers I. Underreporting research is scientific misconduct, *JAMA.* 1990;263:1405–6.
4. Code of Federal Regulation, Title 45, Public Welfare, Part 46—Protection of Human Subjects, Department of Human Services. (<http://www.etsu.edu/ospa/exempt2.htm>). Accessed 4/1/00.
5. Casarett D, Karlawish J, Sugarman J. Should patients in quality improvement activities have the same protections as participants in research studies? *JAMA.* 2000;284:1786–8.

RESOURCES

- The Belmont Report [1976]. (<http://ddonline.gsm.com/demo/consult/belmont-int.htm>). Accessed 5/23/01.
- Committee on Publication Ethics. The COPE Report 1998. (<http://www.bmj.com/misc/cope/tex1.shtml>). Accessed 5/9/01.
- Committee on Publication Ethics. The COPE Report 2000. (<http://www.bmj.com/publicationethics/cope/cope.htm>). Accessed 5/9/01.
- Council of Biology Editors. *Ethics and Policy in Scientific Publication.* Bethesda, MD: Council of Biology Editors, 1990.
- Council for International Organizations of Medical Sciences (CIOMS), *International Guidelines for Ethical Review of Epidemiological Studies,* Geneva, 1991. In: King NMP, Henderson GE, Stein J (eds). *Beyond Regulations: Ethics in Human Subjects Research.* Chapel Hill, NC: University of North Carolina Press, 1999.
- The Hastings Center's Bibliography of Ethics, Biomedicine, and Professional Responsibility. Frederick, MD: University Publications of America in Association with the Hastings Center, 1984.
- Henry RC, Wright DE. When are medical students considered subjects in program evaluation? *Acad Med.* 2001;76:871–5.
- National Research Press. Part 4: Responsibilities. In: *Publication Policy.* (http://www.monographs.nrc.ca/cgi-bin/cisti/journals/rp/rp2_cust_e?pubpolicy). Accessed 6/5/01.
- Roberts LW, Geppert C, Connor R, Nguyen K, Warner TD. An invitation for medical educators to focus on ethical and policy issues in research and scholarly practice. *Acad Med.* 2001;76:876–85.