

The background of the entire page is a white field. A large, dark, textured shape, resembling a splatter or a stylized animal, is positioned in the upper half. Overlaid on this and the rest of the page is a network of thin, green lines that form a grid-like pattern, with some lines intersecting at various angles.

Second International Congress on Peer Review in Biomedical Publication

Sponsored by the American Medical Association

**September 9-11, 1993
Chicago, Illinois**

Welcome!

Welcome to the Second International Congress on Peer Review in Biomedical Publication! The first such Congress, held in 1989, was, judged by the liveliness of the sessions and the large number of written comments I received afterward, a success. Whether this one succeeds is now up to the presenters, who I trust will be clear, inventive, and challenging, and the audience, on whom I rely to be critical, outspoken, and contentious.

In 1989, I referred to the anxieties we went through after George Lundberg and I decided, in 1986, to invite the scientific community to carry out and present research on peer review. Would we receive any abstracts? Would anyone come to hear the abstracts presented? While the last four years have not been free of this sort of worry, the fact that we have received 110 abstracts, some of high quality, suggests that the infant science is beginning to take root and that henceforth, editors will have to get used to the idea that their own practices are a suitable subject for inquiry.

I am very grateful to the American Medical Association for its enlightened support of this endeavor and in particular to James Todd, AMA Executive Vice President; Larry Joyce, Senior Vice President of Publishing and Communications; Robert Kennett, Vice President of Publishing; and George Lundberg, Editor-in-Chief, Scientific Publications. Sharon Kremkau, Mary Lynne Sobotka, and the staff of Meeting Planning and Management have been invaluable. Helga Wutz, Roxanne Young, Sharon Iverson, Vee Bailey, Elaine Williams, Greg Reynolds, Terri Hanley, Cheryl Iverson, and Jeff Molter have contributed in innumerable ways to the Congress. Cheryl Manno has shouldered a huge load and done so efficiently and without complaint.

Marcia Angell, Kathy Case, Cindy Clark, Lewis Gidez, Charles Hancock, Dan Koshland, John Last, John Maddox, Maeve O'Connor, John Overbeke, David Sharp, Pat Woolf, and especially Jane Smith have advised us in many ways and assisted us with advertising in their respective journals and promoting the Congress to their respective organizations. We have all benefited from the advice and flair of Elizabeth Knoll, who did so much to put the first Congress together.

The members of the Advisory Board have been a great help, especially John Bailar, Annette Flanagin, Suzanne Fletcher, Brian Haynes, Elizabeth Knoll, Stephen Lock, and Jane Smith, each of whom I asked to review and grade all the abstracts we received so that the selection could be made as fairly as peer review allows.

Throughout these four years, Annette Flanagin has been enthusiastic, farsighted, dedicated, and good humored. The Congress owes more to her than to anyone. I cannot thank her enough.

It's been fun putting all this together. Once again — I hope you enjoy it as much as I.

Drummond Rennie
Director, Peer Review Congress

Peer Review Congress Advisory Board

Drummond Rennie
Director
Deputy Editor (West)
The Journal of the American Medical Association and
Adjunct Professor of
Medicine, Institute for Health
Policy Studies, University of
California, San Francisco

Annette Flanagin
N American Coordinator
Associate Senior Editor
The Journal of the American Medical Association

Jane Smith
European Coordinator
Deputy Editor
British Medical Journal

Marcia Angell
Executive Editor
The New England Journal of Medicine

John C. Bailar III
Science Advisor
National Academy of
Science

Thomas Chalmers
Associate Director
Harvard School of Public
Health

Suzanne Fletcher
Editor
Annals of Internal Medicine

Renee C. Fox
Annenberg Professor
of the Social Sciences
University of Pennsylvania

Charles C. Hancock
Executive Director
American Society for
Biochemistry and
Molecular Biology

Brian Haynes
Editor
ACP Journal Club

Elizabeth Knoll
Science Editor
University of California
Press

Daniel E. Koshland, Jr
Editor
Science

John Last
Editor
The Annals of the Royal College of Physicians and Surgeons of Canada

Stephen Lock
Editor
British Medical Journal,
1975-1991

George D. Lundberg
Editor
The Journal of the American Medical Association

Harry Marks
Institute for the History of
Medicine
Johns Hopkins University

Leo Offerhaus
Consultant
World Health Organization
European Office

Povl Riis
Editor
Ugeskr Laeger, 1966-1991

M. Roy Schwarz
Senior Vice President
American Medical
Association

David Sharp
Deputy Editor
The Lancet

Kenneth S. Warren
CEO
Comprehensive Medical
Services

Patricia Woolf
Lecturer
Department of Molecular
Biology
Princeton University

Table of Contents

| | |
|---|----|
| Welcome! | 1 |
| Peer Review Congress Advisory Board | 1 |
| Conference Program | 5 |
| Information About Continuing Medical Education | 28 |
| Plenary Session Abstracts | |
| Reliability of an Instrument for Assessing the Quality of Industry-Sponsored Symposia | 8 |
| Mildred Cho and Lisa Bero | |
| Evaluating Peer Reviews: Pilot Testing of a Grading Instrument | 8 |
| Irene Feurer, Gary Becker, Daniel Picus, Estella Ramires, Michael Darcy, and Marshall Hicks | |
| Overall Scientific Quality Is Improved by Peer Review Evaluation | 9 |
| Margaret Abby, Michael Massey, Susan Galandiuk, and Hiram C. Polk, Jr | |
| Evaluating the Quality of Articles Published in Journal Supplements Compared With Those Published in the Regular Journal Issue | 9 |
| P. A. Rochon, J. H. Gurwitz, C. M. Cheung, and T. C. Chalmers | |
| Effect of Peer Review and Editorial Changes on the Quality of Manuscripts Published in the <i>Annals of Internal Medicine</i> | 9 |
| Steven N. Goodman, Jesse A. Berlin, Suzanne W. Fletcher, and Robert H. Fletcher | |
| The Tasks of Peer Review: A Study and a Proposal | 10 |
| Erica Frank | |
| What Can We Learn by Studying the Association Between Referees' Recommendations and Editors' Decisions? | 10 |
| Lowell I. Hargens | |
| Effects of Peer Review and the Editorial Processes on the Readability of Original Articles | 10 |
| John C. Roberts, Robert H. Fletcher, and Suzanne W. Fletcher | |
| Do Journal Readers and Reviewers Agree on Manuscript Quality? | 11 |
| Amy C. Justice, Suzanne W. Fletcher, Robert H. Fletcher, Jesse A. Berlin, and Steven N. Goodman | |
| Characteristics of Peer Reviewers Who Produce Good Reviews | 11 |
| Arthur T. Evans, Robert A. McNutt, Suzanne W. Fletcher, and Robert H. Fletcher | |
| Inappropriate and Appropriate Selection of 'Peers' in Grant Review | 12 |
| Stanton A. Glantz | |
| Do We Need Structured Reporting of Randomized Controlled Trials? | 12 |
| David Moher, Corrine S. Dulberg, and George A. Wells | |
| Assessing the Quality of Randomization From Reports of Controlled Trials Published in Journals of Obstetrics and Gynecology | 12 |
| Kenneth F. Schulz, Iain Chalmers, David A. Grimes, and Douglas G. Altman | |
| The Utilization of Statistical Reviewers in Biomedical Journals | 13 |
| Steven N. Goodman, Stephen George, and Douglas G. Altman | |
| Transfer of Technology From Statistical Journals to the Biomedical Literature: Past Trends and Future Predictions | 13 |
| Douglas G. Altman and Steven N. Goodman | |
| Peer Review in Journals Indexed in <i>Index Medicus</i> | 13 |
| Lois Ann Colaianni | |

| | |
|---|-----------|
| How Well Does the Peer Review Process Function? A Quality Assurance Study From the Author's Perspective | 14 |
| Bobbie Jean Sweitzer and David J. Cullen | |
| Editorial Peer Review: Comparison of US Authors Publishing in Two Groups of Journals | 14 |
| Ann C. Weller | |
| Full Publication of Results Initially Reported in Abstracts: A Meta-analysis | 14 |
| Roberta W. Scherer and Kay Dickersin | |
| Completeness of Textbook Reviews in Medical Journals: A Look at a Neglected Area of Peer Review | 15 |
| Barbara Gastel | |
| Alternatives to Peer Review: An Interdisciplinary Perspective | 15 |
| Jonathan L. Entin | |
| A Case Study of Irresponsible Editorial Behavior | 15 |
| Martin Fichman and Selma Zimmerman | |
| How Should We Deal With Editorial Misconduct? | 16 |
| Douglas G. Altman, Iain Chalmers, and Andrew Herxheimer | |
| Fraudulent and Unethical Practices in Scientific Publishing | 16 |
| Herbert N. Nigg and Gabriela Radulescu | |
| The Scientific Community's Response to Evidence of Fraudulent Misconduct..... | 16 |
| William P. Whitely, Drummond Rennie, and Arthur W. Hafner | |
| Duplicate Publication of Original Articles in and From the <i>Nederlands Tijdschrift voor Geneeskunde</i> (<i>Dutch Journal of Medicine</i>) | 17 |
| Hans Barnard and John A. Overbeke | |
| Publication Bias and Public Policy | 17 |
| Lisa Bero and Drummond Rennie | |
| Effect of Institutional Prestige on Reviewers' Recommendations and Editorial Decisions..... | 17 |
| Joseph M. Garfunkel, Harvey J. Hamrick, Edward E. Lawson, and Martin H. Ulshen | |
| Is There Gender Bias in <i>JAMA's</i> Peer Review Process? | 18 |
| Julie Gilbert, Elaine Williams, and George D. Lundberg | |
| Multiple Blinded Reviews of the Same Two Manuscripts: Referee Characteristics, Language, and Methodologic Analyses | 18 |
| Povl Riis, Magne Nylenna, and Yngve Karlsson | |
| The Relative Efficiency of Double-blind vs Single-blind Reviewing: Empiric Evidence From Economics Journals | 19 |
| David N. Laband and Michael J. Piette | |
| The Effects of Blinding on Acceptance of Research Papers by Peer Review | 19 |
| Martin Fisher, Sanford B. Friedman, and Barbara Strauss | |
| The Effects of Blinding on Editors' Decision Making | 20 |
| Robert A. McNutt, Arthur T. Evans, Suzanne W. Fletcher, and Robert H. Fletcher | |
| Poster Session Abstracts | |
| So Much for Peer Review: Evaluation of Peer Review of Independently Submitted Manuscripts vs Manuscripts Previously Reviewed and Approved by a Society Program Committee | 20 |
| Arthur E. Baue and Donald L. Kaminski | |
| Effects of Peer Review and Editing on Readers' Perceptions of Research Manuscripts | 21 |
| Jesse A. Berlin, Steven N. Goodman, Robert H. Fletcher, and Suzanne W. Fletcher | |

| | |
|--|-----------|
| An Instrument for Assessing the Quality of Reporting of Clinical Research | 21 |
| Jesse A. Berlin, Steven N. Goodman, Suzanne W. Fletcher, and Robert H. Fletcher | |
| More of the Same: Referees' Egregious Mistakes | 21 |
| Juan Miguel Campanario | |
| An Analysis by Country of Scientific Reports in Five General Medical Journals..... | 22 |
| Edward W. Campion and Rachel Hart | |
| Rejection Rates in Biology Journals..... | 22 |
| Lisa DiDonato | |
| Incidence of Acceptance, Rejection, Resubmission, and Nonpublication of Full Reports of Research Presented at Academic Meetings..... | 22 |
| Lynn Dirk | |
| Are Surgical Research Publications Better Peer Reviewed Than Those in Standard Surgical Journals? | 23 |
| James T. Evans, Homayoun Attaran, Serrie Lico, and Patricia Carlin | |
| Source and Acknowledgment of Funding for Articles Published in Medical Journals..... | 23 |
| Timothy C. Fagan, Paula L. Carlson, and Robert J. Goldberg | |
| Editorial Accountability in Peer Review in Nursing Journals | 23 |
| Shirley H. Fondiller | |
| Authorial Selection Criteria for Manuscript Submission | 24 |
| Erica Frank | |
| Perceptions of Plagiarism in the Use of Other Authors' Language | 24 |
| Kell N. Julliard | |
| Standardizing Peer Review: Evaluating the Process With a Sham Paper | 24 |
| P. Dileep Kumar | |
| Readers of Peer-Reviewed Medical Journals: Do They Have a Message for Their Media? A Pilot Study ... | 25 |
| John C. Roberts, Robert H. Fletcher, Suzanne W. Fletcher, Kathleen Case, and Linda H. McClure | |
| Mixed Messages: Biomedical Writing and Peer Review Discourse | 25 |
| Gary L. Miranda | |
| How Long Does Peer Review Take? One Component of Publication Lag Time | 25 |
| Della J. Mundy and Joseph F. Terdiman | |
| Reasons for Rejection at the <i>Nederlands Tijdschrift voor Geneeskunde</i> (<i>Dutch Journal of Medicine</i>) in 1990 | 26 |
| Cees C. Kan, Jaques H. Lockefeer, and John A. Overbeke | |
| Peer Review: An Inquiry Among Referees of the <i>Nederlands Tijdschrift voor Geneeskunde</i> (<i>Dutch Journal of Medicine</i>) | 26 |
| Bouke C. H. De Jong and John A. Overbeke | |
| Mistakes in References in the <i>Nederlands Tijdschrift voor Geneeskunde</i> (<i>Dutch Journal of Medicine</i>) | 26 |
| Sjoerd O. Hobma and John A. Overbeke | |
| Publication Policies for Economic Analyses and Biomedical Ethics | |
| Deborah Roney, Daniel Sulmasy, and Kevin Schulman | |
| Review Articles: The Role of Peer Review | 27 |
| Bernard Schwartz and Susan Erickson | |
| Identification of Randomized Clinical Trials by Means of MEDLINE | 28 |
| Peri Schuyler, Kay Dickersin, Roberta Scherer, and Nancy Wright | |
| The Effect on the Outcome of Statistical Tests on the Decision to Publish and Vice Versa | 28 |
| Theodor D. Sterling, Wilf L. Rosenbaum, and James J. Weinkam | |

Conference Program

Thursday, September 9

All plenary sessions will be held in the International Ballroom. Each presentation will last ten minutes and will be followed immediately by ten minutes of discussion and audience participation.

- 7:30 AM Registration and Continental Breakfast—International Foyer Exhibits—State Room**
- 8:45 AM Welcome and Introduction**
James S. Todd, Executive Vice President, American Medical Association
George D. Lundberg, Editor, *JAMA*
Drummond Rennie, Deputy Editor (West), *JAMA*
- 9:10 AM Crude But Indispensable**
Jerome P. Kassirer, *The New England Journal of Medicine*
- 9:30 AM The Structural Transformation of the Sciences and the End of Peer Review**
Horace Freeland Judson, Program in History of Science, Stanford University

Peer Review and Quality Control

Moderator, Thomas Chalmers, Harvard School of Public Health

- 9:50 AM Reliability of an Instrument for Assessing the Quality of Industry-Sponsored Symposia**
Mildred Cho, University of California, San Francisco
- 10:10 AM Evaluating Peer Reviews: Pilot Testing of a Grading Instrument**
Irene D. Feuer, *Journal of Vascular and Interventional Radiology*
- 10:30 - 11:00 AM Break—International Foyer and Exhibits—State Room**
- 11:00 AM Overall Scientific Quality Is Improved by Peer Review Evaluation**
Margaret Abby, University of Louisville School of Medicine
- 11:20 AM Evaluating the Quality of Articles Published in Journal Supplements Compared With Those Published in the Regular Journal Issue**
Paula Rochon, University of Toronto
- 11:40 AM Effects of Peer Review and Editorial Changes on the Quality of Manuscripts Published in the Annals of Internal Medicine**
Steven Goodman, Johns Hopkins University School of Medicine
- 12:00 - 1:30 PM Luncheon—Imperial Ballroom and Exhibits—State Room**

Mechanisms of Peer Review

Moderator, Povl Riis, University of Copenhagen

- 1:30 PM The Tasks of Peer Review: A Study and a Proposal**
Erica Frank, Emory University School of Medicine
- 1:50 PM What Can We Learn by Studying the Association Between Referees' Recommendations and Editors' Decisions?**
Lowell Hargens, Department of Sociology, Ohio State University
- 2:10 PM Effects of Peer Review and the Editorial Processes on the Readability of Original Articles**
John Roberts, Johns Hopkins University School of Medicine
- 2:30 - 3:00 PM Break—International Foyer and Exhibits—State Room**

Mechanisms of Peer Review continued

Moderator, Povl Riis, University of Copenhagen

- 3:00 PM Do Journal Readers and Reviewers Agree on Manuscript Quality?**
Amy Justice, *Annals of Internal Medicine*
- 3:20 PM Characteristics of Peer Reviewers Who Produce Good Reviews**
Arthur Thomas Evans, University of North Carolina, Chapel Hill
- 3:40 PM Inappropriate and Appropriate Selection of 'Peers' in Grant Review**
Stanton A. Glantz, University of California, San Francisco
- 4:00 PM Thursday's Program Concludes**

Friday, September 10

All plenary sessions will be held in the International Ballroom. Each presentation will last ten minutes and will be followed immediately by ten minutes of discussion and audience participation.

**7:30 AM Continental Breakfast—International Foyer
Exhibits—State Room**

8:30 AM Information and Introduction
Drummond Rennie

Statistics and Peer Review

Moderator, Suzanne Fletcher, *Annals of Internal Medicine*

8:40 AM Do We Need Structured Reporting of Randomized Controlled Trials?
David Moher, Loeb Medical Research Institute, University of Ottawa

**9:00 AM Assessing the Quality of Randomization From Reports of Controlled Trials
Published in Journals of Obstetrics and Gynecology**
Iain Chalmers, UK Cochrane Centre

9:20 AM The Utilization of Statistical Reviewers in Biomedical Journals
Stephen George, Duke University

9:40 AM Transfer of Technology From Statistical Journals to the Biomedical Literature: Past Trends and Future Predictions
Douglas G. Altman, Imperial Cancer Research Fund,

10:00 - Break and Poster Sessions—International Foyer

11:00 AM 23 participants will present their work during open-forum sessions.
Exhibits—State Room

Scientific Evidence, Peer Review, and the Courts

Moderator, George D. Lundberg, *JAMA*

11:00 AM Robert Charrow, Esq, Kroll & Morrin
Robert Fletcher, *Annals of Internal Medicine*
Kenneth Rothman, *Epidemiology*

12:00 - Luncheon—Imperial Ballroom and Exhibits—State Room
1:30 PM

Other Aspects of Peer Review

Moderator, Stephen Lock, *British Medical Journal*, 1975-1991

1:30 PM Peer Review in Journals Indexed in Index Medicus
Lois Ann Colaianni, National Library of Medicine

1:50 PM How Well Does the Peer Review Process Function? A Quality Assurance Study From the Author's Perspective
David Cullen, *Journal of Clinical Anesthesia*

2:10 PM Editorial Peer Review: Comparison of US Authors Publishing in Two Groups of Journals
Ann C. Weller, Library of the Health Sciences, University of Chicago at Illinois

2:30 - Break and Poster Sessions—International Foyer
3:30 PM 23 participants will present their work during open-forum sessions.
Exhibits—State Room

Other Aspects of Peer Review continued

Moderator, Stephen Lock, *British Medical Journal*, 1975-1991

3:30 PM Full Publication of Results Initially Reported in Abstracts: A Meta-analysis
Roberta Scherer, University of Maryland School of Medicine

3:50 PM Completeness of Textbook Reviews in General Medical Journals: A Look at a Neglected Area of Peer Review
Barbara Gastel, Texas A&M University

4:10 PM Alternatives to Peer Review: An Interdisciplinary Perspective
Jonathan L. Entin, Case Western Reserve University School of Law

4:30 PM Friday's Program Concludes

6:00 PM Reception—American Medical Association Lobby, 515 N State St
Buses will transport registrants to and from the reception. Please meet at the B-2 level entrance of the Fairmont Hotel.
Buses will depart at 5:45 PM and every 15 minutes thereafter. Please be sure to wear your Congress name badge.

Saturday, September 11

All plenary sessions will be held in the International Ballroom. Each presentation will last ten minutes and will be followed immediately by ten minutes of discussion and audience participation.

**7:30 AM Continental Breakfast—International Foyer
Exhibits—State Room**

8:45 AM Information and Introduction
Drummond Rennie

Scientific Misconduct and Peer Review

Moderator, Drummond Rennie, JAMA and University of California, San Francisco

8:50 AM *A Case Study of Irresponsible Editorial Behavior*
Selma Zimmerman, Division of Natural Science, York University

9:10 AM *How Should We Deal With Editorial Misconduct?*
Douglas G. Altman, Imperial Cancer Research Fund

9:30 AM *Fraudulent and Unethical Practices in Scientific Publishing*
Herbert N. Nigg, *Bulletin of Environmental Contamination and Toxicology*

9:50 AM *The Scientific Community's Response to Evidence of Fraudulent Misconduct*
William P. Whitley

10:10 AM *Duplicate Publication of Original Articles in and From
the Nederlands Tijdschrift voor Geneeskunde (Dutch Journal of Medicine)*
John Overbeke, *Dutch Journal of Medicine*

10:30 - 11:00 AM Break—International Foyer and Exhibits—State Room

Bias and Peer Review

Moderator, Marcia Angell, *The New England Journal of Medicine*

11:00 AM *Publication Bias and Public Policy*
Lisa Bero, University of California, San Francisco

11:20 AM *Effect of Institutional Prestige on Reviewers' Recommendations and Editorial Decisions*
Joseph M. Garfunkel, *Journal of Pediatrics*

11:40 AM *Is There Gender Bias in JAMA's Peer Review Process?*
Julie Gilbert, Stanford University

12:00 - 1:30 PM Luncheon—Imperial Ballroom and Exhibits—State Room

Blinded Peer Review

Moderator, Elizabeth Knoll, University of California Press

1:30 PM *Multiple Blinded Reviews of the Same Two Manuscripts:
Referee Characteristics, Language, and Methodologic Analyses*
Povl Riis, University of Copenhagen

1:50 PM *The Relative Efficiency of Double-blind vs Single-blind Reviewing:
Empiric Evidence From Economics Journals*
David N. Laband, Perdue School of Business, Salisbury State University

2:10 PM *The Effects of Blinding on Acceptance of Research Papers by Peer Review*
Martin Fisher, Cornell University Medical College

2:30 PM *The Effects of Blinding on Editors' Decision Making*
Robert McNutt, University of North Carolina, Chapel Hill

2:50 PM Wrap-up
Drummond Rennie

3:00 PM Saturday's Program Concludes

Plenary Session Abstracts

Plenary session abstracts are listed in order of presentation on pages 8 through 20. Asterisks indicate the names of presenters. Affiliations are listed for each author and an address is provided for the corresponding author.

Peer Review and Quality Control

Reliability of an Instrument for Assessing the Quality of Industry-Sponsored Symposia

*Mildred Cho and Lisa Bero

Institute for Health Policy Studies, University of California, San Francisco, 1388 Sutter St, 11th Floor, San Francisco, CA 94109

Objective: To determine the interrater reliability and content validity of an instrument to assess the quality of methods used in drug studies published in the biomedical literature. The instrument will be used to compare the quality of drug studies published as proceedings of pharmaceutical industry-sponsored symposia with the quality of matched articles published in peer-reviewed medical journals.

Design: An existing quality assessment instrument (Spitzer et al. *Clin Invest Med*. 1990;13:17-42) was modified for our study because it assessed the quality of a variety of study designs ranging from case reports to randomized controlled trials. Seven reviewers were recruited to use the instrument to assess three randomly selected symposium articles and three peer-reviewed articles matched by date of publication, journal, and class of drug used in the study. Reviewers were blinded to (1) the hypothesis of our study, (2) the authors, journal names, and dates of the articles, and (3) whether the articles were from symposia or not. Reviewers were given written instructions only. The 16-item instrument assesses the quality of the study design, subject selection, sample size, blinding, statistical analysis, and validity of conclusions. Content validity was determined by asking the reviewers about each question, and interrater reliability was measured by percentage agreement and κ values for four possible responses ("no," "partial," "yes," and "not applicable"). Furthermore, an overall quality score for each article was calculated. We tested the effects of three different weighting schemes, roughly based on work by Chalmers et al (*Controlled Clin Trials*. 1981;2:31-49), on the index of correlation of the quality scores.

Results: Overall, reviewers found the content of the instrument to be valid and suggested minor changes to four of 16 items. For seven items, five of seven reviewers gave identical responses on four or more articles. Six items had low percentage agreement and κ scores. Poor reliability was almost always due to ambiguity in the use of the responses "partial" or "not applicable." Possible quality scores ranged from 0 to 1, with 1 representing the best score. For six articles, the SD of mean scores was lowest (0.16) and index

of correlation (W) highest (.73; $P < .01$) when we used weighting scheme 3, which gave more weight to questions on study design, randomization, accounting for bias, and validity of conclusions. Weighting scheme 1 gave equal weight to all questions, and weighting scheme 2 gave more weight to study design only.

Conclusions: We have designed a valid and reliable instrument to assess the quality of drug studies. The instrument is applicable to a variety of research designs. We will be able to determine whether characteristics of symposia (ie, type of sponsorship) are associated with quality.

Evaluating Peer Reviews: Pilot Testing of a Grading Instrument

*Irene Feurer, Gary Becker, Daniel Picus, Estella Ramires, Michael Darcy, and Marshall Hicks

Journal of Interventional Radiology, Miami Vascular Institute, Baptist Hospital of Miami, 8900 N Kendall Dr, Miami, FL 33176

Objective: This study measured the reliability and preliminary validity of an instrument, developed and to be used by editors, to evaluate review quality.

Design: Attributes were scored as follows (possible points/percentage contribution): timeliness (3/21%); summary form (1/7%); etiquette (1/7%); sectional narratives (3/21%); citations (2/14%); narrative summary (2/14%); and insights (2/14%). Three graders (one editor and two reviewers) evaluated reliability. Two editors evaluated validity. The sample (53 reviews of 23 manuscripts) included a cohort of manuscripts mailed to reviewers during 2 months such that potential score ranges were maximal, reviewers were unique, and graders' reviews were excluded. Subsamples were randomly constructed for the reliability ($n=41$; power $> .85$) and validity ($n=12$) analyses. An explicit scoring guide was provided. Graders were privy to reviewers' identities to simulate the instrument's intended use and because scoring of "etiquette" would be confounded by blinding. Grading sequence was random, and graders were requested to disregard reviewers' identities. Statistical analyses included analysis of variance (ANOVA), Kendall's coefficient of concordance (W), correlation (r), and χ^2 .

Results: Data are given for total score, but conclusions are comparable for relevant subscores and transformations. Reliability was demonstrated by ANOVA (group mean \pm SD): 7.8 ± 2.50 ; 7.8 ± 2.27 ; 7.22 ± 2.53 ; $P=.46$. Validity was preliminarily confirmed ($W=.94$, $P<.05$; $r=.91$, $P<.001$). Agreement between reviewers' basic recommendations (accept/reject) and editors' decisions was moderate ($n=53$, $r=.55$, $\chi^2=13.93$, $P<.001$) and unrelated to reviewers' average scores ($n=53$, $r=.13$, $P=.37$).

Conclusions: This instrument is reliable. Its validity is

suggested, and further validation is warranted. Potential refinements and applications include identifying scores and criteria that discriminate among editorial board candidates, helpful reviewers, and minimally helpful reviewers.

Overall Scientific Quality Is Improved by Peer Review Evaluation

*Margaret Abby, Michael Massey, Susan Galandiuk, and Hiram C. Polk, Jr

Department of Surgery, University of Louisville School of Medicine, Louisville, KY 40292, and the *American Journal of Surgery*, New York, NY

Objective: Peer review is intended to improve the quality of published scientific data through the evaluation process. Our goal was to determine what percentage of manuscripts rejected by the *American Journal of Surgery* (AJS) in 1989 were published subsequently, how soon after rejection, and where. Primary reasons for rejection were also noted. We postulated that the majority of manuscripts rejected by AJS would be published in another journal within 3 years after rejection or, rarely, had already been published.

Design: We conducted a MEDLINE search from 1984 through 1992 of work published by the primary author of the rejected manuscript. We excluded manuscripts that were submitted to AJS by international authors because of greater difficulty in tracking subsequent publication.

Results: One hundred twenty-five consecutive North American manuscripts were rejected by AJS in 1989, but only 38% of them were published subsequently. The average duration between AJS rejection and later publication was only 17 months. Of those subsequently published, 54% went to general surgical journals, including 12% that were later reevaluated and accepted by AJS. Twenty-nine percent of the rejected manuscripts were published in specialty medical journals, 10% in state and local journals, and the remainder in general medical journals. Twenty-eight percent of the authors of rejected manuscripts were noted to have published articles on similar subjects repeatedly.

Conclusions: Peer review evaluation is probably helpful in improving clarity and quality of ultimately published manuscripts. These findings reflect more favorably on the peer review process than did previous studies in the sense that some poor work is excluded from the medical literature.

Evaluating the Quality of Articles Published in Journal Supplements Compared With Those Published in the Regular Journal Issue

*P. A. Rochon,¹ J. H. Gurwitz,² C. M. Cheung,¹ and T. C. Chalmers²

¹Baycrest Centre for Geriatric Care, 3560 Bathurst St, North York, Ontario, Canada M6A 2E1; and ²Harvard University, Boston, Mass

Objective: The relationship between the quality of articles and whether they were published in regular journal or supplement issues were evaluated.

Design: We identified all randomized controlled trials (RCTs) of drug therapies in adults between January 1990 and June

1992 by means of a MEDLINE search of the journals publishing the most supplements (*American Heart Journal*, *American Journal of Cardiology*, and *American Journal of Medicine*). A total of 503 abstracts were reviewed; 319 abstracts met initial inclusion criteria. A further 61 were reexcluded: non-RCTs ($n=43$), drug washout ($n=7$), unobtainable ($n=5$), non-drug related ($n=4$), and study of children ($n=2$). Reviewers blinded to supplement status assessed the quality of each of the remaining 258 articles.

Results: Seventy-six articles (29.5%) were published in supplements. The majority of therapies were cardiovascular (186 [72.1%]) and antibiotic (24 [9.3%]). Quality scores, using a standardized scoring system of methods and results, were lower in articles published in supplements ($t[253]=2.79$, $P=.006$) than in articles published in the regular journal issue. Supplements less frequently acknowledged a biostatistician ($\chi^2=7.79$, $df=1$, $P=.005$) or government or foundation support ($\chi^2=8.16$, $df=1$, $P=.004$) and were more likely to be sponsored by a pharmaceutical manufacturer ($\chi^2=28.14$, $df=1$, $P=.00001$). Articles published in supplements had greater discrepancies between the number of patients randomized and those analyzed ($t[224]=2.64$, $P=.009$).

Conclusions: We demonstrated that RCTs published in journal supplements are of inferior quality to those in regular issues of the same journals. Questions about the relative value of journal supplement articles are raised.

Effect of Peer Review and Editorial Changes on the Quality of Manuscripts Published in the *Annals of Internal Medicine*

*Steven N. Goodman,¹ Jesse A. Berlin,² Suzanne W. Fletcher,³ and Robert H. Fletcher³

¹Johns Hopkins University, Oncology Center, Division of Biostatistics, 550 N Broadway, Suite 1103, Baltimore, MD 21205; ²Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia; and ³*Annals of Internal Medicine*, Philadelphia

Objective: To evaluate the effects of peer review and subsequent editorial changes on manuscript quality.

Design: Masked before-after study of 104 original research manuscripts submitted to the *Annals of Internal Medicine* between March 1992 and March 1993 and accepted for publication. A manuscript quality assessment tool of 35 questions was used by 44 expert assessors to evaluate each manuscript as it was at the time of submission and also after revision and editorial changes based on peer review. Assessors were unaware of the design or aims of the study, and different persons evaluated the two versions of each manuscript.

Results: The average of item scores improved by 0.24 point on a five-point scale (95% confidence interval, 0.08 to 0.40), and the percentage of items assigned a score of 3 or higher increased by an absolute 7.3% (confidence interval, 3.3% to 11.3%) from a baseline mean of 75%. The subjective summary score did not change significantly ($P=.2$). After correction for regression to the mean, the bottom 50% of manuscripts showed twofold to threefold larger improvements than did the top 50%. The largest effects were seen in

the discussions of study limitations and in the appropriateness of the strength and tone of the conclusions.

Conclusions: Peer review and editorial changes improve manuscript quality. Reviewers' summary judgments may be of less value for editors than their ratings of specific aspects of a manuscript.



Mechanisms of Peer Review

The Tasks of Peer Review: A Study and a Proposal

*Erica Frank

Department of Community and Preventive Medicine and Department of Medicine, Emory University School of Medicine, 69 Butler St SE, Atlanta, GA 30303-3219

Objective: Despite the pivotal role of peer review in the scientific process, there has never been a formal study of the way in which reviewers' opinions are elicited. This paper describes such a study and makes recommendations for a more standardized approach to the peer review process.

Design: The study population included the top 100 journals as rated by the 1989 Institute for Scientific Information citation frequency index. Only journals with primarily US-based editorial offices were included (n=73).

Results: The response rate for this study was 97.3%. Journals varied substantially in many of their reviewer requests. While 96% of journals asked reviewers to recommend acceptance or rejection, only 72% asked reviewers to assess manuscripts' novelty, 69% requested assessments of clarity, and only 51% asked for assessments of the reasonableness of manuscripts' conclusions. Only 46% of journals reminded reviewers that manuscripts were confidential documents, only 51% provided a separate cover letter, and only 25% provided extensive (one full page or more) reviewer instructions.

Conclusions: While review uniformity is not requisite, differences between review protocols seem to be a function not of specific journals' needs but of some journals' excluding important components of reviews. Based in part on the Council of Biology Editors' recommendations, standard components of reviews and of instructions for reviewers are recommended.

What Can We Learn by Studying the Association Between Referees' Recommendations and Editors' Decisions?

*Lowell I. Hargens

Department of Sociology, 300 Bricker Hall, 190 N Oval Mall, Ohio State University, Columbus, OH 43210

Objective: To show how researchers can use data on the bivariate association between referees' recommendations for a sample of submitted papers and editors' subsequent decisions about those papers (1) to determine the number of latent dimensions needed to account for the observed association, (2) to estimate distances between the categories of each variable, and (3) to test statistical hypotheses about differences between the nominally distinct categories constituting each variable.

Design: Data on cohorts of papers submitted to a biologic science journal (*Physiological Zoology*) and a behavioral science journal (*American Sociological Review*) were examined. Bivariate frequency tables describing the association between referees' assessments and editors' decisions at each journal were analyzed by means of Goodman's RC Association Model II.

Results: Despite large differences in these journals' average referee recommendations and final editorial dispositions, one latent dimension, presumably scholarly merit, can account for the association between referee recommendations and editorial decisions at both journals. The data also show that the referees' "revise and resubmit" recommendations play a different role at the two journals. At the biologic science journal it is more akin to the "accept conditionally" verdict, while at the behavioral science journal it resembles the "reject" verdict. At both journals, editors' "revise and resubmit" decisions are similar to acceptances and quite different from rejections.

Conclusions: Goodman's analysis of association methods hold promise both for studying the structure of journal editorial decision making and for improving peer review systems of manuscript evaluation.

Effects of Peer Review and the Editorial Processes on the Readability of Original Articles

*John C. Roberts,¹ Robert H. Fletcher,² and Suzanne W. Fletcher²

¹Johns Hopkins University School of Medicine, Francis Scott Key Medical Center, 4940 Eastern Ave, B-2 North, Baltimore, MD 21224; and ²*Annals of Internal Medicine*, Philadelphia, Pa

Objective: To measure the effect of the peer review and editorial processes on the readability of original articles.

Design: We assessed the readability of 101 consecutive manuscripts reporting original research submitted to the *Annals of Internal Medicine* between March and December 1992. We studied each manuscript on receipt and after it had passed through the peer review and editorial processes. Using the formulas of Flesch and Gunning, we analyzed

manuscripts for readability and length. (Both formulas are based on sentence length and syllables per 100 words.) Text and abstracts were analyzed similarly but separately. Mean readability scores and word counts of the initial and final manuscripts were compared by means of two-tailed *t* tests for paired observations.

Results: Initial readability scores of manuscripts and abstracts were 17.12 and 16.66, respectively. (Units roughly correlate to years of education.) At publication, scores were 16.82 ($P=.001$) and 15.66 ($P=.0001$). Most of the improvement in readability was attributable to shorter sentence length. These readability scores correspond to "very difficult" to read. By comparison, studies of other print media show scores of about 11 for the *New York Times* editorial page, about 18 for a typical legal contract, and 14.37 for this abstract. Manuscript length averaged 3190.7 words before and 3190.5 words after manuscripts were reviewed and edited. Abstracts, initially 277.2 words, shrank by 0.9 word.

Conclusions: Readability scores of abstracts and manuscripts were rated "very difficult" before the peer review and editorial processes, and they were only slightly better afterward. Previous studies have shown readability scores to correlate with readership, which suggests that if readability scores of medical journals were lower, clinicians' readership of journals might increase.

Do Journal Readers and Reviewers Agree on Manuscript Quality?

*Amy C. Justice,¹ Suzanne W. Fletcher,¹ Robert H. Fletcher,¹ Jesse A. Berlin,² and Steven N. Goodman³

¹*Annals of Internal Medicine*, Sixth and Race streets, Philadelphia, PA 19106-1572; ²Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia; and ³Johns Hopkins University School of Medicine, Baltimore, Md

Objective: General medical journals serve both clinicians and researchers. Peer review is done by researchers with expertise in the manuscript's content or methodologic area. However, little is known about how researchers' judgments of manuscript quality compared with those of readers, who are primarily clinicians. This study assessed whether the quality assessments of clinician-readers (readers) agreed with those of peer reviewers (peers) and expert methodologists (experts) for manuscripts selected for publication in the *Annals of Internal Medicine*.

Design: From March 1992 through March 1993, 104 consecutive original research manuscripts selected for publication were assessed by four readers, two peers, and two experts. Each assessor completed an instrument that included a 10-point ordinal scale of overall manuscript quality. Each reader or peer assessed only one manuscript, while each expert evaluated six manuscripts.

Results: Eighty-eight percent of reader, 95% of peer, and 100% of expert assessments were returned. Readers and peers gave high scores (77% and 73% above 5, respectively), while experts were more critical (52% of scores above 5, $P<.0001$). When scores were assessed by manuscript, opinions varied widely within and between reader, peer, and expert groups (in all cases, agreement $<35\%$, $\kappa<0.1$). Among experts, who assessed multiple manuscripts, scores were more related to

the expert than to the manuscripts ($P=.03$).

Conclusions: Readers and peers are generally satisfied with the quality of manuscripts selected for publication in the *Annals of Internal Medicine*, while methodologic experts are more critical. For a given manuscript, opinions about quality vary widely both within and across reviewer groups. Among expert methodologists, global quality scores are more related to the expert than to the manuscript.



Characteristics of Peer Reviewers Who Produce Good Reviews

*Arthur T. Evans,¹ Robert A. McNutt,¹ Suzanne W. Fletcher,² and Robert H. Fletcher²

¹Cecil G. Sheps Center for Health Services Research, 725 Airport Rd, Suite 210, CB #7590, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7590; and ²*Annals of Internal Medicine*, Philadelphia, Pa

Objective: To determine the characteristics of good peer reviewers.

Design: Cross-sectional analysis of data gathered during a randomized controlled trial at the *Journal of General Internal Medicine*. The sample included 226 reviewers of 131 consecutively submitted manuscripts of original research. The quality of each review was judged on a scale of 1 to 5 by an editor who was blinded to the identity of the reviewer. Reviewer characteristics were taken from the curricula vitae.

Results: Of the reviewers, 201 (91%) completed the review and submitted a curriculum vitae. Eighty-six (43%) of the 201 reviewers produced good reviews (a grade of 4 or 5). Logistic regression showed that when a reviewer was less than 40 years old, from a top academic institution, well known to the editor choosing the reviewer, and blinded to the identity of the manuscript's author, the probability that he or she would produce a good review was 87%. A reviewer without any of those characteristics had a 7% probability of producing a good review. Other characteristics that were significant only on bivariate analysis included previous clinical research training, additional postgraduate degrees, and more time spent on the review. There was a negative but statistically nonsignificant association between academic rank and review quality: 37% of full professors and 51% of fellows or assistant professors produced good reviews ($P=.11$).

Conclusions: Good peer reviewers for this journal tended to be young, from strong academic institutions, well known to the editors, and blinded to the identity of the manuscript's authors.

Inappropriate and Appropriate Selection of 'Peers' in Grant Review

*Stanton A. Glantz

Division of Cardiology, Box 0124, University of California, San Francisco, CA 94143-0124

Objective: To assess the California Tobacco Related Disease Research Program (TRDRP) Behavioral and Public Health Research on Tobacco Study Section members as "peers" to review tobacco policy research.

Design: MEDLINE was searched for the years 1989 through 1993 with the use of the key word *tobacco* for TRDRP reviewers. The National Heart, Lung and Blood Institute (NHLBI) Cardiovascular Study Section, which reviews my NHLBI grant on ventricular function, was searched as a control using the key word *heart*.

Results: Six of 18 TRDRP reviewers (33%) had no citations. The median number of citations was two (interquartile range, zero through four). I had seven publications indexed to "tobacco"; only one member (5%) had more than seven publications indexed on "tobacco." The members' "tobacco" publications concentrated on well-controlled experimental interventions on smoking cessation and prevention strategies. Only one member had primary expertise in tobacco policy research. All 31 (100%) of the NHLBI reviewers had "heart" publications. The median number of citations was nine (interquartile range, seven through 19). I had 11 publications indexed to "heart"; 13 members (42%) had more than 11 publications. Five members had a primary interest in the subject of my NHLBI application.

Conclusions: Members' professional interests play a critical role in the level of interest and enthusiasm they will have for a proposal, which directly affects the priority score. The TRDRP's primary focus is experimental intervention; the single member on tobacco control does not overcome the biases intrinsic in this membership. A similar situation exists at the National Cancer Institute, where the Behavioral Medicine Study Section has not funded one of the last 41 applications for tobacco control research. The membership of these review committees has effectively precluded research on tobacco control policy.

classify a trial as negative and to determine primary outcomes were developed. All RCTs published in three leading general medical journals published in 1975, 1980, 1985, and 1990 were reviewed.

Results: There were 102 negative RCTs published during this 15-year review. Of these, only 33 (32%) reported calculating sample size. Although this number is small, the situation has improved over time, from 0% in 1975 to 43% in 1990. Less than half of the trials reported the control group event rate (45%). Eighteen percent reported whether the alpha was one or two tailed, and no trial reported the statistic on which sample size calculations were based. Only 20 trials made any kind of statement of the clinical significance of the results with respect to the observed differences.

Conclusions: There has been a considerable increase in the number of RCTs published during the period of this review. Our results indicate that there is still room for improvement in the quality of reporting of many statistical issues. Structured reporting of pertinent information in RCTs might improve communication in the same way as have structured abstracts.

Assessing the Quality of Randomization From Reports of Controlled Trials Published in Journals of Obstetrics and Gynecology

Kenneth F. Schulz,¹ *Iain Chalmers,¹ David A. Grimes,² and Douglas G. Altman³

¹UK Cochrane Centre, Summertown Pavilion, Middle Way, Oxford OX2 7LG, England; ²Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco; and ³Medical Statistics Laboratory, Imperial Cancer Research Fund, London, England

Objective: To assess the extent to which reports of controlled trials in journals of obstetrics and gynecology provide details of the way in which trial participants were allocated to treatment comparison groups.

Design: Survey of all reports of parallel group trials (in which allocation was stated to have been randomized) published in the 1990 and 1991 volumes of four journals of obstetrics and gynecology.

Results: Of 206 reports identified, only 32% gave an acceptable account of the method used to generate random numbers, and only 23% contained information showing that steps had been taken to conceal assignment until the point of treatment allocation. In 125 reports in which hypothesis tests were used to compare baseline characteristics, only 2% were statistically significant at the 5% level. In 96 reports of trials using unblocked randomization, the differences in sample sizes between treatment and control groups were much smaller than would be expected by chance ($P < .0001$).

Conclusions: Proper randomization is required to generate unbiased comparison groups in clinical trials. Only 10% of the reports of trials published in a sample of obstetrics and gynecology journals described proper methods of randomization. Additional analyses suggest that there may have been both selective reporting of baseline comparisons and non-random manipulation of comparison groups by investigators.

Statistics and Peer Review

Do We Need Structured Reporting of Randomized Controlled Trials?

*David Moher, Corrine S. Dulberg, and George A. Wells

Clinical Epidemiology Unit, Loeb Medical Research Institute, Ottawa Civic Hospital, 1053 Carling Ave, Ottawa, Ontario, Canada K1Y 4E9

Objective: To examine whether there has been a change in the reporting of statistical power and sample size calculations in published negative randomized controlled trials (RCTs) during the last 15 years.

Design: A structured data collection form and guidelines to

The Utilization of Statistical Reviewers in Biomedical Journals

Steven N. Goodman,¹ *Stephen George,² and Douglas G. Altman³

¹Johns Hopkins University, Oncology Center, Division of Biostatistics, 550 N Broadway, Suite 1103, Baltimore, MD 21212; ²Duke University Medical Center, Durham, NC; and ³Medical Statistics Laboratory, Imperial Cancer Research Fund, London, England

Objective: To document the degree of use and nature of involvement of statistical reviewers in the editorial process of major biomedical journals.

Design: A mailed survey of the editors of approximately 175 biomedical journals that publish original clinical research, chosen on the basis of high impact factor or high clinical readership.

Results: Questions covered by the survey include the level of involvement of the reviewer in the editorial process, the process by which statistical reviewers are chosen, and how often the editor judges statistical review to have produced important changes in manuscripts. The answers to these questions will be related to the journal's type (eg, general medicine, subspecialty), impact factor, circulation, selectivity, and total annual number of original research articles. The results of this survey will be compared with those of a similar survey conducted in 1983 of 100 biomedical journal editors.

Conclusions: For many researchers, journals set the de facto standards for methodologic sophistication and soundness. The manner in which these standards are set and who sets them is therefore of interest to readers, authors, and editors.

Transfer of Technology From Statistical Journals to the Biomedical Literature: Past Trends and Future Predictions

*Douglas G. Altman¹ and Steven N. Goodman²

¹Medical Statistics Laboratory, Imperial Cancer Research Fund, PO Box 123, 61 Lincoln's Inn Fields, London WC2A 3PX, England; and ²Johns Hopkins University, Baltimore, Md

Objective: To investigate the speed and nature of the transfer of new statistical methods into the medical literature and, on the basis of current data, to predict what methods medical journal editors should expect to see in the next decade.

Design: Influential statistical papers were identified from lists of the most cited and most influential papers and from informed opinion. However, for several important methods, there is no single seminal paper. For each paper, the time pattern of citations in the medical literature has been ascertained. Similar analyses have been performed for a few expository statistical papers published in medical journals. In addition, longitudinal studies of the statistical content of papers in medical journals have been reviewed to examine changes in methodology over time.

Results: Annual citations show weak evidence of decreasing lag times between the introduction and widespread use of new statistical methods. Newer technical innovations still typically take four to six years before they achieve 25 citations in the medical literature. Few methodologic

advances of the 1980s seem yet to have been widely cited in medical journals. By contrast, expository papers in medical journals can reach 500 citations within four to five years. Longitudinal studies indicate a large increase since 1979 in the use of more complex methods, notably survival analysis, logistic regression, and Mantel-Haenszel methods.

Conclusions: Time trends suggest that technology diffusion has speeded up over the last 30 years, although there is still a lag of several years before medical citations begin to accrue. Several more complex statistical methods introduced during the 1980s are beginning to be seen more frequently. Although it is not possible to identify recent papers that will turn out to be major breakthroughs, journals should expect to see growing numbers of papers using increasingly sophisticated methods (often computer-intensive methods), such as bootstrapping, Bayesian methods (especially Gibbs' sampling), generalized additive methods, and new methods for analyzing longitudinal and hierarchical data. Statistical reviewers may find it difficult to assess whether these methods have been applied appropriately and correctly.

Other Aspects of Peer Review

Peer Review in Journals Indexed in *Index Medicus*

*Lois Ann Colaizzi

Library Operations, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894

Objective: To determine the proportion and types of journals indexed in *Index Medicus* that were peer reviewed and to what extent peer review can be ascertained by examining an issue of a journal title.

Design: The first issue in 1992 of a sample of English-language journals indexed in *Index Medicus* was examined for statements and evidence of peer review. Also examined were all indexed English-language titles in four subject areas. Such evidence as multiple copies of manuscripts and dates of receipt, revision, and acceptance was noted. A questionnaire to determine the actual peer review process was sent to the editors in the subject areas. Actual practice was noted and correlated with the evidence and statements found in the journal.

Results: Responses have been received from the editors of 66% of the journals in three subject areas. Clear evidence of peer review was found in 50% of these. The most common location for a peer review statement was the instructions to authors. Editors reported not printing a peer review statement in a third of the titles. External reviewers were reported to be

used for peer review of research manuscripts in 84% of the replies; 43% of the editors reported using external reviewers for reviewing invited manuscripts.

Conclusions: Editors' self-reported peer review practices vary, depending on the type of article, and cannot be reliably ascertained from evidence printed in the journal.

How Well Does the Peer Review Process Function? A Quality Assurance Study From the Author's Perspective

Bobbie Jean Sweitzer and *David J. Cullen

Department of Anesthesia, Massachusetts General Hospital, Boston, MA 02114

Objective: The peer review process for evaluating scientific manuscripts is critical to maintain high standards for publication, though information about peer review is lacking. Therefore, we surveyed the authors of manuscripts submitted to a peer-reviewed general anesthesiology journal, the *Journal of Clinical Anesthesia (JCA)*, whether or not their manuscripts were accepted, to determine their satisfaction with the review process.

Design: Authors of all unsolicited original manuscripts submitted to the *JCA* in 1991 and 1992 received anonymous questionnaires. In addition to satisfaction with specific issues, six items were judged by means of a five-point ordinal scale: timeliness of review, quality of review, manuscript improvement, whether the process was educational, whether the article was modified after the review, and specificity of the reviewers' comments. Authors were also asked to comment on ways to improve the process, and if the initial review influenced subsequent submissions. Results were analyzed with analysis of variance and χ^2 statistics.

Results: Of 209 authors, 95 (45%) responded. The status of their manuscripts was as follows: accepted with revision, 69 manuscripts (46 responses; 67%); rejected but may resubmit after revision, 54 manuscripts (23 responses; 43%); and rejected outright, 86 manuscripts (26 responses; 30%). More authors of manuscripts accepted with revision responded to our questionnaire ($P < .001$ by χ^2) and had slightly more favorable opinions of the review process than did authors whose manuscripts were rejected with a possibility of revision or rejected outright. Rejected authors were less likely to modify their articles or sense improvement in their manuscripts as a result of peer review. More authors were satisfied than dissatisfied with the review process regardless of their manuscripts status. Authors with manuscripts accepted with revision chose more satisfying characteristics than did authors whose manuscripts were rejected with a possibility of revision or rejected outright and were more likely to offer positive comments.

Conclusions: This survey informed us about many aspects of our journal's peer review process and gave us useful feedback to retain the benefits and modify unfavorable aspects. Surveying authors, the key consumers in the peer review process, is both feasible and useful.

Editorial Peer Review: Comparison of US Authors Publishing in Two Groups of Journals

*Ann C. Weller

Library of the Health Sciences, University of Illinois at Chicago, 1750 W Polk St, Chicago, IL 60612

Objective: To compare authors' experiences with editorial peer review practices of two groups of indexed US medical journals.

Design: Group 1 journals met two criteria: (1) they were indexed in 1992 *Index Medicus* and (2) they met several reputational measures. Group 2 journals met criterion 1 only. Surveys were sent to the first authors of 616 randomly selected articles (approximately 300 per group). Questions focused on the target article.

Results: The response rate was 75%. Group 1 manuscripts were somewhat more likely ($P < .06$) to be peer reviewed than were those in group 2. About 20% of group 1 and 16% of group 2 articles had been previously rejected. Group 1 journals were more likely ($P < .02$) to accept manuscripts previously rejected by another group 1 journal. Group 1 authors were more likely ($P < .05$) to select a journal for its prestige, article quality, and readership, as well as to subscribe to, have previously published in, and reviewed for that journal. Group 2 authors were more likely ($P < .05$) to have been invited to submit their manuscripts. For both groups, the review process only marginally improved content, organization, or statistical analysis or clarified conclusions. More than 50% of both groups believed the review offered constructive suggestions, but conclusions were seldom changed. Between 3% and 15% of all authors experienced considerable conflicting advice, reviewers' misunderstanding of their manuscripts, or biased review.

Conclusions: Authors from both groups had similar experiences and concerns with peer review; they differed on reasons for journal selection, their connections with the publishing journal, and patterns of resubmission after rejection.

Full Publication of Results Initially Reported in Abstracts: A Meta-analysis

*Roberta W. Scherer and Kay Dickersin

Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, 419 W Redwood St, Suite 460, Baltimore, MD 21012

Objective: To combine data from available studies to determine the rate at which abstracts are subsequently published in full-length form and the association between selected study characteristics and full publication.

Design: Data were combined from seven separate studies, representing seven medical specialties, that described follow-up of abstracts to learn of subsequent full publication. Abstracts selected for the seven studies were associated with presentations at a meeting, were published in meeting proceedings, or appeared in a specified sample of the

literature. All studies examined the rate of full publication, allowing at least 27 months for a study to achieve this. Factors investigated in the combined analysis for possible association with publication were “statistically significant” or positive results and sample size. Such associations would indicate a publication bias. Sample size for an abstract was defined in terms of being above or below the median sample size for the group of abstracts described in a single study. Combined odds ratios (ORs) were calculated by the Mantel-Haenszel method.

Results: Combined results from all seven studies showed that only 46.6% (685/1471) of all abstracts were subsequently published in full. Pooled results of three studies revealed a positive association between “statistically significant” or positive results and publication (OR, 1.43; 95% confidence interval, 0.97 to 2.09). Combined results of two studies indicated a positive association between studies with sample size above the median and publication (OR, 2.13; 95% confidence interval, 1.29 to 3.54).

Conclusions: Less than half of all studies presented in abstract form are subsequently published as full-length reports. Previous presentation of results in abstract form does not appear to protect against publication bias.

Completeness of Textbook Reviews in Medical Journals: A Look at a Neglected Area of Peer Review

*Barbara Gastel

Department of Journalism, Texas A & M University, College Station, TX 77843-4111

Objective: Book reviews in medical journals are readily accessible to examination but have received little systematic assessment. In the hope of improving this little-investigated area of peer review, an exploratory study was performed on the completeness of text-book reviews in general medical journals.

Design: All book reviews published in *JAMA* or *The New England Journal of Medicine* during July through December 1991 were obtained, and the reviews of general textbooks in clinical or basic science disciplines were identified. These textbook reviews were assessed for the presence of information in seven areas: (1) scope of the book’s content, (2) quality of content, (3) quality of presentation, (4) appropriate uses or users of book, (5) comparison with other books or editions, (6) background of book’s author(s) or editor(s), and (7) method used in examining the book (given that book’s length—a median of 1101 pages—often precluded full reading).

Results: In all, 37 reviews were assessed. For each of items 1 through 4, at least 34 reviews provided some information, although often the information seemed fragmentary. A total of 31 reviews included some information in item 5, and 22 included some in item 6. Only five reviews contained any information in item 7.

Conclusions: Textbook reviews in medical journals often do not address, or only address incompletely, some important areas. As a solution, journal editors should consider the use of

“structured book reviews” with standardized headings for various categories of content. Use of such reviews, resembling structured abstracts, could simplify the task of reviewing, help ensure that reviews are complete, and make information in reviews more accessible.

Alternatives to Peer Review: An Interdisciplinary Perspective

*Jonathan L. Entin

Journal of Legal Education, Case Western Reserve University, School of Law, 11075 East Blvd, Cleveland, OH 44106-7148

Much of the debate about the peer review process compares current practices with a hypothetical ideal. In assessing the criticisms of the existing system of peer review, however, we should consider how well that system works compared with alternative approaches. Peer review, whatever its faults, is justified as a second-best system of scientific and intellectual publication. A very different model of research publication dominates legal scholarship. Several hundred journals exist in that field, few of which use anything remotely resembling peer review. After examining this system in some detail and contrasting it with the editorial system in the sciences, I suggest that adopting such a system would result in substantially more problems than currently exist. It would probably encourage deleterious changes in the organization, length, and content of scientific papers and almost certainly would create incentives for publication decisions to be influenced significantly by institutional affiliation and other intellectually extraneous author characteristics. In short, whatever is wrong with peer review, the system is preferable to the likely alternatives.

Scientific Misconduct and Peer Review

A Case Study of Irresponsible Editorial Behavior

Martin Fichman and *Selma Zimmerman

Multidisciplinary Studies, Glendon College, York University, 2275 Bayview Ave, Toronto, Ontario, Canada M4N 3M6

We present a current example of the abrogation of editorial decision making, disregard of scholarly standards, and a lack of willingness on the part of a federally funded publisher to be responsible for the scholarly standards of its academic journals.

This case study centers around an article on the relationship of women in the labor force and society’s ills. It was published in a special conference issue on chaos theory in a physics journal, of which the article’s author was also guest editor. This paper’s publication came about as a result of (1) inappropriate conduct of author/guest editor (ie, submitting a paper for a conference issue that was not given at the

conference); (2) the journal editor's abrogation of responsibility to the journal's mandate (ie, restricted to contributions to physics) and to scholarly standards (ie, the publication of a paper that contains no clear hypothesis, no methodologic or data collection structure, no rigorous analysis for its conclusions, and no reference to the relevant scientific literature), and (3) an inadequate policy of one peer review per paper.

Given that the scientific journal is now the dominant mode of scientific communication, editorial action and peer review remain the most potent preemptive weapons to combat research error and/or misconduct in the reporting of scholarly scientific activity. This case study highlights the absolute need for a more rigorous procedure for selection of editors and reviewers for scholarly scientific journals and for more definitive, and enforceable, guidelines for editors, authors, and reviewers.

How Should We Deal With Editorial Misconduct?

*Douglas G. Altman,¹ Iain Chalmers,² and Andrew Herxheimer²

¹Medical Statistics Laboratory, Imperial Cancer Research Fund, PO Box 123, 61 Lincoln's Inn Fields, London WC2A 3PX, England; and ²UK Cochrane Centre, Oxford, England

The pressures on medical researchers to publish and the consequent lapses of scientific standards have been well documented. Less often discussed are failures by editors (and referees) to behave honestly and honorably toward authors. Examples include rejecting a paper after acceptance, or after all the conditions of a conditional acceptance have been met, and not allowing authors the right of reply to published critical correspondence. Authors are not always successful in bringing about an editorial change of heart.

To focus discussion of the issue of how aggrieved authors might seek redress, we describe a particularly serious case of editorial abuse of power. Some years ago, an editor republished a previously published paper solely to attack it in an editorial, did this without the authors' permission (despite publishing a statement to the contrary), published a further editorial attack when, after much pressure, a year later he published an edited version of the response by the authors, refused to publish any correspondence about the editorial attacks, and finally gave another editor a dishonest account of events to dissuade him from publishing an account of the whole sorry saga. We had full documentation to support these charges. We submitted our account of these events to six journals, including three that had been involved, but were unable to get it published.

There is no appeal mechanism for authors when they think that they have been badly treated by editorial abuse of power. We suggest that the International Committee of Medical Journal Editors should turn its attention to editorial misconduct and investigate possible mechanisms for allowing authors' grievances to be heard, including the possibility of an International Scientific Press Council.

Fraudulent and Unethical Practices in Scientific Publishing

*Herbert N. Nigg¹ and Gabriela Radulescu²

¹*Bulletin of Environmental Contamination and Toxicology*, University of Florida, 700 Experiment Station Rd, Lake Alfred, FL 33850; and

²Springer-Verlag, New York, NY

Objective: To document conclusively cases of scientific fraud and unethical practices in the *Bulletin of Environmental Contamination and Toxicology*.

Design: The editor-in-chief and six associate editors used computer databases, comparison of articles in similar journals, and reviewer queries to provide data.

Results: Two flagrant cases of scientific fraud were discovered. The first incident was a dry laboratory case and was uncovered by an associate editor. The second case was reported to the editors by authors whose work had been plagiarized. The first case was discovered before publication; the second case was published before discovery. Other cases of unethical practices discovered were categorized as (1) split studies, (2) inclusion of previously published data without reference, and (3) duplicate publication. No case in these categories was definitive because authors denied the accusation and the proof of unethical practice was not conclusive.

Conclusions: The editors concluded that unethical and fraudulent practices in scientific publishing are more widespread than initially suspected and that scientific fraud and unethical practices are difficult to detect and, in most cases, to prove.

The Scientific Community's Response to Evidence of Fraudulent Misconduct

*William P. Whitely,¹ Drummond Rennie,² and Arthur W. Hafner³

¹Chicago, Ill; ²*JAMA*, 515 N State St, Chicago, IL 60610; and

³Glenview, Ill

Objective: (1) To determine whether scientists can detect the presence of fraudulent results in published research papers and (2) to identify the sources of information that are most likely to lead to the fullest censure of fraudulent scientific papers.

Design: Case-control study comparing papers by an author known to have published fraudulent papers, Robert A. Slutsky, with control papers by other authors. Each Slutsky paper was scored relative to two non-Slutsky papers from the same issue of the same journal. The main outcome measure was an ordinal score (3, 2, or 1) showing the relative number of non-self-citations received by Slutsky papers.

Results: (1) Slutsky's papers were cited as frequently as control papers before Slutsky's misconduct was exposed. (2) Slutsky's papers were cited less frequently than control papers after news of Slutsky's misconduct became generally known. (3) Slutsky's papers that were retracted in print or in the MEDLINE database were not cited less frequently than unretracted papers by Slutsky. (4) Slutsky's papers that were identified as having been fraudulent or questionable in a

scientific review article were cited significantly less often than other papers by Slutsky.

Conclusions: Scientists did not suspect that Slutsky's papers contained fabrications until news of Slutsky's misconduct became public. Afterward, scientists looked to Slutsky's academic institution for guidance as to which articles were tainted. Academic institutions can play a significant role in correcting the scientific literature by thoroughly investigating suspected cases of fraud and disclosing their findings.

Duplicate Publication of Original Articles in and From the *Nederlands Tijdschrift voor Geneeskunde* (Dutch Journal of Medicine)

Hans Barnard and *John A. Overbeke

Nederlands Tijdschrift voor Geneeskunde, PO Box 75971, 1070 AZ Amsterdam, the Netherlands

Objective: To determine the numbers of correct and incorrect duplicate publications of original articles in the *Nederlands Tijdschrift voor Geneeskunde* (Dutch Journal of Medicine) based on the statement on duplicate publication by the International Committee of Medical Journal Editors (ICMJE).

Design: Retrospective bibliometric study followed by interviews. Of the first and second authors of 172 original articles (all 95 articles from the first half of 1990 and all 77 articles from the first half of 1992), all biomedical publications having the same purport published in 1990 through 1992 were looked up. The authors were subsequently approached by telephone to find out the background of these duplicate publications.

Results: For two articles the corresponding articles were not found. Of the remaining 94 articles from the first half of 1990, 11 (12%) proved to have been published again without this fact being mentioned in a footnote (for which omission the journals were responsible in four cases and the authors in seven cases). Of the remaining 76 articles from the first half of 1992, 12 were found to have been published previously, five in accordance with and seven against the ICMJE statement. Duplicate publication against the statement was attributed to incorrect interpretation of the ICMJE statement by the authors.

Conclusions: Of the original articles published in the *Nederlands Tijdschrift voor Geneeskunde*, a minimum of 14% are being or have already been published elsewhere, 11% without this fact being stated according to the ICMJE statement on duplicate publication.

Bias and Peer Review

Publication Bias and Public Policy

*Lisa Bero and Drummond Rennie

Institute for Health Policy Studies, University of California, San Francisco, 1388 Sutter St, 11th Floor, San Francisco, CA 94109

Objective: To examine the tobacco industry's claim that publication bias makes risk assessments of environmental tobacco smoke (ETS) invalid. Public health policies regulating exposure to hazardous compounds, such as ETS, are often based on risk assessments. The tobacco industry has argued, most recently in a lawsuit against the the US Environmental Protection Agency (EPA), that the EPA's risk assessment of ETS, which concluded that ETS is a cause of lung cancer and respiratory disorders, is invalid. The industry claims that the meta-analysis conducted by the EPA was flawed because publication bias against "negative" studies resulted in exclusion of negative studies from the meta-analysis.

Design: To determine if negative studies on ETS are published, the numbers of articles that supported ("positive studies") or refuted ("negative studies") the hypothesis that ETS exposure is associated with adverse health effects were assessed. All articles that tested this hypothesis were selected from the peer-reviewed medical literature and tobacco industry-sponsored symposia proceedings. Articles were identified by a computerized search of the medical literature supplemented with material obtained from the tobacco industry.

Results: Eighty-two percent (40/49) of the peer-reviewed articles were positive, compared with 37% (24/65) of symposia articles ($P = .0001$). Ninety-four percent (46/49) of the peer-reviewed articles contained a "Methods" section, compared with 43% (28/65) of symposia articles ($P = .0001$).

Conclusions: These findings raise the policy question of whether negative studies should automatically be included in meta-analyses. We recommend that the quality of the studies, their sponsorship, and the review process to which they have been subjected be carefully evaluated.

Effect of Institutional Prestige on Reviewers' Recommendations and Editorial Decisions

*Joseph M. Garfunkel, Harvey J. Hamrick, Edward E. Lawson, and Martin H. Ulshen

Journal of Pediatrics, CBN 7230, Medical School Wing C, University of North Carolina, Chapel Hill, NC 27599-7230

Objective: To determine whether manuscripts from institutions with greater "prestige" are more likely to be recommended for publication by reviewers and to be accepted for publication.

Design: Retrospective study of manuscripts from the United States received at the *Journal of Pediatrics* between January 1 and July 31, 1992. Manuscripts were classified either as major papers or as brief reports. Institutions were ranked in quintiles according to the monetary value of grants funded by the National Institutes of Health. Reviewers' recommendations were classified as reject, reconsider, or accept, and editorial decisions were classified as accept or reject, without regard to qualifying recommendations.

Results: For the 147 brief reports, lower institutional rank was associated with lower rates of recommendation for acceptance and of selection for publication. The acceptance rate for the highest quintile was 32% vs 6% and 11% for the fourth and fifth quintiles, respectively ($P=.004$). For the 258 major papers, however, there was no significant relationship between institutional rank and either the reviewers' recommendations or the acceptance rate. Similar results were found when the manuscripts were divided into five numerically equal groups according to institutional rank. The power to detect a 20% difference in acceptance rate was 85%.

Conclusions: There was no evidence of reviewer or editorial bias in the evaluation of major papers. The differences observed for brief reports probably reflect the submission of more case reports and brief papers of lesser quality from lower-ranking institutions, most likely related to the lower levels of research funding for more substantive studies.

Is There Gender Bias in JAMA's Peer Review Process?

*Julie Gilbert,¹ Elaine Williams,² and George D. Lundberg²

¹Department of Public Policy and Department of Biological Sciences, Stanford University, Palo Alto, Calif; and ²Scientific Publications Group, American Medical Association, 515 N State St, Chicago, IL 60610

Objective: To assess whether research articles submitted to *JAMA* have a variable chance of acceptance, or possess differing peer review and manuscript processing characteristics, associated with the gender of the corresponding author, the assigned editor, or the reviewers.

Design: A retrospective, controlled analysis of all original contributions (OC), preliminary communications (PC), clinical investigations (CI), and brief reports (BR) received by *JAMA* in 1991. Statistical analysis assessed whether the gender of the corresponding author, the assigned editor, or the reviewers had an effect on, or an association with, the acceptance rate, peer review, or manuscript processing characteristics.

Results: Five female editors and eight male editors reviewed and made decisions on 1851 articles. Female editors handled 847 articles; male editors handled 1004 articles. Editors requested 3382 reviews for all OCs, PCs, CIs, and BRs received in 1991. Female editors used female reviewers 462 times and male reviewers 1042 times. Male editors used female reviewers on 468 occasions and male reviewers on 1410. Of the 1851 articles considered, 243 (13.1%) were eventually accepted. Of submissions assigned to female editors, 76 (12.6%) of 605 articles with male corresponding authors and 25 (10.3%) of 242 articles with female corresponding authors were accepted. Of articles assigned to male

editors, 115 (14.4%) of 800 articles with male corresponding authors and 27 (13.2%) of 204 articles with female corresponding authors were accepted. The average turn-around time for all reviews was 26.3 days (median, 24). Female reviewers took an average of 26.9 days (median, 25) while male reviewers took an average of 24.7 days (median, 22). On a scale of 1 (lowest score) through 5 (highest score), female editors gave female reviewers an average score of 3.5 and male reviewers an average score of 3.4. Male editors gave both male and female reviewers an average score of 3.7.

Conclusions: Articles submitted to *JAMA* in 1991 did not experience a significant difference ($P>.4$) in acceptance rates based on the gender of the corresponding author and assigned editor. Female editors used female reviewers more often than did male editors ($P<.001$).

Blinded Peer Review

Multiple Blinded Reviews of the Same Two Manuscripts: Referee Characteristics, Language, and Methodologic Analyses

*Povl Riis,¹ Magne Nylemma,² and Yngve Karlsson³

¹Medical Department C, Herlev University Hospital, DK-2730 Herlev, Denmark; ²*Journal of the Norwegian Medical Association*, Oslo, Norway; and ³*Swedish Medical Journal*, Stockholm, Sweden

Objective: To study the quality and extent of referees' methodologic analyses, the usefulness of referees' statements to editors and authors, and the influence of English vs national language.

Design: Two nonauthentic but realistic papers with an approach familiar to all clinicians were sent for blind review to 20 internists, 20 surgeons, and 20 general practitioners in Norway, Sweden, and Denmark, respectively, to evaluate a new structured form for referee judgments. Each of the two papers was prepared in both an English and a national-language version. Through randomization, each referee received one in the English version and the other in the national language. A structured assessment of the quality of manuscripts was expressed on a five-point scale, and the impact of referee characteristics (age, gender, and experience) was analyzed by a multiple linear regression model. Methodologic flaws were inserted in both manuscripts (wrong sampling unit, inadequate choice of statistical methods, or inappropriate use of control groups), and the present analysis aimed at investigating to what extent these invalidating weaknesses were mentioned by the referees.

Results: A total of 156 reviews could be used for the study of referee characteristics and language. With increasing experience, the referees reached a lower quality score ($P<.05$). A tendency toward stricter assessment with younger referee age was seen ($P<.05$). No influence of the referees' gender,

speciality, or nationality was found. For the test paper judged to be of the lowest quality, the English-language version was assessed to be better than the national-language version ($P < .05$). A total of 159 reviews were applicable for the methodology study, yet in 54 (34%), no methodologic comments accompanied the assessment, and in six there were only very incomplete ones. In all, 200 comments on individual manuscripts were usable for analysis. On a four-point qualitative scale created for this purpose, the average was 1.7 points, without any significant differences between specialties, languages, or countries. Wrong sampling units were mentioned by only 24% (95% confidence interval, 15 to 35), and only one referee mentioned the incorrect use of a parametric test in nonparametric material. More referees might have noticed, but did not mention, these grave statistical errors.

Conclusions: The more experienced and younger referees gave a stricter assessment of the papers than their less experienced and their older colleagues. An English version seems to be more acceptable than a national-language version of the same manuscript. Most referees mentioned only the shortcomings of manuscripts as part of their overall judgment. Such comments might help editors but will not serve the education of scientists in scientific methods, which is the important spinoff of review.

The Relative Efficiency of Double-blind vs Single-blind Reviewing: Empiric Evidence From Economics Journals

*David N. Laband¹ and Michael J. Piette²

¹Department of Economics and Finance, Salisbury State University, Salisbury, MD 21801; and ²Economic Research Services Inc, Tallahassee, Fla

Objective: To determine whether articles published in double-blind journals receive significantly more or fewer citations than those published in single-blind journals.

Design: Drawing from a sample of 1051 full articles published in 28 economics journals during 1984, we used nonlinear regression and ordered probit techniques to estimate the impact of double-blind reviewing on citations, controlling for a variety of author, article, and journal characteristics.

Results: Citation analysis of 1051 articles published in 28 top economics journals showed that the single-blind review process was plagued by type I error bias. Articles published in journals using a double-blind review process attracted significantly more citations than did articles published in journals using a single-blind review process.

Conclusions: Single-blind review apparently suffers from type I error, at least to a greater extent than does double-blind review. That is, single-blind journals publish a larger fraction of articles that should not have been published than do double-blind journals. When reviewers know the identity of the author(s) of a paper, they are able to (and apparently do) substitute particular criteria for universal criteria in their evaluative process.



The Effects of Blinding on Acceptance of Research Papers by Peer Review

*Martin Fisher,¹ Sanford B. Friedman,² and Barbara Strauss²

¹Division of Adolescent Medicine, North Shore University Hospital, 300 Community Dr, Manhasset, NY 11030; and ²*Journal of Developmental and Behavioral Pediatrics*, New York, NY

Objective: To study whether reviewers aware of author identity are more biased than reviewers unaware of author identity.

Design: The 57 consecutive manuscripts submitted to the *Journal of Developmental and Behavioral Pediatrics* between September 1991 and March 1992 were sent to two blinded and two nonblinded reviewers. Spearman rank correlation coefficients were used to compare the sum of rating scores of 1 through 5 (1, accept; 5, reject) given by the two blinded reviewers, the two nonblinded reviewers, and the editor with the number of papers published previously by the first and senior authors (as determined from requested curricula vitae). Blinded reviewers were sent a questionnaire asking whether they could determine the identity of the authors, how they knew, and whether they thought blinding changed the quality or difficulty of their review.

Results: Forty-two percent of manuscripts were accepted; all articles that received a composite score greater than 16 from the four reviewers (16%) were rejected, all articles with scores less than 13 (28%) were accepted, and all articles scoring 13 through 16 (56%) were decided by the editors. The Wilcoxon sign rank test showed no differences between blinded and nonblinded scores. The number of previous articles by the senior author was significantly correlated ($P < .01$) with blinded scores ($r = .45$) and editor decisions ($r = .45$), but not with nonblinded scores. The number of articles by the first author was correlated ($P < .05$) with editor decisions ($r = .35$) but not with blinded or nonblinded scores. Fifty (46%) of 108 blinded reviewers correctly guessed the identity of the authors, mostly from self-references and knowledge of the work; 86% believed blinding did not change the quality of their review, and 73% believed it did not change the difficulty of performing a review.

Conclusions: The blinded reviewers and editors in this study, but not the nonblinded reviewers, gave better scores to authors with more previous papers. These results imply that blinded reviewers may provide more unbiased reviews, and that editors, but not nonblinded reviewers, may give a benefit of the doubt to more experienced researchers.

The Effects of Blinding on Editors' Decision Making

*Robert A. McNutt,¹ Arthur T. Evans,¹ Suzanne W. Fletcher,² and Robert H. Fletcher²

¹Division of General Medicine and Clinical Epidemiology, Department of Medicine, CB #7590, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7590; and ²*Annals of Internal Medicine*, Philadelphia, Pa

Objective: To determine whether blinding reviewers to the identity of manuscript authors and institutions affects editors' decisions to publish.

Design: Data were obtained during a randomized trial of blinded peer review at the *Journal of General Internal Medicine*. One hundred thirty-one consecutive, original research manuscripts were sent to two external reviewers, one of whom was randomized to receive a manuscript with the authors' and institutions' names removed. Editors were blinded to the identity of the authors and reviewers but were unblinded before deciding whether to publish the manuscript.

Results: Editors agreed more often with blinded reviewers; blinded reviewers were 100% sensitive and 75% specific for detecting which manuscripts would be published, while unblinded reviewers were 75% sensitive and 60% specific. When reviewers' recommendations differed, no manuscript was published when the blinded reviewer recommended rejection, whereas 19% were published when the unblinded reviewer recommended rejection ($P=.02$, Fisher's Exact Test). The effects of blinding were most powerful for manuscripts from high-prestige authors and institutions; blinded reviewers gave better methodologic critiques and recommended rejection more often (both $P<.05$). Blinded reviewers' recommendations and their critiques of study methods were associated with editors' decisions, even after controlling, by multivariable analysis, for the editor's independent assessment of the manuscript. Recommendations by unblinded reviewers, on the other hand, explained none of the variation in editors' decisions.

Conclusions: Blinding reviewers to authors and institutions removes bias and affects editors' decisions to publish.

Poster Session Abstracts

Poster session abstracts are listed in alphabetical order (by presenting author's last name) on pages 20 through 28. Asterisks indicate the names of presenters. Affiliations are listed for each author and an address is provided for the corresponding author.

So Much for Peer Review: Evaluation of Peer Review of Independently Submitted Manuscripts vs Manuscripts Previously Reviewed and Approved by a Society Program Committee

*Arthur E. Baue and Donald L. Kaminski

Department of Surgery, St Louis University Medical Center, 3635 Vista, PO Box 15250, St Louis, MO 63110-0250

Objective: To evaluate the peer review process in a surgical journal distinguished by relationships with a number of surgical societies.

Design: Reviews and recommendations of editorial board members (EBMs) and invited reviewers (IRs) of manuscripts submitted to the *Archives of Surgery* during a 6-month period in 1989 were compared for agreement, consistency, and patterns of reviewers. These results were then compared with those for manuscripts submitted after presentation at surgical society meetings.

Results: Each of 241 independently submitted manuscripts

had two reviews by EBMs and IRs. Agreement on publication recommendations was as follows: EBMs and IRs, 55%; two IRs, 35%; one EBM and one IR, 57%; and two EBMs, 65%. Only 11% had one vote to accept and one to reject.

Two EBMs recommended accepting all manuscripts reviewed. Seven EBMs and four IRs rejected all manuscripts reviewed. Eight EBMs and 15 IRs made an equal number of accept-reject recommendations. Thirty percent of the manuscripts were accepted. The abstracts that were initially reviewed and accepted by a society's program committee, in contrast, were already subjected to peer review. After the presentations and discussions, final manuscripts were reviewed by a publication committee and/or EBMs, a triple process with 85% agreement and acceptance. Discussions were also published.

Conclusions: The high acceptance rate for society manuscripts and the high level of agreement and rejection rate for independent submissions indicate that a majority of surgeons prefer presenting their best work to a society audience, followed by publication, rather than independent submission. This peer review process works. Enlightened, careful editors provide a control mechanism by recognizing tough, charitable, and balanced reviewers.

Effects of Peer Review and Editing on Readers' Perceptions of Research Manuscripts

*Jesse A. Berlin,¹ Steven N. Goodman,² Robert H. Fletcher,³ and Suzanne W. Fletcher³

¹Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA 19104-6095; ²Johns Hopkins University, Baltimore, Md; and ³*Annals of Internal Medicine*, Philadelphia, Pa

Objective: The peer review and editing process for clinical journals is intended to improve articles for readers, most of whom are clinicians and not researchers. Although this practice is supported by tradition and argument, its effectiveness has not been formally tested by research. We therefore studied how readers rate manuscripts before and after peer review and editing.

Design: We first surveyed a random sample of *Annals of Internal Medicine* subscribers to assess their interest in reading articles (described by titles) reporting original research to be published in the *Annals of Internal Medicine*. Next, we asked 416 readers to grade 104 manuscripts they said they would read. Each read one of two versions in identical format, as originally submitted and just before publication (after external peer review, statistical and technical editing, and revisions). Readers were asked to grade the written presentation, not the importance or scientific strength, of the articles, on a five-point ordinal scale on several dimensions (easy to read, concise, enjoyable, informative, and relevant) and the understandability of the article's introduction, methods, results, and conclusions. They also gave an overall score.

Results: The response rate from readers surveyed was more than 80%. Most readers assigned high scores to most manuscripts; for many items the majority of readers gave a rating of 4 or 5. Readers' perceptions of manuscripts were virtually unchanged from submission to publication. In the manuscripts with below-average grades, the understandability of the conclusions was improved by an amount approaching statistical significance (mean change, 0.19; confidence interval, -0.04 to 0.42).

Conclusions: Readers' opinions of published manuscripts, except for the clarity of the conclusions, seemed unaffected by the peer review process. This may result from the fact that most readers were satisfied with most manuscripts, even in the unrevised form.

An Instrument for Assessing the Quality of Reporting of Clinical Research

*Jesse A. Berlin,¹ Steven N. Goodman,² Suzanne W. Fletcher,³ and Robert H. Fletcher³

¹Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA 19104-6095; ²Johns Hopkins University, Baltimore, Md; and ³*Annals of Internal Medicine*, Pa

Objective: To develop an instrument for measuring the quality of reporting of clinical research and to measure its properties.

Design: We designed a 35-item questionnaire as part of a study of the effects of peer review and editing on the quality of reporting of clinical research (not the quality of the research itself). The questions were based on the literature on research methods, the authors' experience as editors, and pretesting with experts in clinical research methods. Between March 1992 and March 1993, 42 experts in clinical research methods used the questionnaire instrument to evaluate 104 original research manuscripts accepted by the *Annals of Internal Medicine*. Manuscripts were evaluated before and after the peer review process (ie, as submitted and when accepted for publication). A random sample of 32 manuscripts were evaluated independently by a second expert.

Results: The questionnaire instrument was able to detect effects of the peer review process, confirming construct validity. Study participants completed the instrument with virtually no missing data, indicating feasibility. Intraclass correlation coefficients were .12 for the average score of all items and .17 for a 10-point summary judgment score. The range of intraclass correlation coefficients for various subscales ranged from a high of .32 to near zero.

Conclusions: Despite its low reliability, the instrument was able to detect changes related to the peer review process. The instrument could be useful to teach reviewers how to review and as a guide for editors. If it is used for research on peer review, however, it will be necessary to compensate for the low reliability by using more reviewers for each manuscript.

More of the Same: Referees' Egregious Mistakes

*Juan Miguel Campanario

Departamento de Fisica, Universidad de Alcala, 28871 Alcala de Henares, Madrid, Spain

In the process of spreading scientific knowledge, professional journals play a key role. Most learned journals have editorial board members and/or external referees to evaluate the quality of submitted papers. Peer-reviewed journals have become symbols of excellence and certified knowledge. Sometimes, however, referees and editorial member boards have rejected or posed objections to papers that have become very important or have been highly cited.

A new approach to the study of the rejection of important papers and the scientists' resistance to discovery is proposed, and cases of rejection of very important discoveries or papers by the scientific establishment (especially by referees) are described.

Referees seem most prone to question or directly reject new ideas or methods and seem more willing to accept second-class work than innovative ideas. A systematic research program to identify important papers or discoveries that have had problems in reaching publication is needed.

An Analysis by Country of Scientific Reports in Five General Medical Journals

*Edward W. Campion and Rachel Hart

The New England Journal of Medicine, 10 Shattuck St, Boston, MA 02115-6094

Objective: To determine the percentage of scientific reports submitted to and published in US and British general medical journals by authors from other countries.

Design: Retrospective tabulation of the countries of origin for scientific reports published in five general medical journals in 1992.

Results: For *The New England Journal of Medicine (NEJM)*, 89 (34%) of 265 reports came from outside the United States, compared with 18 (7.6%) of 237 in the *Journal of the American Medical Association (JAMA)* and 39 (21%) of 188 in the *Annals of Internal Medicine (AIM)*. Of 380 reports in the *Lancet*, 226 (59%) came from outside Britain, compared with 89 (29%) of 310 for the *British Medical Journal (BMJ)*. For all five journals combined, the reports from non-English-speaking countries most frequently came from the Netherlands (41 reports), France (34), Sweden (32), Italy (29), Germany (21), Switzerland (21), Finland (19), and Japan (19). After adjusting for population, authors from the following countries published most frequently in these five journals: Finland (1.46 times the output of the United States), Sweden (1.42), Switzerland (1.22), and the Netherlands (1.07). The five journals received 15 201 manuscripts, of which 4174 (27%) were from non-English-speaking countries. A total of 1380 reports were published, for an acceptance rate approximated as 9.1% overall and as follows by country: Switzerland, 12.2%; United States, 12.0%; Sweden, 11.3%; Belgium, 9.5%; the Netherlands, 9.0%; Italy, 8.7%; Finland, 8.3%; United Kingdom, 8.2%; Canada, 8.1%; France, 7.2%; Australia, 6.2%; Germany, 5.6%; and Japan, 4.5%. From 1985 to 1992, the fraction of published reports from non-English-speaking countries increased: *NEJM*, from 12% to 23%; *JAMA*, 3% to 5%; *AIM*, 8% to 15%; *Lancet*, 29% to 36%; and *BMJ*, 19% to 21%.

Conclusions: In these US and British general medical journals, a substantial fraction of scientific reports came from non-English-speaking countries. The increasingly international character of these journals brings new sources for publication bias and reviewer bias along with greater demands on the peer review process and editorial process for fairness, accuracy, and honesty.

Rejection Rates in Biology Journals

*Lisa DiDonato

University of Arizona Press, 1230 N Park, Suite 102, Tucson, AZ 85719

Objective: To assess the effect of subdiscipline, submission volume, page count, backlog, number of reviewers per manuscript, relative academic competition, and journal prestige on rejection rates in biology periodicals.

Design: Editors of 130 North American, European, and Asian journals were surveyed about submission volume, page count, and rejection rates for the years 1989, 1990, and 1991. Editors

also described review processes and rated other journals within their biologic subdiscipline on a "prestige" scale. Two subdisciplinary groupings were based on taxonomic criteria and the paradigmatic divisions used by the National Science Foundation (NSF). Data on backlog (acceptance to publication time) and number of society members (for sponsored journals) were gathered at the library. Numbers were averaged over the 3 years for each journal and regressed against rejection rate.

Results: The questionnaire had a 48% response rate. Significant ($P < .05$) relationships were found between average rejection rate and number of submissions, number of pages published, and taxonomic subdivisions. No relationships were found between rejection rate and a journal's backlog ($P > .28$), prestige rating ($P > .34$), number of reviewers per manuscript ($P > .22$), number of society members ($P > .81$), or NSF subdisciplinary division ($P > .53$). Average annual statistics for all periodicals were as follows: submissions, 503.45±604.73 manuscripts; rejection rate, 40%±14%; page count, 2214.82±2294.05; and backlog, 6.16±2.46 months. Journals used 2.11±0.46 reviewers per manuscript and received (on a scale of 3 to 1) prestige ratings of 2.06±0.61.

Conclusions: Rejection rates in biology journals seem to be affected more by space restrictions than by other factors.

Incidence of Acceptance, Rejection, Resubmission, and Nonpublication of Full Reports of Research Presented at Academic Meetings

*Lynn Dirk

Editorial Office, University of Florida College of Medicine, Box 100254, Gainesville, FL 32610-0254

Objective: To determine outcome of scientific research, including submission to peer review journals.

Design: Information on publication of research in one academic medical department with an editorial office was obtained from files or survey of authors. The study sample consisted of abstracts that reported research presented at annual meetings of professional societies and that were published in society journals between 1980 and 1989.

Results: Of 55 authors surveyed about unpublished research, 50 responded (89%). Of 147 abstracts, 48% (70/147) were followed by submission of full reports to a peer-reviewed journal, 38% (56/147) of which were published; data from 16% (23/147) were published in book chapters or combined with other data; and 46% (68/147) of abstracts were not followed by publication in any form. Of 70 submitted manuscripts, data on submission to peer-reviewed journals were available for 59: 44% (26/59) were accepted by the first journal to which they were submitted; 25% (15/59), the second; 3% (2/59), the third; and 2% (1/59 each), the fourth, fifth, and sixth. Also, of 22% (13/59) submitted but not published, seven were rejected by one journal, three by two journals, two by three journals, and one by five journals. Of 33% (49/147) not written or submitted as full reports, authors of eight intended eventually to write and submit full reports. Authors of 41 reported that they would not write and submit

full reports, primarily because of low priority/loss of interest ($n=17$) or relocation of coauthors ($n=9$). When rejecting manuscripts that were never published, editors most often cited reviewers' comments ($n=9$) and methodologic problems ($n=4$).

Conclusions: Peer review keeps a small proportion of work unpublished, the quality of which must be evaluated outside of peer review to determine whether it suppresses innovation.



Are Surgical Research Publications Better Peer Reviewed Than Those in Standard Surgical Journals?

*James T. Evans, Homayoun Attaran, Serrie Lico, and Patricia Carlin

Department of Surgery, Nassau County Medical Center, 2201 Hempstead Turnpike, East Meadow, NY 11554

Objective: To use citation and quotation errors as indirect outcome measures of peer review of articles published in surgical publications specifically designated as research. These measures were then compared with the known citation and quotation errors in standard surgical journals.

Design: Three surgical research publications were chosen for review: *Journal of Surgical Research*, volume 51, No. 2, August 1991; *Journal of Investigative Surgery*, volume 3, No. 2, 1990; and *Surgical Forum*, volume 41, 1990. The references in each were consecutively numbered. Fifty random numbers were generated by computer to select references for study. Each original article was analyzed by means of previously published criteria. These results were compared with data from standard surgical journals by χ^2 tests.

Results: The *Journal of Surgical Research* had nine citation errors (four major and five minor) and 10 quotation errors (five major and five minor). The *Journal of Investigative Surgery* had 11 citation errors (six major and five minor) and 25 quotation errors (14 major and 11 minor). *Surgical Forum* had 16 citation errors (four major and 12 minor) and 14 quotation errors (10 major and four minor). There were no significant differences in major citation and quotation errors. Overall, there were more minor quotation errors ($P=.003$) and fewer minor citation errors ($P=.002$).

Conclusions: The peer review process is no more rigorous in specifically designated surgical research journals than in standard surgical journals.

Source and Acknowledgment of Funding for Articles Published in Medical Journals

*Timothy C. Fagan, Paula L. Carlson, and Robert J. Goldberg
Clinical Research Unit, Arizona Health Sciences Center, 1501 N Campbell Ave, Room 6335, Tucson, AZ 85724

Objective: To determine the percentage of original investigations published in major medical journals that did not acknowledge any source of funding and to determine if sources of funding changed during a 20-year period.

Design: All original investigations involving at least 10 human subjects for the calendar years 1971, 1981, and 1991 in two general medical journals, two internal medicine journals, and two clinical pharmacology journals were reviewed for acknowledged sources of funding. These 1549 articles were classified into government, industry, private foundation, other, or no stated source of funding.

Results: Overall, 28% of articles acknowledged no source of funding. This percentage increased from 22% in 1971 to 32% in 1981 and to 28% in 1991 ($P<.01$). Government funding decreased from 55% of articles published in 1971 to 39% in 1991 ($P<.001$). Industry funding increased from 9% in 1971 to 14% in 1991 ($P<.001$).

Conclusions: Manuscripts may be submitted with no acknowledged source of funding because of a perception that knowledge of an industrial source of funding may lessen the likelihood of acceptance by the journal or the likelihood of belief by readers. Alternatively, the study may have been supported by a variety of sources, eg, state salary, excess funds from other grants, or patient charges, which were not designated for the study. Readers deserve to know the source of funding, and medical journals should require a statement of the source of funding for each manuscript accepted for publication.

Editorial Accountability in Peer Review in Nursing Journals

*Shirley H. Fondiller

Publishing for Health Dimensions, 605 W 112th St, Suite 5D, New York, NY 10025

Objective: To determine the practices and policies of nursing journals in their relationship with peer reviewers and authors.

Design: Eighty nurse editors were sent a 42-item questionnaire. Questions were based on the assumption that quantity and not quality had become the rule rather than the exception and designed to assess whether excellence in publishing was at risk, with accountability the principal issue.

Results: The response rate was 66%. More than 70% of the sample reported no written policies on the expectations and responsibilities of editors, reviewers, and authors. Statistical analysis showed many similarities among editors in their peer review practices, but a marked diversity appeared in the use or existence of formal orientation programs for reviewers, evaluation mechanisms, review of revised manuscripts, and policy toward multiple authors.

Conclusions: Respondents reported that the study heightened an awareness of increasing problems in peer review. Without impinging on the autonomy of nurse editors, it is recommended that some standardization among nursing journals be explored to ensure more accountability in the peer review system and to legitimize authorship.

Authorial Selection Criteria for Manuscript Submission

*Erica Frank

Department of Community and Preventive Medicine and Department of Medicine, Emory University School of Medicine, 69 Butler St SE, Atlanta, GA 30303-3219

Objective: To assess authors' methods of selecting journals to which they submit manuscripts.

Design: In 1992, all of the active clinical and research faculty of Stanford University School of Medicine (n=491) was sent up to three mailings of a one-page questionnaire.

Results: The response rate was 62.1% (n=305). On a scale of 6 (very important) to 1 (unimportant), respondents ranked factors affecting initial submission to a journal in the following (descending) order: the journal's prestige (5.2); readership composition (4.8); frequency with which it publishes articles on the topic (4.8); likelihood of manuscript acceptance (4.4); journal's circulation size (4.1); rapidity of manuscript turnaround (4.1); presence of good editors (3.9); likelihood of useful reviews (3.4); a history of having published in that journal previously (3.4); colleagues' recommendations (3.1); likelihood of useful biostatistical comments (2.6); presence of familiar editors (2.0); and likelihood of press attention (1.9). For subsequent submissions, the most important factors were the likelihood of acceptance (5.0) and the frequency with which the journal publishes on the topic (4.7).

Conclusions: Such factors as journal prestige, usual journal topics, and readership composition were found to be the most important factors for initial manuscript submissions. For subsequent submissions, more pragmatic factors, such as likelihood of acceptance, gained importance. These findings have significance for editors soliciting manuscripts and structuring policies for their journals.

Perceptions of Plagiarism in the Use of Other Authors' Language

*Kell N. Julliard

Christine M. Kleinert Institute for Hand and Micro Surgery, 225 Abraham Flexner Way, Suite 850, Louisville, KY 40202

Objective: To determine if physicians (1) perceived as plagiarism two sentences cited from another author's work used almost verbatim but without quotation marks and (2) considered this type of plagiarism important, as compared with the opinions of other academics.

Design: Participants compared two sentences from a recent medical journal article with three samples and identified whether the samples contained plagiarism. All samples cited the original but closely copied it without quotation marks. Participants also indicated whether using another author's language without quotation marks was a serious form of plagiarism. Eighty-four medical school faculty members, 28 editors-in-chief of medical journals, 121 English teachers, and 34 editors-in-chief of scholarly literary journals, randomly selected from printed lists, and 119 sophomore medical students received questionnaires.

Results: Rates of return were 60% for medical school faculty, 50% for medical editors, 34% for English faculty, 53% for literary editors, and 84% for medical students. Of physicians, 36% believed plagiarism occurred in any sample; 16% considered the use of an author's language verbatim without quotation marks a serious form of plagiarism. Of medical students, 68% believed plagiarism to have occurred; 51% believed it to be serious. Of English teachers and literary editors, 89% regarded plagiarism to have occurred; 75% believed it to be serious. Statistical differences between these groups ranged from $P=.003$ to $P<.00001$.

Conclusions: Physicians may have lower standards regarding the use of other authors' language in their articles than English specialists or medical students do.

Standardizing Peer Review: Evaluating the Process With a Sham Paper

*P. Dileep Kumar

Department of Medicine, Medical College Hospital, Calicut 673 008, Kerala, S India

Objective: The peer review process lacks standardization and absolute measurements in general. To make peer review more objective, a scoring system is suggested.

Design: A sham paper about a study on fluorosis in Calicut, India, with known deficiencies was mailed to 32 peer reviewers, who were asked to award marks to different parts of the paper (title, abstract, introduction, materials and methods, results, discussion, tables, references, and statistical analysis). The scores (y) were compared with a set of predetermined scores (x) for the sham article, and the constant x/y was calculated for each section. Ratings of an individual reviewer were also determined.

Results: Ten reviewers (31.25%) responded, and seven of these responses were usable. The constant x/y ranged from 0 to 2.5. The ratings of the reviewers ranged from 0.34 to 0.45.

Conclusions: The ratings indicate significant deviations from the set standard, which may not be ideal. If a particular reviewer is asked to award marks (z) for each section of a future paper, the real worth of the paper may equal $z(x/y)$.

Readers of Peer-Reviewed Medical Journals: Do They Have a Message for Their Media? A Pilot Study

John C. Roberts,¹ Robert H. Fletcher,² Suzanne W. Fletcher,² Kathleen Case,² and *Linda H. McClure³

¹Johns Hopkins University, Francis Scott Key Medical Center, 4940 Eastern Ave, Baltimore, MD 21224; ²*Annals of Internal Medicine*, Philadelphia, Pa; and ³Johns Hopkins University School of Hygiene and Public Health, Baltimore

Objective: To learn more about how practicing physicians use peer-reviewed journals and whether journals meet their needs in day-to-day practice.

Design: After conducting five focus groups of 41 clinicians, we constructed and pretested a survey. The resulting pilot survey was mailed in June and July 1993 to a systematic sample of 150 internists who subscribe to the *Annals of Internal Medicine*. Univariate descriptive methods of the results of the first wave of responses were carried out.

Results: Fifty-five of 150 physicians responded to the first of three mailings. Medical journals ranked highest in both use and usefulness in practice, followed by textbooks, colleagues in other specialties, and colleagues in respondents' specialties. Mass media sources ranked low, with television ranking lowest. Correlation between "use" and "usefulness" was high ($r > .50$), $P < .0001$ among all sources except radio. Respondents agreed strongly that "reading medical journals is necessary to good practice," but they equally as strongly said they were "overwhelmed by the number of journal articles presented" to them. Despite these pressures, 29.6% said they had "changed practice" based on a journal article, and 94% of that group could cite the article when probed.

Conclusions: In this pilot survey, these physicians perceived medical journals as being their most important source of medical information, and many could cite specific articles that changed their practices. These data suggest a discrepancy between physicians' perceptions about their information resources and previous observations that show colleagues to be the primary resource.

Mixed Messages: Biomedical Writing and Peer Review Discourse

*Gary L. Miranda

Kaiser Permanente Center for Health Research, 3800 N Kaiser Center Dr, Portland, OR 97227-1098

Since the turn of the century, the style of biomedical publications has shifted from that of journalism to that of natural science writing, a model grounded in Francis Bacon's 17th-century ideal of scientists as "faithful secretaries" of nature. By conforming to certain stylistic conventions—the passive vs the active voice, observation vs narrative or anecdote, abstract nouns vs verbs—such writing draws attention away from the writer/observer and creates the impression of words that refer "objectively" to quantifiable events unaffected by human intervention.

By contrast, peer reviews of submitted papers employ a radically different style from that of the papers themselves. Scientific integrity and exclusion of bias remain ideals, but the elaborate stylistic scaffolding that supports the impression of objectivity is temporarily dismantled. Thus, rhetoric, personal involvement, and even passion are tolerated to a degree that would be unacceptable in the paper itself. The impression is not that of neutral observers bearing witness to a physical event, but of concerned professionals discussing a matter of mutual interest.

These differences between peer review discourse and the formal style of scientific articles underscore that the latter is essentially a rhetorical device to bolster scientific credibility. Incorporating some of the elements of peer review discourse into the style of biomedical journals might better serve the needs of an enterprise whose purpose is not to perpetuate an abstract ideal of scientific "objectivity" but to share knowledge that will ultimately help to alleviate human suffering.

How Long Does Peer Review Take? One Component of Publication Lag Time

*Della J. Mundy¹ and Joseph F. Terdiman²

¹Department of Medical Editing, Kaiser Foundation Research Institute, 1800 Harrison St, 16th Floor, Oakland, CA 94612-3429; and ²Division of Research, Kaiser Permanente Medical Care Program, Oakland

Objective: To describe peer review lag time for a group of manuscripts from a health maintenance organization.

Design: We calculated mean times taken for peer review (peer review lag time), ie, total time taken from submission of a manuscript for publication to receipt of any reviewers' requests for its revision(s) to final acceptance. We excluded publisher processing time after peer review from acceptance to print publication and considered only those manuscripts published in each of the years analyzed. Using manuscript tracking software with a relational database, written in the Paradox Application Language for personal computers, we generated alerts if journal acknowledgment was not received 30 days after submission, and again if no response from reviewers was received after 90 days. After an alert was received, a form letter of inquiry was sent to the journal. Similarly, alerts were generated and form letters sent to authors if an edited draft (at initial or revision stage) was not returned in 30 days.

Results: Overall mean times from submission to acceptance for published manuscripts for the years 1989 ($n=32$), 1990 ($n=30$), 1991 ($n=41$), and 1992 ($n=36$), ie, for all manuscripts, whether accepted at first submission (including tentatively accepted manuscript revised and submitted to the same journal) or submitted to second and subsequent journals, were 5.84, 6.10, 6.07, and 4.62 months, respectively. Mean times from submission to rejection for each previous rejection of a published manuscript (approximating peer review lag time but including institutional processing time) were 1.48, 3.00, 2.95, and 3.66 months, respectively.

Conclusions: We observed a general decrease in time from

submission to acceptance and, as expected, an increase in peer review lag time with number of submissions. This descriptive study could not evaluate complex effects resulting from our alerts, publishers' reviewer tracking systems, and increased difficulty in obtaining acceptance caused by crowding from pressure to publish and decreased publishers' resources in this decade.

Reasons for Rejection at the *Nederlands Tijdschrift voor Geneeskunde* (Dutch Journal of Medicine) in 1990

Cees C. Kan, Jaques H. Lockefeer, and *John A. Overbeke

Nederlands Tijdschrift voor Geneeskunde, PO Box 75971, 1070 AZ Amsterdam, the Netherlands

Objective: To assess editorial board members' recommendations for rejection.

Design: A retrospective investigation of the reasons mentioned in 1990 by the editorial board of the *Nederlands Tijdschrift voor Geneeskunde* for rejecting original articles, reviews, clinical lessons, and case reports, and whether the length of the article, professional status of the first author, most relevant specialty, and origin of the article influenced the rejection probability.

Results: Of 547 articles, 38% (208) were rejected. Case reports were rejected most often (49%); reviews were rejected least often (33%). Peer review of original articles was the most thorough, and that of case studies the least thorough. The most frequent reason for rejection in all categories was "substantial shortcomings," particularly in the original articles (80%). Rejection of clinical lessons and case studies was relatively often due to criteria concerning the clinical message. Both rejected and accepted articles on average exceeded the length limit. Professional status and specialty of the authors were associated with a difference in rejection probability (χ^2 , $P < .05$). Specialists had the lowest rejection rates (31%). Articles originating from nonacademic institutions were more often rejected than articles from academic centers.

Conclusions: We advise authors to judge their articles by means of the criteria mentioned. Consultation of methodologic experts, literature databases, and/or the editorial board can improve an article's chances for publication. Nonspecialists can benefit from the experience and expertise of specialists. In general, authors' attitudes and willingness to alter their articles according to editorial advice appear to have beneficial results.

Peer Review: An Inquiry Among Referees of the *Nederlands Tijdschrift voor Geneeskunde* (Dutch Journal of Medicine)

Bouke C. H. De Jong and *John A. Overbeke

Nederlands Tijdschrift voor Geneeskunde, PO Box 75971, 1070 AZ Amsterdam, the Netherlands

Objective: To determine the manner of reviewing used by the referees of this journal.

Design: Inquiry by mail of 43 members of the editorial board (who are the usual referees) and 34 "external" referees.

Results: All 43 board members and 30 external referees responded. There were no differences in answers between the two groups. Mean refereeing time was 2.1 hours (range, 1.1 to 4.2 hours). Of the 73 responders, 66 were referees for four (range, one to 25) other journals. Compared with the time spent to referee for this journal, six spent less time, 30 spent the same time, and 28 spent more time for the other journals. The methods used and the validity of the conclusions were criteria determined to be important in judging original articles. Criteria determined to be unimportant included the institution of investigation, the illustriousness of the authors, knowing the authors, and the source of grants. Of 69 responders, 16 (23%) preferred blinding of authors and institutions, 24 (35%) did not prefer blinding, and 29 (42%) did not have a distinct preference. Of 72 referees, only 11 (15%) thought their judgment would be more objective if blinding was used.

Conclusions: In general, the responders had the opinion that most of the factors that would improve subjectivity in the refereeing process did not influence their recommendations.

Mistakes in References in the *Nederlands Tijdschrift voor Geneeskunde* (Dutch Journal of Medicine)

Sjoerd O. Hobma and *John A. Overbeke

Nederlands Tijdschrift voor Geneeskunde, PO Box 75971, 1070 AZ Amsterdam, the Netherlands

Objective: To determine the accuracy of quotations and references of manuscripts submitted to and articles published in the *Nederlands Tijdschrift voor Geneeskunde*.

Design: The accuracy of 100 randomly selected references of manuscripts submitted for publication during a period of 4 weeks was checked by means of *Index Medicus*. The citation and quotation of another 100 references (50 with a reference list of less than 13 articles and 50 with a reference list of more than 24 articles) of published original articles were checked, using the original article.

Results: In the references of manuscripts submitted for publication, 70% contained one or more inaccuracies. In the published articles, 31% of references were not completely correct. In 5% these errors prevented immediate identification of the cited article. In one of every three articles a citation

error was made: five trivial errors, 16 slightly misleading errors, and 12 seriously misleading errors. There was no association between the number of citation errors and the length of the reference list.

Conclusions: The number of inaccuracies in references and citations is high. Authors should check their references and citations more carefully.

Publication Policies for Economic Analyses and Biomedical Ethics

*Deborah Roney, Daniel Sulmasy, and Kevin Schulman

Clinical Economics Research Unit, Georgetown University Medical Center, 4000 Reservoir Rd NW, Washington, DC 20007

Objective: To assess several aspects of the publication process as they affect the quality of the published literature in economic analysis and biomedical ethics: peer review by experts in the field, appropriate assignment of the article to reviewers, and publication policies regarding the relationships between the author and the funding source.

Design: We developed a survey to compare these aspects of the review process for clinical, economic, and ethics contributions. We mailed the survey to 15 major medical journals, with 80% responding.

Results: In determining the reviewers for a particular submission, respondents reported that the journal had editors who specialized in the field 67% of the time for clinical articles, 42% of the time for economics articles, and 42% of the time for ethics articles. Respondents scored the frequency with which clinicians in the specialty field are selected as primary reviewers by means of a Likert scale, with 1 representing never and 5, always. Clinical submissions received a mean score of $4.6 \pm .5$, economics $4.3 \pm .9$, and ethics $4.4 \pm .5$. However, when describing reviewers, journals reported little understanding of the formal training of its reviewers in these fields. Eighty-two percent of journals had statisticians among their editors to address concerns in clinical manuscripts, only 17% had health economists among their editors, and none had ethicists among their editors. Only 8% of journals were able to classify correctly a philosophical ethics article as an original contribution. Disclosure of the financial relationship between the authors and the sponsor is requested by 75% of the publications. Only 33% inquired whether the investigator had a publication guarantee from the sponsor; only 8% knew if the sponsorship agreement required written approval of the manuscript by the sponsor before submission; only 8% knew if there was an independent steering committee for the study; and only 8% knew if the sponsor had the ability to delay manuscript submission for other than patent reasons. Further, 91% of journals would accept a manuscript reporting results of primary data collection studies performed under a consulting agreement, while 64% would accept a manuscript reporting results of secondary data collection studies performed under a consulting agreement.

Conclusions: The findings of this study suggest that there are several ways in which the peer review process can be strengthened to improve the quality of the medical literature in the areas of economic analyses and biomedical ethics.

Journals need to understand better the financial arrangements surrounding manuscript submission to protect the integrity of the medical literature.



Review Articles: The Role of Peer Review

*Bernard Schwartz and Susan Erickson

Survey of Ophthalmology Inc, 7 Kent St, Brookline, MA 02146

Objective: To evaluate the standards of peer review for review articles published in biomedical journals.

Design: Questionnaires were sent to 1650 publishers and editors of biomedical journals, requesting information about the characteristics (eg, length, number of references, authorship, and frequency) of literature-based review articles and the peer review policies applied to them.

Results: As of July 27, 1993, 193 responses had been received. Of these, 27 journals that do not publish review articles were excluded from the study. Twenty journals contain over 70% review articles and were considered to be "review journals." Nineteen (95%) of the review journals reported that they are peer reviewed. Only seven (39%) of these routinely use two or more referees other than the editors, thus meeting our criteria for "stringent" peer review. The remaining 146 journals are primarily original-report journals that publish some review articles. In 135 (92%) of the original-report journals, review articles are peer reviewed, but in only 73 (54%) of these are they stringently reviewed. Of the respondents, 144 answered the questions regarding source of review articles (invited vs volunteered), peer review procedures, and acceptance rates: 39 journals (group A) have invited reviews only; 82 journals (group B) have both invited and volunteered reviews; and 23 journals (group C) have volunteered reviews only. Stringent peer review is imposed by 14%, 44%, and 87% in groups A, B, and C, respectively. An average of 64% of reviews in group A, 37% in group B, and 25% in group C are accepted without revision. Rejection rates are 6% in group A, 23% in group B, and 38% in group C.

Conclusions: Review articles, especially those with invited authors, are not subject to the same standards of peer review that are applied to original reports. With the increasing complexity and volume of biomedical literature, review articles play a crucial role in the critical interpretation and integration of information. Therefore, high editorial standards should be set for the peer review of these articles.

Identification of Randomized Clinical Trials by Means of MEDLINE

*Peri Schuyler,¹ Kay Dickersin,² Roberta Scherer,² and Nancy Wright¹

¹National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894; and ²Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore

Objective: To determine (1) the causes of under ascertainment of randomized clinical trials (RCTs) with the use of MEDLINE searching and (2) whether indexing policy changes and vocabulary expansion beginning in 1989 at the National Library of Medicine (NLM) have led to improved application of Medical Subject Heading (MeSH) terms.

Design: Using a combination of electronic (MEDLINE) and hand search (66 journals), we created a "gold standard" of reports of RCTs in vision science, published in 1988 in an NLM-indexed journal. Reports were classified as RCTs if the intervention was assigned by means of randomization or quasi-randomization (eg, alternation). Each article was classified as to the location in the article (ie, title, abstract, or text) of explicit methodologic terms (eg, randomly allocated, randomized assignment) that led us to categorize it as an RCT. A subset of the 1988 RCTs was created, limited to reports containing some form of the truncated term *random* in the title or abstract. A similar file was created for vision trials published in 1991, to compare the application of MeSH indexing between the two years.

Results: We identified 196 reports describing RCTs published in 1988. Only 38% (74/196) of the reports were indexed under the MeSH heading Clinical Trials or the publication type Clinical Trial. Eighty-one percent of the articles so indexed contained methodologic terms in the title or abstract, and 18% contained those terms in the Methods section. In contrast, of 122 articles not indexed by either of these terms, only 49% contained explicit methodologic terms in the title or abstract and 46% did so in Methods. Comparison of the application of MeSH indexing showed an increase in the use of the MeSH clinical trial terms between 1988 (48%) and 1991 (78%).

Conclusions: Complete identification of RCTs by means of MEDLINE could be improved by authors' providing explicit methodologic information in the title and abstract of the

article. For articles using a term such as *randomized* in the title or abstract, indexing appears to have improved between 1988 and 1991. Recent indexing policy changes at the NLM are the likely reason for this.

The Effect on the Outcome of Statistical Tests on the Decision to Publish and Vice Versa

*Theodor D. Sterling, Wilf L. Rosenbaum, and James J. Weinkam

School of Computing Science, Faculty of Applied Sciences, Simon Fraser University, Burnaby, British Columbia, V5A 1S6 Canada

Objective: A preference to publish papers with positive rather than negative results has long been recognized. It appears to be confounded with the statistical test for accepting or rejecting a null state, focusing on the assumption that the probability that null state is true is no greater than 5 in 100 (ie, $P \leq .05$). This practice creates uncertainty about the meaning of reported probabilities and leads to the neglect of important negative findings.

Design: We reviewed 10 (behavioral, psychological, public health, epidemiologic, and medical) research journals to determine some of the factors in publication bias. All research papers for 1 year's publication were classified according to whether statistical hypothesis testing was employed and if the null state was rejected.

Results: The percentage of research articles that used statistical tests ranged from 44% to 98% for different journals. Of investigations that used so-called statistical tests, 81% to 98% rejected the null state with $P \leq .05$. There were few replications of previous studies. No null state was rejected with $P > .05$.

Conclusions: Publication bias appears not to have changed since 1958, when one of us first reviewed the use of statistical tests. The present system of selecting reports for publication favors to an unknown extent research results that may be due to chance. Negative findings seem to be ignored. Consequently, results of most meta-analyses are probably invalid. The uncertainty created by publication bias calls for far-reaching revisions in the peer review process.

Information About Continuing Medical Education

The American Medical Association is accredited by the Accreditation Council for Continuing Medical Education to sponsor continuing medical education for physicians.

The American Medical Association designates this Continuing Medical Education activity for 18 credit hours of Category 1 of the Physician's Recognition Award of the AMA.

You must attend all 3 days to receive all 18 credit hours. To obtain your CME credit, please sign the CME credit forms that will be available at the Registration Desk on Saturday, from 12:00 to 3:30 PM. Notification of your CME credit will be mailed to you 6 weeks after the Congress.

