

Testing for the Presence of Positive-Outcome Bias in Peer Review

A Randomized Controlled Trial

Gwendolyn B. Emerson, MD; Winston J. Warme, MD; Fredric M. Wolf, PhD; James D. Heckman, MD; Richard A. Brand, MD; Seth S. Leopold, MD

Background: If positive-outcome bias exists, it threatens the integrity of evidence-based medicine.

Methods: We sought to determine whether positive-outcome bias is present during peer review by testing whether peer reviewers would (1) recommend publication of a “positive” version of a fabricated manuscript over an otherwise identical “no-difference” version, (2) identify more purposefully placed errors in the no-difference version, and (3) rate the “Methods” section in the positive version more highly than the identical “Methods” section in the no-difference version. Two versions of a well-designed randomized controlled trial that differed only in the direction of the finding of the principal study end point were submitted for peer review to 2 journals in 2008-2009. Of 238 reviewers for *The Journal of Bone and Joint Surgery* and *Clinical Orthopaedics and Related Research* randomly allocated to review either a posi-

tive or a no-difference version of the manuscript, 210 returned reviews.

Results: Reviewers were more likely to recommend the positive version of the test manuscript for publication than the no-difference version (97.3% vs 80.0%, $P < .001$). Reviewers detected more errors in the no-difference version than in the positive version (0.85 vs 0.41, $P < .001$). Reviewers awarded higher methods scores to the positive manuscript than to the no-difference manuscript (8.24 vs 7.53, $P = .005$), although the “Methods” sections in the 2 versions were identical.

Conclusions: Positive-outcome bias was present during peer review. A fabricated manuscript with a positive outcome was more likely to be recommended for publication than was an otherwise identical no-difference manuscript.

Arch Intern Med. 2010;170(21):1934-1939

Author Affiliations:

Departments of Orthopaedics and Sports Medicine (Drs Emerson, Warme, and Leopold) and Medical Education and Biomedical Informatics (Dr Wolf), University of Washington School of Medicine, Seattle; *The Journal of Bone and Joint Surgery* (Dr Heckman); and *Clinical Orthopaedics and Related Research* (Dr Brand).

POSITIVE-OUTCOME BIAS (POB) is defined as the increased likelihood that studies with a favorable or statistically significant outcome will be published than will studies of similar quality that show unfavorable or “no-difference” results.¹⁻⁶ Although POB is not limited to the peer review process, manuscript review is considered an important locus of this phenomenon. When investigators^{1,3,6-9} have looked downstream at the published literature, they have suggested the presence of bias in peer review, but the authors of more rigorous studies^{4,5,10-17} have drawn mixed conclusions on this point. Because of numerous confounding factors (including differences in study quality, sample size, and clinical relevance), the existence of this phenomenon is almost always inferred rather than proved, even when appropriate denominators are compared.

To better characterize whether POB exists, we posed 3 hypotheses and tested them by submitting 2 versions of a fabricated

manuscript to peer review at 2 cooperating journals. The manuscript versions were substantially identical except that one version was “positive” in that it found a difference between the treatment groups and the other version concluded with a no-difference finding. First, we hypothesized that the percentage of reviewers who would recommend publication of the positive manuscript version would be higher than the percentage who would recommend publication of the no-difference version. Second, to determine whether reviewers would scrutinize the no-difference versions more stringently, we inserted identical purposefully placed “errors” in each of the 2 versions of the test manuscript and hypothesized that reviewers would identify fewer of those errors in the positive version. Third, to ascertain whether study outcomes affected reviewers’ perceptions of the (identical) “Methods” sections, we hypothesized that reviewers would score the “Methods” section in the positive-outcome manuscript ver-

sion higher than they would score the identical section in the no-difference version.

METHODS

STUDY INFRASTRUCTURE, INCLUSION, AND TEST MANUSCRIPT DISTRIBUTION

The institutional review board of the University of Washington School of Medicine, Seattle, approved this CONSORT (Consolidated Standards of Reporting Trials)-conforming study (Figure).¹⁸ Two versions of a nearly identical fabricated manuscript describing a randomized controlled trial were created (eAppendices 1 and 2; <http://www.archinternmed.com>) and were sent to peer reviewers at 2 leading orthopedic journals; the reviewers were blinded to the manuscript's authorship and other administrative details, which is the standard practice for both journals. The 2 manuscript versions were identical except that in the positive version, the data point pertaining to the principal study end point favored the primary hypothesis, and the conclusion was worded accordingly, whereas in the no-difference version, the data did not show a statistically significant difference between the 2 study groups, and the conclusion was worded accordingly. We intentionally placed 5 errors in each manuscript.

The editors in chief of the 2 participating journals, *The Journal of Bone and Joint Surgery (American Edition)* (JBJS) and *Clinical Orthopaedics and Related Research* (CORR), identified a large number of experienced reviewers with expertise in the subject area of the manuscript (general orthopedics, spine, and joint replacement) and then sent all of them an e-mail notifying them that some time in the next year they might receive a manuscript as part of a study about peer review and that if they wanted to decline to participate they should contact the editor. Potential reviewers were not made aware of the study's hypotheses, that the manuscript they received would be fabricated, or when they might receive the manuscript. The university-based study researchers were blinded to all identifying information about the reviewers themselves.

FABRICATED TEST MANUSCRIPTS

Two versions of the fabricated test manuscript on the subject of antibiotic prophylaxis for clean orthopedic surgery were created (eAppendices 1 and 2), one with a positive conclusion (showing that the administration of an antibiotic for 24 hours postoperatively, in addition to a preoperative dose, was more effective than the single preoperative dose alone in the prevention of a surgical-site infection) and the other with a no-difference conclusion. Both manuscript versions were amply, and identically, powered. The manuscripts consisted of identical "Introduction" and "Methods" sections, "Results" sections that were identical except for the principal study end point (and data tables) being either statistically significantly different or not, and "Comment" sections that were substantially the same. To test the second hypothesis of this project (that error detection rates might differ according to whether a positive or a no-difference manuscript was being reviewed), 5 errors were placed in both versions of the fabricated manuscript. These consisted of 2 mathematical errors, 2 errors in reference citation, and the transposition of results in a table; these errors were identical, and identically placed, in both manuscript versions. Because the "Methods" sections in the positive and no-difference manuscript versions were verbatim identical, in principle they should have received equal scores from reviewers who rated the manuscripts for methodological validity.

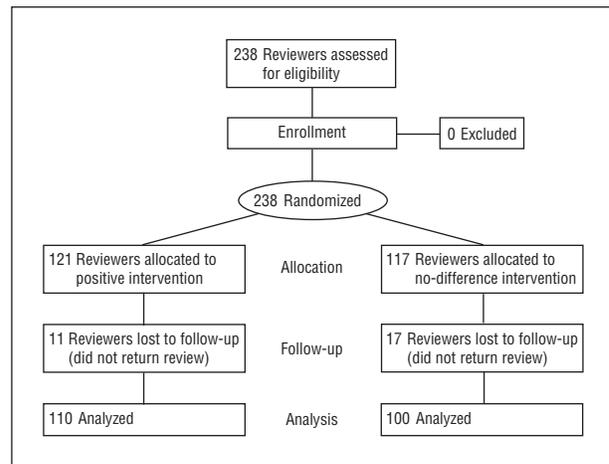


Figure. Randomized controlled trial flowchart for positive-outcome bias study.

The test manuscript was created purposefully to represent an extremely well-designed, multicenter, surgical, randomized controlled trial. It was circulated to reviewers before the journals involved began requiring prospective registration of clinical trials, and, thus, the fact that the trial was not so registered would not have been a "red flag" to peer reviewers. At both journals, peer review was blinded, and funding sources for blinded manuscripts under review are not disclosed to peer reviewers.

RANDOMIZATION AND REVIEW

Participating reviewers were randomized to receive either the positive or the no-difference version of the fabricated test manuscript. Block randomization was used, with blocks of 20 manuscripts (10 positive and 10 no-difference) used to assign reviewers for each journal approximately the same number of each manuscript version to review overall. Once a reviewer was invited to review a version of the manuscript, that reviewer's name was removed from the eligible pool at both journals (for those reviewers who review at both journals) to ensure that no reviewer was contacted twice during the study. The manuscripts were distributed to participating reviewers between December 1, 2008, and February 28, 2009.

Reviewers at CORR were given 3 weeks to complete the reviews, and those at JBJS were given 25 days. These are the usual periods for review at these journals. At the end of the review period, the reviews were forwarded by each journal to the university-based investigators, who were blinded to identifying information about the reviewers and to which version of the manuscript was being reviewed while they were analyzing the reviews. Once all the reviews had been received, each reviewer was sent a second notification indicating that he or she had participated in the study and identifying the test manuscript explicitly to prevent inappropriate application of its content to clinical practice.

STUDY END POINTS

The 3 hypotheses were tested by assessing the difference between the 2 groups of reviews with respect to 3 outcomes: the acceptance/rejection recommendation rates resulting from the peer reviews of the 2 versions of the manuscript (accept or reject; the a priori primary study end point), the reviewers' methods quality scores (range, 0-10), and the number of purposefully placed errors in each manuscript that were detected (range, 0-7). The maximum number of errors that could be detected was 7, not 5, because subsequent to manuscript distribution we found 2 inadvertent errors in addition to the 5 intentional errors in both

manuscript versions; the 2 inadvertent errors were minor discrepancies between the contents in the abstract and the body of the manuscript, and they were identical, and identically located, in both manuscript versions. We included these errors in the error detection end point of the present study, rendering the denominator 7 errors for detection by the reviewers.

We had to accommodate some differences between the reviewer recommendation formats of the 2 participating journals. At both journals, reviewers are asked to use a similar 4-grade scale regarding recommended manuscript disposition, and at both journals it is the editors, not the reviewers, who ultimately determine manuscript disposition. Although the exact verbiage varies slightly between the journals, the editors agreed that in practice, the grading process is similar at the 2 journals: A indicates accept or accept with minor revisions; B, accept with major revisions; C+, major revision needed (but publication unlikely); and C, reject. Both journals solicit free-text comments from reviewers. In addition, *CORR* reviewers are asked to give a numerical score of 1 to 10 for various elements of the manuscript, including the validity of the methods used. To generate numeric methods scores for *JBS* reviews that could be compared with the numeric grades of the "Methods" sections generated by reviewers at *CORR*, 2 of us (G.B.E. and W.J.W.) read each *JBS* review, which had been blinded by a third one of us (S.S.L.) to redact the overall recommendation for publication and to remove any indication in the "Comment" section of which version of the manuscript was being reviewed; each review was then assigned a numerical score of 1 to 10 assessing methodological validity by each of the 2 readers. This scoring process was conducted independently by each of the 2 readers, not as part of a discussion or consensus-driven process. Differences of less than 2 points on the 10-point Likert scale were averaged; differences of 2 points or more were to be adjudicated by the senior author (S.S.L.). None required adjudication. Error detection was evaluated by having 1 study investigator review the free-text fields of all the reviews for mention of the 5 intentionally placed errors plus the 2 inadvertent ones.

POWER ANALYSIS

A power analysis was conducted to estimate the number of subjects (peer reviewers at *JBS* and *CORR*) needed to achieve a power of 0.80 and an α value (1-tailed) of 0.05 to discern a difference in rejection rates of 15% (eg, 5% vs 20% and 10% vs 25%) between the 2 versions of the manuscript.¹⁹ One-tailed testing was chosen because to this point, there has been no evidence in the literature of a publication bias favoring no-difference results. This resulted in an estimate of the need to recruit a minimum of 118 peer reviewers for each version of the test manuscript (for a difference between 5% and 20%) to 156 peer reviewers (for a difference between 10% and 25%). The fabricated manuscript was sent to 124 reviewers at *JBS* and to 114 reviewers at *CORR*, for a total of 238 peer reviewers. At *JBS*, 102 of the 124 reviewers (82.3%) returned a review of the manuscript, and at *CORR*, 108 of the 114 reviewers (94.7%) returned a review of the manuscript. Of the 124 reviewers at *JBS*, 59 (47.6%) received the positive version and 65 (52.4%) received the no-difference version. Of the 114 reviewers at *CORR*, 62 (54.4%) received the positive version and 52 (45.6%) received the no-difference version. Differences in outcome with respect to the primary study hypothesis were observed between the 2 participating journals; thus, the results were both pooled and analyzed separately for each journal.

STATISTICAL ANALYSIS

A logistic regression analysis was performed to examine differences in the proportions of acceptance/rejection rates for (1) re-

views of each version of the manuscript, (2) each journal, and (3) an interaction effect between version and journal. Odds ratios (ORs) with accompanying 95% confidence intervals (CIs) are reported as tests for statistical significance, $P < .05$, 2-tailed.²⁰ Analysis of variance was used to test for significant differences in methods scores and number of errors detected in a 2 (version) \times 2 (journal) design. All the analyses were run using a software program (SPSS, version 17.0; SPSS Inc, Chicago, Illinois).

RESULTS

We observed consistency between the 2 journals in reviewing the positive-outcome manuscripts more favorably than the no-difference manuscripts; however, the magnitude of this effect varied and was somewhat stronger for one journal than for the other (**Table**). Overall, across both journals, 97.3% of reviewers (107 of 110) recommended accepting the positive version and 80.0% of reviewers (80 of 100) recommended accepting the no-difference version ($P < .001$; OR, 8.92; 95% CI, 2.56-31.05), indicating that the positive version of the test manuscript was more likely to be recommended for publication by reviewers than was the no-difference version.

At *CORR*, the percentages of reviewers recommending publication of the positive and no-difference versions did not differ with the numbers available (96.7% [58 of 60] vs 89.6% [43 of 48], respectively; $P = .28$; OR, 3.37; 95% CI, 0.62-18.21). In contrast, at *JBS*, more positive versions than no-difference versions of the test manuscript were recommended for publication by the reviewers (98.0% [49 of 50] vs 71.2% [37 of 52], respectively; $P = .001$; OR, 19.87; 95% CI, 2.51-157.24).

Reviewers for both journals identified more errors in the no-difference version (mean, 0.85; 95% CI, 0.68-1.03) than in the positive version (0.41; 95% CI, 0.23-0.57) ($P < .001$) (**Table**). When examining the results for each journal separately, we found that reviewers at *CORR* detected more errors in the no-difference manuscript version (mean, 1.00; 95% CI, 0.74-1.26) than in the positive version (0.52; 95% CI, 0.29-0.75) ($P = .02$). The same finding held at *JBS*; reviewers detected more errors in the no-difference version (mean, 0.71; 95% CI, 0.47-0.96) than in the positive version (0.28; 95% CI, 0.03-0.53) ($P = .005$).

Reviewers' scores for methodological validity likewise suggested the presence of POB, despite the "Methods" sections of the 2 versions being identical (**Table**). The analysis of variance for methods scores again indicated a significant effect across both journals based on outcome (positive vs no-difference) and between the 2 journals but no significant interaction effect (outcome \times journal), again showing that although the magnitude of the finding differed between the 2 journals, the direction of the finding was the same: positive-outcome manuscripts received higher scores for methodological validity than did no-difference manuscripts. Methods scores assigned by reviewers for both journals were higher for the positive version (mean, 8.24; 95% CI, 7.91-8.64) than for the no-difference version (7.53; 95% CI, 7.14-7.90) ($P = .005$).

Examining the results for each journal separately, with the numbers available we observed no difference between the *CORR* reviewers' methods scores awarded to the positive manuscript version (mean, 7.87; 95% CI, 7.38-

Table. Rates of Reviewers' Recommendations for Acceptance, Error Detection, and Methods Scores of Manuscripts With Positive vs No-Difference Findings at 2 Orthopedic Journals

Journal	Positive Version, No./Total No. (%)	No-Difference Version, No./Total No. (%)	P Value	OR (95% CI)	
Accept Manuscript					
<i>CORR</i>	58/60 (96.7)	43/48 (89.6)	.28	3.37 (0.62-18.21)	
<i>JBJS</i>	49/50 (98.0)	37/52 (71.2)	.001	19.87 (2.51-157.24)	
Total	107/110 (97.3)	80/100 (80.0)	<.001	8.92 (2.56-31.05)	
Error Detection					
	Positive Version		No-Difference Version		
	Reviewers, No.	Score, Mean (SD) [95% CI]	Reviewers, No.	Score, Mean (SD) [95% CI]	P Value
<i>CORR</i>	60	0.52 (0.68) [0.29-0.75]	48	1.00 (1.34) [0.74-1.26]	.02
<i>JBJS</i>	50	0.28 (0.45) [0.03-0.53]	52	0.71 (0.96) [0.47-0.96]	.005
Total	110	0.41 (0.60) [0.23-0.57]	100	0.85 (1.16) [0.68-1.03]	<.001
Methods Scores					
<i>CORR</i>	60	7.87 (1.81) [7.38-8.36]	48	7.38 (2.37) [6.83-7.93]	.22
<i>JBJS</i>	50	8.68 (1.21) [8.14-9.22]	52	7.66 (2.17) [7.14-8.19]	.005
Total	110	8.24 (1.61) [7.91-8.64]	100	7.53 (2.26) [7.14-7.90]	.005

Abbreviations: CI, confidence interval; *CORR*, *Clinical Orthopaedics and Related Research*; *JBJS*, *The Journal of Bone and Joint Surgery*; OR, odds ratio.

8.36) vs the no-difference manuscript (7.38; 95% CI, 6.83-7.93) ($P = .22$). In contrast, at *JBJS*, scores for methodological validity were higher for the positive manuscript version (mean, 8.68; 95% CI, 8.14-9.22) than for the no-difference version (7.66; 95% CI, 7.14-8.19) ($P = .005$).

COMMENT

We found evidence of POB in the review processes of both journals studied. Although the strength of this framing effect varied between the 2 journals, the overall effect for the pooled sample of reviewers at both journals favored the positive-outcome manuscript version in terms of more frequent recommendations to publish, less-intensive error detection, and higher methods scores.

Significant differences in the frequency of error detection suggested heightened scrutiny of the no-difference manuscript compared with that of the positive-outcome manuscript, and methods scores were significantly higher for the positive version despite the "Methods" sections being identical between the 2 test manuscript versions. Although the magnitude of these findings varied between the 2 journals surveyed, the direction of the findings was consistent across both journals and for all 3 primary study end points (recommendation to publish, error detection, and methods scores). This study was designed for analysis of the pooled sample (both journals combined); although the final sample size was slightly below the desired power threshold, the results were significant, indicating that there was no type II error in the pooled sample. The sample size was underpowered for analysis at the level of the individual journals, although despite this, significant differences also were detected at the level of the individual journals for 4 of the 6 primary study end points surveyed. It is possible that had the sample size been larger, other differences would have been significant as well.

To the extent that POB exists, it would be expected to compromise the integrity of the literature in many important ways, including, but not limited to, inflation of apparent treatment effect sizes when the published literature is subjected to meta-analysis. To our knowledge, although numerous studies^{1,3,5-10,13-16} have inferred POB by comparing denominators of studies submitted (or initiated) with those published, we identified no other experimental studies in the biomedical literature with which we could directly compare these results.

Mahoney,²¹ in 1977, reported on a form of confirmatory bias that he defined as the tendency to emphasize and believe experiences that support one's views and to ignore or discredit those that do not and used a qualitative analysis to conclude that it was present during peer review. A POB is not the same thing as a confirmatory bias as defined by Mahoney, but there may be some overlap in terms of the psychological influences, in particular, in the way that the results of a study might exert a framing effect on how a reader perceives that study's methodological rigor.²¹ A study using methods superficially similar to ours was published in 1990,²² and the data were republished with some additional reviews of a test manuscript²³; that work was widely criticized for obvious methodological and ethical shortcomings.^{24,25} The present study avoided the methodological and ethical concerns (which centered on the topic of informed consent) raised^{24,25} about that earlier work.^{22,23}

There were subtle differences between the journals studied in terms of the strength of the evidence for POB (Table). Possible explanations for the observed differences between the 2 journals are necessarily somewhat speculative but could include insufficient sample size for per-journal analysis and potential differences in manuscript management, reviewer pools, and review processes.

Reviewers detected relatively few of the intentionally placed errors. We expected this to be the case because it

was important for the study design to place enough errors to allow reasonable statistical comparisons, but we did not want the test manuscripts to suffer from the appearance of careless science or poor proofreading. Accordingly, the errors inserted were relatively subtle, and error detection was, accordingly, relatively infrequent overall. Still, the difference was statistically significant at both journals, and the effect size (comparing the frequency of error detection in the positive with the no-difference manuscript versions) was large.

It is possible that the Hawthorne effect (the idea that observed study subjects behave differently because they are being observed) may have played a role in the findings.²⁶ Two of the 238 reviewers invited to participate made comments or queries that revealed a suspicion about the test manuscript's authenticity based on a feature of the electronic manuscript distribution program that struck them as unusual. Both of these reviewers were at *CORR* (although the same program is used by *JBS*), and both returned reviews. It is possible that more reviewers noticed this finding but did not query the editors. If anything, however, one might surmise that heightened scrutiny might cause reviewers to be more, rather than less, careful about POB, so the Hawthorne effect in this form should have diminished rather than increased the effect of POB. The reviewers were not informed of the purpose of the study or the experimental hypotheses, and we received no communication at any point that suggested that they learned of these hypotheses through other means.

The journals use slightly different manuscript review formats, which required us to convert *JBS* reviewers' qualitative comments about the test manuscript's "Methods" section into quantitative scores to evaluate those comments statistically. The fact that this conversion was performed by us, and not by the reviewers themselves, could be perceived as a potential limitation of this experimental approach. However, this is not likely to have been a major limitation because the scoring was done independently by 2 investigators on blinded redacted manuscripts (so that the scorers did not know whether the comments had been made about the positive or the no-difference manuscript version), and the scorers had a high degree of interobserver agreement.

It has been proposed that bias of the sort observed by Mahoney²¹ and herein is not just a part of evidence-based medicine or peer review but is part of human cognitive behavior (finding what one seeks); indeed, Mahoney²¹ pointed this out and suggested that Francis Bacon identified the phenomenon nearly 400 years ago. Previous studies^{14,27,28} have found that the "newsworthy" (defined as a positive finding) is more likely to draw a favorable response from peer reviewers and, indeed, that work with positive outcomes is more likely to be submitted to peer review in the first place.⁴ If so, then that, along with the evidence identified in this experimental study, highlights the importance of sensitivity to this issue during peer review. It is possible that registries of prospective trials will mitigate POB at the level of scientists; journal editors should consider providing specific guidance to reviewers on the subject of the review of no-difference manuscripts to minimize the impact of POB on manuscript acceptance. Journals should specifically

encourage authors to submit high-quality no-difference manuscripts and should look for opportunities to publish them, whether in the "traditional" print versions of the journal, in online journal appendices, or in partnership with open-source media.

Accepted for Publication: May 25, 2010.

Correspondence: Seth S. Leopold, MD, Department of Orthopaedics and Sports Medicine, University of Washington, 1959 NE Pacific St, HSB BB 1053 (356500), Seattle, WA 98195-6500 (leopold@u.washington.edu).

Author Contributions: Dr Leopold had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. **Study concept and design:** Emerson, Warme, Wolf, Heckman, Brand, and Leopold. **Acquisition of data:** Emerson, Heckman, and Brand. **Analysis and interpretation of data:** Emerson, Warme, Wolf, Heckman, Brand, and Leopold. **Drafting of the manuscript:** Emerson, Wolf, Heckman, and Leopold. **Critical revision of the manuscript for important intellectual content:** Emerson, Warme, Wolf, Heckman, Brand, and Leopold. **Statistical analysis:** Wolf. **Administrative, technical, and material support:** Emerson, Heckman, and Leopold. **Study supervision:** Heckman and Leopold.

Financial Disclosure: During this study, Dr Heckman was the editor in chief at *JBS* and Dr Brand was the editor in chief at *CORR*.

Online-Only Materials: eAppendices 1 and 2 are available at <http://www.archinternmed.com>.

Additional Contributions: Joseph R. Lynch, MD, provided considerable help in the early design process and test manuscript creation; Leslie Meyer, BA, provided considerable organizational efforts; Matthew Cunningham, PhD, assisted with data analysis; and Amy Bordiuk, MA, edited the manuscript.

REFERENCES

1. Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA*. 1998;280(3):254-257.
2. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457-2465.
3. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith HJ Jr. Publication bias and clinical trials. *Control Clin Trials*. 1987;8(4):343-353.
4. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA*. 1992;267(3):374-378.
5. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008; 3(8):e3081. doi:10.1371/journal.pone.0003081.
6. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337(8746):867-872.
7. Misakian A, Bero L. Publication bias and research on passive smoking: comparison of published and unpublished studies. *JAMA*. 1998;280(3):250-253.
8. Moscati R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Acad Emerg Med*. 1994;1(3):267-271.
9. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev*. 2007;(2):MR000005.
10. Lee KP, Boyd EA, Holroyd-Leduc JM, Bacchetti P, Bero LA. Predictors of publication: characteristics of submitted manuscripts associated with acceptance at major biomedical journals. *Med J Aust*. 2006;184(12):621-626.
11. Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other char-

- acteristics associated with citation of published studies in peer-reviewed journals. *JAMA*. 2002;287(21):2847-2850.
12. Cunningham MR, Warme WJ, Schaad DC, Wolf FM, Leopold SS. Industry-funded positive studies not associated with better design or larger size. *Clin Orthop Relat Res*. 2007;457:235-241.
 13. Hasenboehler EA, Choudhry IK, Newman JT, Smith WR, Ziran BH, Stahel PF. Bias towards publishing positive results in orthopedic and general surgery: a patient safety issue? *Patient Saf Surg*. 2007;1(1):4.
 14. Lynch JR, Cunningham MR, Warme WJ, Schaad DC, Wolf FM, Leopold SS. Commercially funded and United States-based research is more likely to be published; good-quality studies with negative outcomes are not. *J Bone Joint Surg Am*. 2007;89(5):1010-1018.
 15. Okike K, Kocher MS, Mehlman CT, Heckman JD, Bhandari M. Publication bias in orthopaedic research: an analysis of scientific factors associated with publication in the *Journal of Bone and Joint Surgery (American Volume)*. *J Bone Joint Surg Am*. 2008;90(3):595-601.
 16. Weber EJ, Callahan ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA*. 1998;280(3):257-259.
 17. Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. *Stat Methods Med Res*. 2005;14(5):515-524.
 18. Moher DSK, Schulz KF, Altman DG; CONSORT GROUP (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med*. 2001;134(8):657-662.
 19. Dupont WDPW, Plummer WD Jr. Power and sample size calculations: a review and computer program. *Control Clin Trials*. 1990;11(2):116-128.
 20. Altman D. *Practical Statistics for Medical Research*. New York, NY: Chapman & Hall; 1991.
 21. Mahoney M. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognit Ther Res*. 1977;1:161-175.
 22. Epstein WM. Confirmational response bias among social work journals. *Sci Technol Human Values*. 1990;15(1):9-38.
 23. Epstein WM. Confirmational response bias and the quality of the editorial processes among American social work journals. *Res Soc Work Pract*. 2004;14(6):450-458.
 24. Feinstein AR. Construction, consent, and condemnation in research on peer review. *J Clin Epidemiol*. 1991;44(4-5):339-341.
 25. Kemper KJ. Pride and prejudice in peer review. *J Clin Epidemiol*. 1991;44(4-5):343-345.
 26. Jones SRG. Was there a Hawthorne effect? *Am J Sociol*. 1992;98(3):451-468.
 27. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA*. 1990;263(10):1392-1395.
 28. Chalmers I. Underreporting research is scientific misconduct. *JAMA*. 1990;263(10):1405-1408.

Call for Photographs

The *Archives* is seeking photographs to be included as fillers in our journal. We believe that our readers may be an excellent source of interesting and thoughtful photographs. If you would like us to consider your photography for publication, we invite you to submit your photograph to our Web-based submission site under the category Images From Our Readers at <http://manuscripts.archinternmed.com>. Please upload photograph submissions in .jpg or .tif format. Hard copy photographs are not acceptable. For more information please e-mail archinternmed@jama-archives.org.