4-2008

The Ombudsman: Verification of Citations: Fawlty Towers of Knowledge?

Malcom Wright
University of South Australia, malcolm.wright@marketingscience.info

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

Suggested Citation:
Wright, M., J. S. Armstrong. 2008. The Ombudsman: Verification of Citations: Fawlty Towers of Knowledge? (with commentary and reply). Interfaces 38(2) 125-139.

http://interfaces.journal.informs.org/

The Ombudsman: Verification of Citations: Fawlty Towers of Knowledge?

Abstract
The prevalence of faulty citations impedes the growth of scientific knowledge. Faulty citations include omissions of relevant papers, incorrect references, and quotation errors that misreport findings. We discuss key studies in these areas. We then examine citations to "Estimating nonresponse bias in mail surveys," one of the most frequently cited papers from the Journal of Marketing Research, to illustrate these issues. This paper is especially useful in testing for quotation errors because it provides specific operational recommendations on adjusting for nonresponse bias; therefore, it allows us to determine whether the citing papers properly used the findings. By any number of measures, those doing survey research fail to cite this paper and, presumably, make inadequate adjustments for nonresponse bias. Furthermore, even when the paper was cited, 49 of the 50 studies that we examined reported its findings improperly. The inappropriate use of statistical-significance testing led researchers to conclude that nonresponse bias was not present in 76 percent of the studies in our sample. Only one of the studies in the sample made any adjustment for it. Judging from the original paper, we estimate that the study researchers should have predicted nonresponse bias and adjusted for 148 variables. In this case, the faulty citations seem to have arisen either because the authors did not read the original paper or because they did not fully understand its implications. To address the problem of omissions, we recommend that journals include a section on their websites to list all relevant papers that have been overlooked and show how the omitted paper relates to the published paper. In general, authors should routinely verify the accuracy of their sources by reading the cited papers. For substantive findings, they should attempt to contact the authors for confirmation or clarification of the results and methods. This would also provide them with the opportunity to enquire about other relevant references. Journal editors should require that authors sign statements that they have read the cited papers and, when appropriate, have attempted to verify the citations.

Comments
Suggested Citation:
Wright, M., J. S. Armstrong. 2008. The Ombudsman: Verification of Citations: Fawlty Towers of Knowledge? (with commentary and reply). Interfaces 38(2) 125-139.

http://interfaces.journal.informs.org/

Interfaces
Vol. 38, No. 2, March–April 2008, pp. 125–139
ISSN 0092-2102 | EISSN 1526-551X | 08 | 3802 | 0125

informs

DOI 10.1287/inte.1070.0317

The Ombudsman: Verification of Citations: Fawlty Towers of Knowledge?

Malcolm Wright
Ehrenberg-Bass Institute, University of South Australia, Adelaide, South Australia, malcolm.wright@marketingscience.info
J. Scott Armstrong
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, armstrong@wharton.upenn.edu

The prevalence of faulty citations impedes the growth of scientific knowledge. Faulty citations include omissions of relevant papers, incorrect references, and quotation errors that misreport findings. We discuss key studies in these areas. We then examine citations to "Estimating nonresponse bias in mail surveys," one of the most frequently cited papers from the *Journal of Marketing Research*, to illustrate these issues. This paper is especially useful in testing for quotation errors because it provides specific operational recommendations on adjusting for nonresponse bias; therefore, it allows us to determine whether the citing papers properly used the findings. By any number of measures, those doing survey research fail to cite this paper and, presumably, make inadequate adjustments for nonresponse bias. Furthermore, even when the paper was cited, 49 of the 50 studies that we examined reported its findings improperly. The inappropriate use of statistical-significance testing led researchers to conclude that nonresponse bias was not present in 76 percent of the studies in our sample. Only one of the studies in the sample made any adjustment for it. Judging from the original paper, we estimate that the study researchers should have predicted nonresponse bias and adjusted for 148 variables. In this case, the faulty citations seem to have arisen either because the authors did not read the original paper or because they did not fully understand its implications. To address the problem of omissions, we recommend that journals include a section on their websites to list all relevant papers that have been overlooked and show how the omitted paper relates to the published paper. In general, authors should routinely verify the accuracy of their sources by reading the cited papers. For substantive findings, they should attempt to contact the authors for confirmation or clarification of the results and methods. This would also provide them with the opportunity to enquire about other relevant references. Journal editors should require that authors sign statements that they have read the cited papers and, when appropriate, have attempted to verify the citations.

Key words: citation errors; evidence-based research; nonresponse bias; quotation errors; surveys.

The growth of scientific knowledge requires the *correct* reporting of relevant studies. Unfortunately, current procedures give little assurance that authors of papers published in leading academic journals follow this practice. Instead, the evidence suggests that researchers often do not read the relevant research papers. This manifests itself in two ways: First, researchers overlook relevant papers. Second, they make errors when reporting on the papers, either through incorrect referencing or incorrect quotation of the contents of the cited paper.

This problem is described in other scientific disciplines (e.g., MacRoberts and MacRoberts 1989); however, there is little work on reporting errors in the management science literature. We review prior literature relevant to these problems and then analyze a highly cited methodological paper to identify the tendency toward faulty citations in management research.

Prior Evidence: Do Researchers Read Relevant Papers?

Omissions: Authors often overlook relevant research. Sometimes this occurs because they search for evidence only within their own discipline. In addition, they often ignore papers that provide contradictory evidence or views. For example, in a study on escalation bias, papers that supported commonly held beliefs were cited nine times as frequently as those that conflicted with common beliefs (Armstrong 1996). Franke (1996) reports a similar finding for the Hawthorne studies, in which papers with opposing

views have little impact on management thinking. We confirmed this claim by analyzing the citation rates for key papers on the Hawthorne studies using the ISI Citation Index in July 2006. Roethlisberger and Dickson's (1939) original book showed over 350 citations for that and subsequent editions. Work that criticized these results, Parsons (1974) and Franke and Kaul (1978), showed 71 citations. We checked with Franke to verify that we cited his work correctly. He directed us to broader literature, and noted that Franke (1980) provided a longer and more technically sophisticated criticism; this later paper has been cited in the ISI Citation Index just nine times as of August 2006. MacRoberts and MacRoberts (1986) analyzed overlooked research by examining 15 articles on the history of genetics. They found that these 15 articles required 719 references for adequate coverage of prior research; however, only 216 (30 percent) of these 719 were actually cited in their sample. Individual articles cited between zero and 64 percent of relevant references.

Incorrect references: Errors in the citation of references are common. For example, we found that 14 percent of the 350 citations to Roethlisberger and Dickson (1939) incorrectly reported Roethlisberger's initials. This problem has been extensively studied in the health literature. More generally, Eichorn and Yankauer (1987) found that 31 percent of the references in public health journals contained errors, and three percent of these were so severe that the referenced material could not be located. Doms (1989) found that 42 percent of references in dental journals were inaccurate—30 percent of these were major errors, such as incorrect journal titles, article titles, or authors. Evans et al. (1990) studied 150 randomly selected references cited in three medical journals and found a 48 percent error rate. Other studies have found error rates of 56 to 67 percent in obstetrics and gynecology journals (Roach et al. 1997), 32 percent in nursing journals, including 43 major errors in the 180 references examined (Schulmeister 1998), 40 percent in otolaryngology/head and neck surgery journals, with 12 percent major errors (Fenton et al. 2000), 36 percent in manual therapy journals (Gosling et al. 2004), and 34 percent in biomedical informatics journals (Aronsky et al. 2005). Schulmeister (1998) includes a summary of earlier literature in this area. This problem is serious even for the most prestigious journals. Lok et al. (2001) found that highly rated journals contained fewer minor mistakes but just as many major errors.

Quotation errors: Substantive errors that misreport findings are more damaging than errors in references. We refer to these as *quotation errors*. DeLacey et al. (1985) found quotation errors in 15 percent of the references cited in six medical journals. Twelve percent of references involved errors that were misleading or seriously misrepresentative. Eichorn and Yankauer (1987) found that authors' descriptions of previous studies in public health journals differed from the original copy in 30 percent of references; half of these descriptions were unrelated to the quoting authors' contentions. The detailed analysis that Evans et al. (1990) did of quotation errors in surgical journals raised concern, in many cases, that the original reference was not even read by the authors. Schulmeister (1998) found 12 out of 180 nursing articles examined contained major quotation errors. In another medical specialization, Fenton et al. (2000) found quotation errors for 17 percent of references including major quotation errors for 11 percent of references. Wager and Middleton (2003), in a systematic review of medical journals, concluded that 20 percent of the quotations were incorrect. Lukic et al. (2004) examined three anatomy journals and found that 19 percent of the quotations were incorrect: shockingly, nearly all of these involved major errors. Gosling et al. (2004) found quotation errors in 12 percent of references in a study of manual therapy journals.

Analysis of a Highly Cited Paper

We examined the citation history of "Estimating nonresponse bias in mail surveys" by Armstrong and Overton (1977)—we will refer to this as A&O. This was the third-most-cited article in the *Journal of Marketing Research* with 963 citations in the ISI Citation Index at the time of our analysis in 2006. This was a suitable article for our exploratory analysis of citation errors because it is highly cited and because it makes clear, methodological recommendations that are easy to verify.

A&O Recommendations

A&O sought to develop methods for dealing with nonresponse bias in mail surveys. They relied on the concept that nonresponders are more similar to late responders than to early responders. Those who respond initially to a mail survey are most interested in the topic; thus, nonresponse bias would only apply to those items that are most closely related to the topic. For example, if the survey dealt with intentions to purchase a new product, those most interested in the product would be in the first wave to respond. Those in the second wave (that is, they respond to a follow-up plea) would presumably be less interested in the new product. Nonresponse bias would not be expected for other items such as demographic questions.

A&O recommended an adjustment for nonresponse bias only when the direction of bias that experts expected is consistent with the observed trend across response waves. They assessed their method by analyzing previously published results for 136 items from 16 studies. These studies had median sample sizes of 1,000 for the first wave and 770 for the second wave. Of these items, 54 percent showed statistically significant biases or differences between the waves. A consensus of judges correctly predicted the direction of 64 percent of these biased items, with 32 percent of items overlooked and 4 percent predicted incorrectly. A combination of judgment and extrapolation correctly predicted the direction of 60 percent of biased items. Incorrect predictions dropped to two percent.

When the consensus of judges and extrapolation agreed, indicating that adjustments should be made for nonresponse bias, A&O undertook correction by extrapolating from the first and second wave responses. They assessed the accuracy of the extrapolated figures by comparing them with the results of a third response wave. A&O's method reduced the mean absolute percentage error (MAPE) due to nonresponse from 4.8 to between 3.3 and 2.5, depending on the particular method of extrapolation. This represents an error reduction of between 31 percent and 48 percent, respectively.

Failure to Include Relevant Studies

In survey research, it has been standard practice for well over half a century to report on sampling error. In contrast, few studies assess errors due to nonresponse. Because such errors are likely to occur in nearly every survey, and because they are often large, it would seem that survey studies should report on the possibility of nonresponse bias and adjust accordingly using proper procedures.

To assess whether papers involving mail surveys report on nonresponse bias, we conducted Google searches in August 2006. First, we looked at surveys that commercial firms as well as academics conducted. We expected that the volume of commercial studies would be enormous in comparison to the academic studies. However, both cases warrant careful scientific analysis. Using the terms "(mail OR postal) survey" and either "results OR findings," we obtained slightly over one million results from our Google searches. We expect that this underestimates the number of surveys conducted because most studies are not posted on the Internet.

To determine the attention given to nonresponse bias, we then added "(nonresponse OR non-response) (error OR bias)" to the search criteria. This yielded 24,900 sites. Thus, fewer than three percent of the one million surveys made obvious attempts to mention, let alone address, the issue of nonresponse bias.

To our knowledge, the A&O paper is the only source of an evidence-based procedure for adjusting for nonresponse bias; thus, it presents an ideal test of the percentage of papers that should have cited it. We refined the above search criteria by including "Armstrong" and "Overton." This search yielded 348 sites, merely 1.4 percent of the 23,000 websites. In other words, more than 98 percent of these studies do not mention A&O's evidence-based procedure for adjusting for nonresponse bias even when they recognize nonresponse bias as a potential problem.

In contrast, we would expect academic researchers to be more thorough. Furthermore, experts review their work. Thus, we investigated academic citations of A&O by conducting identical searches using Google Scholar. We located 27,300 websites initially. Of these, 1,600 (about six percent) mentioned nonresponse. While this is an improvement over the general search results, 94 percent of academic research still failed to mention nonresponse bias. Of those that did, we found 339 (2.1 percent) articles that also mentioned A&O.

Our method for assessing the extent to which A&O was improperly excluded is quite unrefined. For example, the above search on Google Scholar accounted for only about one-third of the A&O citations. Equally, some authors who did not mention A&O might argue that nonresponse bias is less relevant for theoretical tests than for population estimates, or that A&O provides no assistance for correcting correlations. However, the findings are so extreme that we can confidently state that researchers routinely fail to consider even the possibility of nonresponse bias. Of those who do consider it, few look for evidence on how to address the issue.

Incorrect References

We examined errors in the references of papers that cite A&O. To do this, we used the ISI Citation Index (in August 2006). We expected this index to underrepresent the actual error rate because the ISI data-entry operators may correct many minor errors. In addition, articles not recognized as being from ISI-cited journals do not have full bibliographic information recorded; therefore, they will also omit errors in the omitted information. Despite this, we found 36 variations of the A&O reference. Beyond the 963 correct citations, we found 80 additional references that collectively employed 35 incorrect references to A&O. Thus, the overall error rate was 7.7 percent.

Quotation Errors

A&O is ideal for assessing the accuracy of how the findings were used because it provides clear operational advice on how to constructively employ the findings. We examined 50 papers that cited A&O, selecting a mix of highly cited and recently published papers. We included the 30 most frequently cited papers of the 1,184 that cited A&O (as provided by a Google Scholar search). Unlike the ISI Citation Index, Google Scholar allowed us to sort citing papers by the number of citations they had received in turn. In sum, our sample of 50 papers received a total of 3,024 Google Scholar citations at the time of analysis in May 2006. The typical article citing A&O said something similar to: "Assuming that nonresponders will be similar to late responders, we tested for differences between early and late respondents on key variables; we found no significant differences, suggesting that nonresponse bias is not a problem in this study."

We instructed a research assistant to obtain copies of the articles in our sample and create a database that recorded the articles' bibliographical details, sample size, response rate, and the sentence or paragraph that cited A&O. The first author coded the records in the database to determine the following information:

(1) Whether the article mentioned A&O's procedures (expert judgment, time-series extrapolation, and consensus between expert judgment and extrapolation).

(2) Whether the article mentioned possible differences between early and late respondent groups.

(3) Whether the article reported significance testing to check for nonresponse bias.

(4) How many biased variables the article identified.

(5) How many biased variables the article corrected.

We then asked a second research assistant to independently repeat the coding as a reliability check. Inter-coder agreement was 94 percent. The second author resolved the remaining 21 (six percent) disagreements with a further blind-coding of these items. Details are provided at jscottarmstrong.com under "publications;" see "codings" following the working paper version.

Of the articles in our sample, 46 mentioned differences between early and late respondents. This indicates some familiarity with the consequences of the interest hypothesis. However, only one mentioned expert judgment, only six mentioned extrapolation, and none mentioned consensus between techniques. In short, although there were over 100 authors and more than 100 reviewers, all the papers failed to adhere to the A&O procedures for estimating nonresponse bias. Only 12 percent of the papers mentioned extrapolation, which is the key element of A&O's method for correcting nonresponse bias. Of these, only one specified extrapolating to a third wave to adjust for nonresponse bias.

In contrast, the techniques we employed within our sample were quite different than those that A&O recommended. Forty-two of the studies (84 percent) reported statistical testing for differences between early and late responses and seven of the other eight studies reported looking for "noticeable patterns," "differences," or conducting "tests" between early and late

respondents without specifying the exact procedures they used.

A&O did not recommend the use of statistical tests to detect nonresponse bias. Such tests would be expected to harm decision making in this situation as Armstrong (2007) explains; he cites prior research showing misrepresentation of significance testing by researchers and reviewers, and notes dangers arising from (1) bias against nonsignificant findings (in this case, bias would be against significant findings), (2) inappropriate selection of a null hypotheses, and (3) distraction from key issues.

A&O did use statistical tests to assess the accuracy of judgment in predicting the direction of bias. This was part of their validation of the accuracy of judgment, not part of their recommendation for detecting bias and adjusting for nonresponse. In A&O's validation of judgment, the combined sample sizes for the two waves had a median of 1,770. The studies we examined had a median sample size of 197. These studies exhibited variation in the division of their samples; some samples were divided into thirds rather than halves, some into early and late quartiles, and some used other percentage divisions smaller than a half. A test for differences between such small subgroups is pointless. Its purpose appears to be to assure reviewers that there is no significant difference; yet, the null hypothesis has no reasonable chance of being rejected. This procedure distracts from the more important issue of improving the survey estimates.

Was nonresponse bias likely to be a problem in these studies? In a review article on the problem of nonresponse, Gendall (2000) concluded that a rough rule of thumb was that a response rate of 50 percent was a minimum acceptable level. However, he noted that this did not apply to all surveys or all variables. For example, surveys with response rates of up to 70 percent could still have the potential for serious nonresponse bias on particular topics, such as contentious social issues. Gendall (2000) stated that the only certain way to reduce the potential for nonresponse bias was to increase response rates. (Gendall did not cite the A&O procedure.)

Despite Dillman's (2000) long-established findings that demonstrate how to achieve high response rates in mail surveys, the median response rate for our sample was 30 percent. Only six studies had response rates of 50 percent or greater. Thus, there is a prima facie case for nonresponse bias among the 88 percent of surveys with response rates of less than 50 percent (note that two studies reported two surveys). Prior knowledge supports this expectation. A&O found nonresponse bias present for 54 percent of the 136 items from 16 studies that they analyzed. In contrast, only 12 studies (24 percent) in our sample reported nonresponse bias and only one attempted a correction. Based on A&O's results, we would expect 4.6 (that is, $0.54 * 136/16 = 4.6$) biased items per study. A&O's procedures would detect and adjust bias for 62 percent or 2.9 of these items per study. Therefore, the studies in our sample should have made adjustments for nonresponse bias to 148 variables in total. Such adjustments would have substantially improved the accuracy of the findings.

Clearly, when respondents are more likely to reply because they are interested in a key variable, researchers should try to (1) increase response rates, and (2) estimate the effect of nonresponse. Prior research has shown that, on average, about five such biased variables exist in each mail survey.

Discussion

Our findings raise questions that do not have good answers. Did the authors actually read the A&O paper? If they read the paper did they understand it? Why didn't the reviewers understand that the authors were not correctly adjusting for nonresponse bias?

The A&O paper seemed to be understandable. The readability index for this paper is 19 on the Gunning-Fog index, and 12 on the Flesch-Kincaid grade level. On that basis, it would be well within the capability of those (often PhDs) who conducted the studies that cited A&O. Had the citing authors been confused, one would have expected them to contact Armstrong or Overton. None did so.

To ensure that the recommendations from A&O were clear, we presented a problem to four marketing faculty members and two undergraduate research assistants. We asked them to read excerpts from the paper and to then take appropriate action given the results from two waves of a survey on a proposed

"minicar mass transit system." They reported spending from 5 to 20 minutes on the problem. One faculty member and one research assistant were not able to understand our summary. The others all properly applied the A&O adjustments. None of them used tests of statistical significance in approaching the problem.

Given the understandability of the recommendations and the fact that no one contacted Armstrong or Overton for clarification, one might question whether the citing authors read the A&O paper. To present their studies in a more favorable light, some authors may have wanted to dispel concerns about nonresponse bias; thus, they cited A&O for support for their own procedures. Interestingly, one of our colleagues said that it is common knowledge that authors add references that they have not read in order to gain favor with reviewers. One wonders: If it is possible to write a paper without reading the references, why should the authors expect readers to read the references?

When we circulated an earlier version of this paper, we received further comments about faulty citations. We show some of these below:

> "I know from my own experience that quotation errors often occur; if you want to know what someone has found, you have to go back to the original paper."

> "I've been amazed by what citation errors I've uncovered . . . less than 50% of (subsequently) cited articles 'get it' (i.e., . . . one of the main findings), or in some cases justify their whole paper's approach on an unsubstantiated propositional paragraph in another article."

> "One search for the source citation of a brand-extension 'fit' dimension . . . cited directly by three, cited in turn by hundreds, is stalled, with a (retiring) senior working paper collections librarian recalling that the paper was never lodged, let alone currently held."

> "I probably did not pay attention in graduate school and so was unaware of your 1977 article on nonresponse, but when I was doing the study described in the attached article, I consulted standard MR text books where the trend analysis is described. Could it be that many other authors simply look up how to handle nonresponse in the MR text books and that is a source of their blunders?"

> "Occasionally, journal referees complain that one of my manuscripts lacks a report on nonresponse bias. If I receive such a complaint, then I trot out the A&O reference and state something like the following in my exposition: *T*-tests revealed that the last 10% of returned questionnaires did not differ meaningfully from the first 90% of returned questionnaires: therefore, the effect of nonresponse bias is minimal. In other words, I only resort to citing A&O and making such a report because I've seen such reports repeatedly in other articles and they seem to satisfy reviewer concerns about nonresponse bias. I've read A&O . . . I agree it's been misused. However, if I believe a referee is mistaken in his/her concern, and I know a way to defuse that mistaken concern without telling the referee that he/she is mistaken, then I will use that way because the probability of surviving the review process decreases when referee concerns are challenged rather than accepted."

Speculating on Possible Solutions

The primary problem is that researchers fail to build upon prior evidence-based research and the journal reviewing process does not require them to do so. Researchers may sometimes not be aware of all the relevant work. However, a large percentage of researchers apparently fail to read many of the papers of which they are aware and do cite. As a result, we expect that most references in papers are spurious.

The Internet offers a solution to problems of omission. Journals should open websites (free to nonsubscribers) that allow people to post key papers that have been overlooked, along with a brief explanation of how the findings relate to the published study.

The problem of quotation errors has a simple solution: When an author uses prior research that is relevant to a finding, that author should make an attempt to contact the original authors to ensure that the citation is properly used. In addition, authors can seek information about relevant papers that they might have overlooked. Such a procedure might also lead researchers to read the papers that they cite. Editors could ask authors to verify that they have read the original papers and, where applicable, attempted to contact the authors. Authors should be required to confirm this prior to acceptance of their paper. This requires some cost, obviously; however, if scientists expect people to accept their findings, they should verify the information that they used. The key is that reasonable verification attempts have been made. Despite the fact that compliance is a simple matter, usually requiring only minutes for the cited author to

respond, Armstrong, who has used this procedure for many years, has found that some researchers refuse to respond when asked if their research was being properly cited; a few have even written back to say that they did not plan to respond. In general, however, most responded with useful suggestions and were grateful that care was taken to ensure proper citation.

We attempted to contact via e-mail 12 authors that we cited in this paper. Six replied, most with useful comments. One author noted that it was very challenging to represent all the papers in this area due to the high volume of work. Another provided us with a list of 60 relevant references, as well as an updated version of her own systematic review, which we cite. One of the authors disagreed with our proposed solution due to the perceived likelihood of contact information becoming obsolete and the potential drain on researchers' time. Our own contact attempts were successful enough that we remain confident in our recommendations.

Conclusions

As we expected, researchers fail to cite relevant research studies. Prior research suggests that there are many problems in reporting on prior research. This includes both omissions of relevant papers and a failure to understand (or even read) many of the papers that researchers cite.

In the case of the A&O paper, we estimated that far less than one in a thousand mail surveys consider evidence-based findings related to nonresponse bias. This has occurred even though the paper was published in 1977 and has been available in full text on the Internet for many years. Furthermore, the paper is easy to find; if one searches Google for "nonresponse bias" and "mail surveys," the A&O paper turns up as the first of over 21,000 websites.

When we investigated a sample of studies that cited A&O, we found 98 percent did so in an improper manner. Instead of following A&O's procedures, 84 percent of our sample inappropriately used statistical-significance tests to examine nonresponse bias. Only 24 percent of our sample detected nonresponse bias, and only one attempted a correction. As a result, most of these papers provided inadequate estimates and falsely claimed that their findings were properly supported. Collier and Bienstock (2007) obtained similar findings; in their examination of three leading marketing journals from 1999 through 2003, only four percent of the 481 studies with surveys "found a statistically significant difference between respondents and nonrespondents" (p. 177). One might think that nonresponse bias is rare.

The net result is that whereas evidence-based procedures for dealing with nonresponse bias have been available since 1977, they are properly applied only about once every 50 times that they are mentioned, and they are mentioned in only about one out of every 80 academic mail surveys. Thus, we estimate that only one in 4,000 academic mail surveys properly applies A&O's adjustments for nonresponse bias. It may be that some of the other 3,999 studies rely on high response rates, demographic comparisons where expectations about the direction of bias are judged to be obvious, or some other evidence-based procedure to address the threat of nonresponse bias. The first author, Wright, has adopted such approaches in a number of studies, having previously overlooked A&O's correction procedure, and having disregarded the reported method of statistical tests for differences between response waves as wrong. Yet, even if our estimates are too pessimistic by a factor of 1,000, we still face a major problem. It also raises questions about the quality of data in over a million commercial mail-survey research studies.

In many respects, the A&O paper was ideal for identifying any tendency toward faulty citation. However, we believe that this problem is pervasive in the social sciences. We find it difficult to read papers in our areas without noting that the researchers have overlooked key papers. In addition, reference lists include a large number of irrelevant papers, raising the question of whether the authors had read or understood those papers. This raises questions about the adequacy of the quality-control system used in science publications. Procter & Gamble advertised "$99\frac{44}{100}$% Pure" for Ivory soap and supported the claim with regular laboratory tests. In contrast, our research on the use of evidence-based findings in mail-survey research shows that it is more than "$99\frac{44}{100}$% Impure" with respect to nonresponse bias.

Authors should read the papers they cite. In addition, authors should use the verification of citations procedure. This means that they should attempt to contact original authors to ensure that they properly cite any studies they rely on to support their main findings. Journal editors should require authors to confirm that they have read the papers that they have cited and that they have made reasonable attempts to verify citations. This will help to reduce errors in the reference list, reduce the number of spurious references, and reduce the likelihood of overlooking relevant studies. Finally, once a paper has been published, journals should make it easy for researchers to post relevant studies that have been overlooked. These procedures should help to ensure that new studies build properly on prior research.

Acknowledgments

We thank D. Aronsky, R. H. Franke, P. Gendall, R. Goldsmith, B. Healey, R. N. Kostoff, M. Hyman, I. Lukic, D. Mather, D. Russell, S. Shugan, and L. Wager for comments on earlier versions of this paper. We also thank the Wharton School and the Ehrenberg-Bass Institute for their research-assistant support.

References

Armstrong, J. S. 1996. Management folklore and management science: On portfolio planning, escalation bias, and such. *Interfaces* **26**(4) 28–42.

Armstrong, J. S. 2007. Significance tests harm progress in forecasting. *Internat. J. Forecasting* **23** 321–327.

Armstrong, J. S., T. S. Overton. 1977. Estimating nonresponse bias in mail surveys. *J. Marketing Res.* **14**(3) 396–402.

Aronsky, D., J. Ransom, K. Robinson. 2005. Accuracy of references in five biomedical informatics journals. *J. Amer. Medical Informatics Assoc.* **12**(2) 225–228.

Collier, J. E., C. C. Bienstock. 2007. An analysis of how nonresponse error is assessed in academic marketing research. *Marketing Theory* **7**(2) 163–183.

DeLacey, G., C. Record, J. Wade. 1985. How accurate are quotations and references in medical journals? *British Medical J.* **291** 884–886.

Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method*, 2nd ed. John Wiley & Sons, New York.

Doms, C. A. 1989. A survey of reference accuracy in five national dental journals. *J. Dental Res.* **68**(3) 442–444.

Eichorn, P., A. Yankauer. 1987. Do authors check their references? A survey of accuracy of references in three public health journals. *Amer. J. Public Health* **77**(8) 1011–1012.

Evans, J. T., H. I. Nadjari, S. A. Burchell. 1990. Quotational and reference accuracy in surgical journals: A continuing peer review problem. *J. Amer. Medical Assoc.* **263**(10) 1353–1354.

Fenton, J. E., H. Brazier, A. De Souza, J. P. Hughes, D. P. McShane. 2000. The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. *Clinical Otolaryngology* **25** 40–44.

Franke, R. H. 1980. Worker productivity at Hawthorne. *Amer. Sociol. Rev.* **45**(6) 1006–1027.

Franke, R. H. 1996. Advancing knowledge, a comment from Richard H. Franke. *Interfaces* **26**(4) 52–57.

Franke, R. H., J. D. Kaul. 1978. The Hawthorne experiments: First statistical interpretation. *Amer. Sociol. Rev.* **43**(5) 623–643.

Gendall, P. 2000. Responding to the problem of nonresponse. *Australasian J. Marketing Res.* **8**(1) 3–18.

Gosling, C. M., M. Cameron, P. F. Gibbons. 2004. Referencing and quotation accuracy in four manual therapy journals. *Manual Therapy* **9** 36–40.

Lok, C. K. W., M. T. V. Chan, I. M. Martinson. 2001. Risk factors for citation errors in peer-reviewed nursing journals. *J. Adv. Nursing* **34**(2) 223–229.

Lukic, I. K., A. Lukic, V. Gluncic, V. Katavic, V. Vucenik, A. Marusic. 2004. Citation and quotation accuracy in three anatomy journals. *Clinical Anatomy* **17** 534–539.

MacRoberts, M. H., B. R. MacRoberts. 1986. Quantitative measures of communication in science: A study of the formal level. *Soc. Stud. Sci.* **16** 151–172.

MacRoberts, M. H., B. R. MacRoberts. 1989. Problems of citation analysis: A critical review. *J. Amer. Soc. Inform. Sci.* **40**(5) 342–349.

Parsons, H. M. 1974. What happened at Hawthorne? *Science* **183** 922–932.

Roach, V. J., T. K. Lau, D. Warwick, N. Kee. 1997. The quality of citations in major international obstetrics and gynecology journals. *Amer. J. Obstetrics and Gynecology* **177**(4) 973–975.

Roethlisberger, F. J., W. J. Dickson. 1939. *Management and the Worker*. Harvard University Press, Cambridge, MA.

Schulmeister, L. 1998. Quotation and reference accuracy of three nursing journals. *Image: J. Nursing Scholarship* **30**(2) 143–147.

Wager, E., P. Middleton. 2003. Technical editing of research reports in biomedical journals. The Cochrane Database of Methodology Reviews. **1** Art. MR000002, DOI: 10.1002/14651858MR000002. [This version first published online: January 20, 2003 in Issue 1. Date of most substantive amendment: November 4, 2002.]

Comment: Errors Galore

Don A. Dillman
Washington State University, Pullman, Washington 99164-4014, dillman@wsu.edu

In 1978, I authored a book, *Mail and Telephone Surveys: The Total Design Method* (Dillman 1978). According to the ISI Citation Indexes, it has now been cited in the scientific literature approximately 4,000 times. When reviewing a summary of its citations, I discovered citations showing publication dates in 24 different years, including 1907 (once) and 1908 (three times). Citations erroneously listed it as having been published in all but three of the years between 1971 and 1995; there were 102 such citations. In addition, 10 citations showed it as having been published in 1999 or 2000. I attribute the latter two years to authors who intended to cite the second edition—although I had changed the title to *Mail and Internet Surveys: The Tailored Design Method* (Dillman 2000).

I discovered 29 different titles for the book, including mail descriptors such as *main, mall, mial, mailed*, and *mailback*. The telephone descriptor also had creative spellings; they included *telephon, teleophone, telephones, telephone*, and *elephone*. Not surprisingly, my name was also frequently misspelled as Dillon, Dilman, Dill, and probably others than I was unable to find. I also discovered that I had been given many new middle initials. A similar pattern of inaccuracies has also emerged with the second edition of the book.

When a friend introduced me to the Social Science Citation Index in the late 1980s, I thought it would be useful to track citations of the book as a way of learning about its impact on research by other authors. My intent was to collect new ideas for incorporation into the second edition. To accomplish that, I asked a graduate student to take the list of citations for a recent year to the University library and copy every article he could find that cited my book.

When I began to review the large stack of articles he placed on my desk, I became dismayed quickly. Most of the citations were perfunctory, e.g., "I used the TDM (or Dillman) method to collect data." The TDM that I had developed and methodically tested involved simultaneously focusing on all aspects of survey design that seemed likely to influence response rates and quality. Specific features of the TDM ranged from number and timing of contacts and details of questionnaire design to the personalization of all communications and use of stamped return envelopes. Examination of the survey procedures that the citing authors used revealed that they often bore little resemblance to the TDM procedures I had described in the book. I concluded that, in some instances, it was unlikely the authors had read the book. After spending a few hours looking at the results of this preliminary foray through the literature, I decided to discontinue the effort and reassign the research assistant to other work. As a consequence of this exercise in frustration, I also developed considerable skepticism toward citation counts in the annual review process as a way of evaluating the actual impact of an author's work.

Wright and Armstrong are right! There is a substantial citation problem, which must be corrected. As Wright and Armstrong emphasize, the problem includes inaccurate use of procedures described in the publication being cited. In addition, errors in citations are also disconcerting, although in some cases they may be humorous, e.g., I was cited as authoring a book written 34 years before I was born! Their paper caused me to reflect on why journals emphasize the use of citations so heavily and also on the process for generating such citations.

Sometimes, an author adds citations to help the reader find literature that is relevant to a research problem; there is no intent to use the particular procedure that the authors being cited advocate. In other cases, such as the use that Wright and Armstrong focus on, authors ostensibly, but often inaccurately, use citations to describe their own work. Or, they add citations because editors and reviewers ask for them and these aspiring authors look for a path of least resistance to successful publication. In addition, editors often assign page limits to rewritten papers. Sometimes, this restriction encourages the addition of certain citations to eliminate detailed descriptions in a paper judged too long. In other cases, it leads to

the deletion of relevant citations to meet the obligatory page limits. In addition, authors may cite papers (particularly their own and those of colleagues) gratuitously to draw attention to this work. The reasons for citations to a particular work vary greatly.

Wright and Armstrong have reported an interesting case study. They showed that virtually all of a sample of 50 articles that cite a much-referenced paper by Armstrong and Overton (1977) failed to adhere to the procedures that the paper recommended for estimating nonresponse bias. It is not clear to me why the authors cited the A&O paper, whether they intended to estimate nonresponse bias, even whether the procedure was relevant to their study findings. I suspect authors and editors could disagree on the importance of doing so. Nonetheless, the analysis does suggest that there can be a huge discrepancy between citing an article and the impact that citing has on author decisions.

Wright and Armstrong propose that whenever an author uses prior research to support a finding, that author should attempt to contact the original authors to ensure that the citation is properly used. My initial reaction is that I hope this does not happen. I do not relish the thought of having authors ask (and expect) me to spend time responding individually to questions about whether they have cited my work correctly, especially because the objectives of articles and citations vary tremendously. Sometimes, nonresponse bias may be central to an analysis; in other instances, it may be tangential to the hypothesis being tested. I can envision such blanket requests evolving into sidebar debates between author and editor about what constitutes proper citation and proper use of other people's work. I would prefer to leave that activity to the editorial process, which often handles it by sending papers for review to authors who are cited on the issues most central to the paper.

Nonetheless, Wright and Armstrong have written a valuable paper. It brings attention to the problem of gratuitous and inaccurate citations that should not be part of scientific writing. The problem has many dimensions, only a few of which the authors develop in this paper. However, I hope its publication will heighten wider awareness of these potential problems and stimulate more work on finding solutions. I also hope this paper causes authors and editors to be more vigilant about citations and how their use supports the writing process.

Meantime, this author has already vowed to check his submitted papers once more for relevance to each cited paper and reasons for the citation. He will also check spellings and ensure that he is not citing another author for having written a paper several decades before he was born.

References

Armstrong, J. S., T. S. Overton. 1977. Estimating nonresponse bias in mail surveys. *J. Marketing Res.* **14**(3) 396–402.

Dillman, D. A. 1978. *Mail and Telephone Surveys: The Total Design Method*. John Wiley & Sons, New York.

Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. John Wiley & Sons, New York.

Comment: Omission and Redundancy in the Use of Citations

Mark D. Uncles
School of Marketing, Australian School of Business, University of New South Wales, Sydney NSW 2052, Australia,
m.uncles@unsw.edu.au

Omitted Citations

Wright and Armstrong claim that "authors often overlook relevant research." I share their concern. One reason is that people search for evidence only within their own discipline. As a geographical scientist now working in marketing and management, this is particularly apparent to me in areas such as diffusion research. It is curious how marketers with an interest in the diffusion of ideas and products make few direct references to work published in epidemiology or geographical

science. There appears to be ignorance of seminal work by Bailey, Bartholomew, Bartlett, Cliff, Hagerstrand, Haggett, Kendall, Kermack, McKendrick, and Ord (among others). Unfamiliarity with other disciplines explains this partially. In addition, people might consciously filter out work from other disciplines in the belief that cross-disciplinary citations unduly complicate the peer-review process. Will the editor feel obliged to send the paper to reviewers from different disciplines? Will this expose the paper to a variety of controversies across disciplines? Will this, in turn, lessen the chances of successful publication? A less generous explanation is that people omit evidence from other disciplines to make their own works appear more innovative and original than they would otherwise seem. The pressure is particularly strong in disciplines such as marketing and management, which place a premium on innovative and original work.

Another factor is that people often overlook relevant research published in regions and languages outside their normal frames of reference. The absence of non-English-language citations in leading marketing and management journals is noticeable even from a casual inspection of reference lists. It is rare to see citations to papers that are written in Mandarin, Arabic, French, German, Spanish, and so forth. However, such papers exist. As with English-language papers, some are poor, others are good. Some deserve to be ignored, others do not. Through field work in China, I have become aware of hundreds of Mandarin-language papers in my discipline; yet, English-language journals cite little of this work. In response to issues such as these, the major European-based marketing journal, the *International Journal of Research in Marketing*, asked referees to use polycentric citation as an assessment criterion—the intent was to encourage citation of papers published in the rich and varied languages of Europe. Unfortunately, this inspired initiative ceased several years ago when the journal's editorship changed.

Relevant research, it appears, is being overlooked. To this extent, I sympathize with the views that Wright and Armstrong express. However, they are being idealistic in calling for comprehensive citation because the sheer volume of work published across disciplines and geographies is so vast and disparate. For example, whereas it is desirable for marketers working in the area of diffusion to be aware of the research of epidemiologists and geographical scientists around the world, it is unreasonable to expect them to have a comprehensive knowledge of all this work. It is difficult enough for marketers to keep abreast of work in their own discipline. The issue, therefore, is not simply the importance of referring to relevant research, but of referring to *the most important and valid research evidence, regardless of disciplinary or geographic boundaries*. In view of improved access to journals through online databases, this more limited and circumscribed goal seems realistic.

Redundant Citations

An insidious problem, which Wright and Armstrong do not address, is that of spurious and redundant citations. When I look at recent journal articles in my discipline, it is the great abundance of references that is striking—not their absence. This is understandable if papers are positioned as reviews, syntheses, or meta-analyses; however, references are as abundant in empirical papers. Compare this to earlier decades. Famously, Einstein in his path-breaking, 1905 paper on the Special Theory of Relativity merely thanked Michelangelo Besso, a colleague in the Swiss Patent Office, but provided no additional footnotes or references (Clark 1979, p. 96). Even at the time of publication, there was work that he could, and perhaps should, have cited. Today, the pendulum has swung decisively in the opposite direction.

Token referencing is commonplace. Authors seem compelled to cite classics—works by people such as Drucker and Levitt in marketing or Luce and Tversky in behavioral decision theory. I wonder how often people have actually read these works, let alone assessed the validity of the research evidence (or lack of evidence) that these works contain. When bibliographic and quotation errors are perpetuated, it is easy to see that the cited papers have not been read and that people are merely copying the errors of previous citing authors. Unnecessary citations can arise when authors are unsure of their ground and need to bolster their viewpoint. Perhaps, the sample size of a survey is dubiously small, but the decision to proceed might appear to be justifiable if the author can

cite five publications that have equally small sample sizes. Evidence from all these papers might be suspect, but the weight of citation is hard for referees and editors to dismiss easily. Similarly, authors sometimes include references to satisfy referees, rather than because of any belief in the merits of the cited work. I must confess to doing this myself. The alternative would have been to receive a negative referee's report and have my paper rejected merely because I stubbornly refused to add a few irrelevant references.

The sheer density of referencing associated with this pendulum swing is out of control. Even readers of empirical papers must penetrate forests of authors and dates as they try to appreciate the main ideas and results. This can be daunting and off-putting, especially for general readers. And it is distracting. Authors and readers alike find themselves devoting mental effort to redundant citations, instead of concentrating their efforts on the really important and most valid research evidence. Ironically, the same online databases that give us such easy access to relevant papers are also feeding the process of redundant citation. Authors are using Google searching to generate reference lists rather than basing their lists on judicious, scholarly reading.

Recommendations

Cite better. Ensure that you cite the most important and valid research evidence, regardless of disciplinary or geographic origins. Cite less. Cut redundant and irrelevant citations. Freed from the distractions of these citations, authors can concentrate on being more accurate in the way they use the smaller numbers of references. This means actually reading the papers, assessing the validity of the evidence, having an awareness of caveats and contingencies, and being open about differences of interpretation. If there is a genuine need for a few supplementary references, dispatch these to footnotes, endnotes, or appendices so they do not become a distraction for time-pressured readers.

Reference

Clark, R. W. 1979. *Einstein: The Life and Times*, revised British ed. Hodder & Stoughton, London, UK.

Comment: Citation Shortcomings: Peccadilloes or Plagiarism?

Brian Martin

School of Social Sciences, Media and Communication, University of Wollongong, New South Wales, Australia, bmartin@uow.edu.au

Citation shortcomings seldom become a burning issue among scholars. One possible reason is the difficulty of studying the problem. While it is relatively straightforward to assess whether the words and letters in citations are correct, it is far more difficult to determine whether authors have correctly chosen and used sources.

Wright and Armstrong show one way to do this: examine the presence or absence of citations to a particular source known to be essential—in their case, Armstrong and Overton (1977)—and, when it is present, assess whether it has been correctly interpreted and implemented. This exemplary study reveals near-universal neglect or misuse of a relevant source.

Another approach is to know the sources in a field comprehensively and to assess all the citations of papers in the field to look for both omissions and inappropriate inclusions. MacRoberts and MacRoberts (1989) used this method and reported a substantial level of citation bias and inaccuracy. They concluded that citations capture only a small proportion of the influence on a scientific paper: many sources that influence a paper are not cited.

Before assessing the significance of these findings, it is worth outlining some purposes of using citations: (1) to support an argument; (2) to indicate to readers the most important and useful studies in the field; (3) to acknowledge sources of ideas, methods, or

quotations; and (4) to impress readers, referees, and editors. Ideally, the fourth purpose would be superfluous because citations made for the first three purposes would be sufficiently impressive.

For each of these purposes, there are conventions and violations of these conventions. For example, research students are often told to acknowledge sources of ideas. However, they soon learn that convention dictates that they omit some types of sources, such as media stories and informal conversations. Ravetz (1971) recommends giving detailed acknowledgements, for example, thanking the person who recommended a particular source. This level of precision in acknowledgement is rare. Most who contribute ideas informally are lucky to receive a mention in a list of people thanked.

Wright and Armstrong argue that an author should read all sources cited. This is reasonable when papers are short. However, is it necessary to read an entire book, or just enough of it to get the basic idea or to check a quotation and its context? This also applies to an article: is it necessary to read it thoroughly and carefully, or is reading the abstract and conclusion sufficient? Much depends on the purpose of the citation. If the source is methodologically central, then careful reading is essential. However, if the purpose is to indicate the place of the source in a survey of the field, then a more superficial understanding may suffice. How essential is it that, in citing A&O in this comment, I carefully read it? (This question is doubly rhetorical because I added this citation to highlight this point!)

Wright and Armstrong are concerned primarily with faulty citation as a factor in poor research. There are other important issues. One is gift citation—an author cites a supervisor, patron, editor, potential referee, or other person in an attempt to curry favor. Another is rivalry citation omission—an author does not cite obvious sources to prevent a rival or enemy from gaining proper credit. I have heard of a number of such cases. However, care must be taken in making a judgment: it is easy to attribute malice when ignorance is a possible explanation.

When an author cites an unsighted source—namely, a source the author has not looked up, seen, or read—the proper practice is to reference the secondary source. For example, I might refer to Eichorn and Yankauer (1987), as cited in Wright and Armstrong (2007). However, if I cite Eichorn and Yankauer without obtaining or reading the paper, instead just copying the citation from Wright and Armstrong without acknowledging them as a source, this would be a type of plagiarism—plagiarism of secondary sources. Given the great number of citation errors in the literature, this seems to be common; nonetheless, it is seldom given the stigmatizing label "plagiarism."

Why are citation shortcomings not taken more seriously? One explanation is that fraud in scholarship is defined in a narrow fashion, with only extensive plagiarism, manufactured data, or alteration of results deemed to be fraud. By this definition, only a few cases of fraud come to light; by denouncing them, the rest of the scholarly community is absolved.

A broader definition of fraud would include exaggerations in grant applications, exploitation of subordinates, omissions and misleading claims in curricula vitae, sloppy scholarship—and plagiarism of secondary sources. According to this broader definition, many scholars are guilty, including some who have risen to high ranks by exploiting the work of others. Arguably, a narrow definition of fraud helps to maintain the hierarchy within research institutions (Martin 1992).

Promoting better citation practices would raise the quality of scholarship. It would also increase the amount of work required to make a scholarly contribution, thereby preventing many researchers from publishing as much, and some from publishing at all. Therefore, the institutionalized pressure to publish for career reasons is one of the biggest obstacles to improved citation practices.

Acknowledgments

I thank Juan Miguel Campanario, John Lesko, Michael MacRoberts, and Malcolm Wright for their useful comments.

References

Armstrong, J. S., T. S. Overton. 1977. Estimating nonresponse bias in mail surveys. *J. Marketing Res.* **14**(3) 396–402.

MacRoberts, M. H., B. R. MacRoberts. 1989. Problems of citation analysis: A critical review. *J. Amer. Soc. Inform. Sci.* **40**(5) 342–349.

Martin, B. 1992. Scientific fraud and the power structure of science. *Prometheus* **10**(1) 83–98.

Ravetz, J. R. 1971. *Scientific Knowledge and Its Social Problems*. Clarendon Press, Oxford, UK.

Interfaces 38(2), pp. 125–139, © 2008 INFORMS

Citation House Rules: Reply to Commentators

J. Scott Armstrong

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, armstrong@wharton.upenn.edu

Malcolm Wright

Ehrenberg-Bass Institute, University of South Australia, Adelaide, South Australia, malcolm.wright@marketingscience.info

How serious are citation errors in research? If they are serious, what actions should we take? The three commentators have confirmed widespread problems. Don Dillman tells a tale of woe about citations of his seminal work, *Mail and Telephone Surveys: The Total Design Method* (Dillman 1978). Mark Uncles comments on the widespread failure to consider relevant research published in related fields or other languages. Brian Martin extends our arguments to consider issues such as gift citations and plagiarism of secondary sources. They all acknowledge the need for better citation practice. Yet Dillman also suggests that the process of contacting authors of primary research would be onerous for the cited authors. Uncles argues for fewer and better citations. Martin asks whether it is reasonable to expect citing authors to read all citations in full.

We do not think it is onerous to check with cited authors. Don Dillman is speaking as an author who has been cited thousands of times. But, in general, few authors are highly cited. The Lippincott Library at the University of Pennsylvania did an unpublished analysis of the citation rates for 31 full professors at Wharton, all of whom were employed at Wharton from 1975 through 1988. During this period, the median first-author citation rate per year was 11. Furthermore, it is only necessary to check when a citation is used to support a finding; in our judgment, this means that it is unnecessary to check the majority of citations. Nonetheless, we appreciate it when people contact us about citations of our papers, and wish that it would happen more often. It is rewarding to know that people rely on our research results. It is part of our role as researchers and educators to support their efforts, and to ensure that we are quoted correctly. Currently, the problem is not what would happen if authors checked every citation; it is to get authors to do *any* verifications.

Mark Uncles suggests that authors should cite better and cite less; rather than seek comprehensive literature reviews, they should cite the "most important and valid research evidence." This is useful advice. However, there are no clear-cut rules for truncation of a citation search. Researchers face problems of *time, availability, influence*, and *relevance* when deciding to truncate a search. Sometimes it may be important to go back, e.g., as far as 1950, to cite unpublished working papers, to record ideas from obscure and little-read journals, or to delve into the geography, psychology, or health economics literature. Journal websites for overlooked citations would allow overly truncated citation searches to be supplemented, and citations from other disciplines to be added. In addition, if authors read what they cite, they will become aware of literature that previous contributors considered important. These authors will be less likely to use unnecessary citations because reading such papers might reveal their irrelevance. Therefore, elimination of unread citations should result in substantial progress toward meeting Mark Uncles' goals. We also urge authors to inform readers of the relevant evidence provided in each citation. If there is no evidence, authors should explain the purpose of the citation. When the first author of this paper does reviews, he strikes out references that do not meet these criteria. This means that he typically strikes out most references.

It is incumbent on citing authors to ensure the accuracy of their citations. Brian Martin is right to emphasize that authors should not be expected to read the entire book that they cite; however, they should at least read the pages that they cite. We could apply this principle to journals by citing only the pages that contain the relevant points. More generally, Brian Martin's rule of avoiding plagiarism of secondary sources is a useful guideline in support of citation accuracy.

But could it be that citation problems are simply a feature of all scientific endeavors? In our paper, we cite a number of studies that show similar problems in other disciplines. Therefore, we might argue that any publication process will exhibit such errors. We believe the situation that we describe is worse. Consider our study: Armstrong and Overton (1977) is highly cited and its citation rate is increasing. It makes specific methodological recommendations, with most citations claiming to apply these recommendations; however, virtually all of the citations are quotation errors.

Based on our paper and on the commentary, we suggest that authors follow these rules:

(1) Cite all papers that you rely upon for evidence (or for direct quotations).

(2) Inform the reader of the content of each citation.

(3) If only part of the cited work has been read, include the applicable page numbers in the citation.

(4) Avoid citing any other papers. In particular, it is unnecessary to cite ideas unless they are part of a direct quotation. Rather, it is detrimental to do so because readers are likely to assume that the citation provides evidence. Moreover, it is difficult to trace ideas back to their origin; Stigler's Law of Eponymy states that, "no scientific discovery is named after its original discoverer."

(5) If you rely on the findings of an author, attempt to contact that author to confirm the citation accuracy.

Surely it is more efficient for editors to take steps to ensure accuracy than to expect readers to check original sources. Leading newspapers and magazines routinely check that sources have been properly quoted. We believe that scientific publications should have procedures for ensuring accurate quotations that are at least as good as those used by the popular press. The simplest and most direct way to do this is to ask lead authors to sign statements that they (or, where relevant, coauthors) have read every source cited.

References

Armstrong, J. S., T. S. Overton. 1977. Estimating nonresponse bias in mail surveys. *J. Marketing Res.* **14**(3) 396–402.

Dillman, D. A. 1978. *Mail and Telephone Surveys: The Total Design Method*. John Wiley & Sons, New York.