

MPRA

Munich Personal RePEc Archive

Verification of Citations: Fawlty Towers of Knowledge?

Malcolm Wright and J. Scott Armstrong

Ehrenberg-Bass Institute, University of South Australia

July 2007

Online at <http://mpra.ub.uni-muenchen.de/4149/>
MPRA Paper No. 4149, posted 19. July 2007

The Ombudsman: Verification of Citations: Faulty 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, as an exploratory study 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 1988); 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 towards 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 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 on 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 3 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, Nadjari, and Burchell (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, Cameron, and Gibbons 2004), and 34 percent in biomedical informatics journals (Aronsky, Ransom, and Robinson 2005). Schulmeister (1998) includes a summary of earlier literature in this area. This problem is serious even for the most prestigious journals. Lok, Chan, and Martinson

(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, Record, and Wade (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 are unrelated to the quoting authors' contentions. The detailed analysis that Evans, Nadjari, and Burchell (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, Cameron, and Gibbons (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 is 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 is 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 adjustment 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 (1.24 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 (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 (6 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 of 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 non-significant 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: 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 ($.54 * 136 / 16 = 4.6$) 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 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.

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 non-response, 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 non-response 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. In fact, we expect that most references in papers are spurious.

The Internet offers a solution to problems of omission. Journals should open websites (free to non-subscribers) 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 attempted to contact the authors prior to submission of their manuscript to the journal (where applicable). 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 many researchers refuse to respond when asked if their research is properly cited; a few have even written back to say that they do not plan to respond. In general, however, most respond with useful suggestions and are grateful that we have taken the care to ensure that we have cited them properly.

We attempted to contact via email 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. However, authors dealing with a large number of citations may prefer to contact a selection rather than a census of their references, or to restrict contact attempts to the most relevant studies.

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 well 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.” 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 1 out of every 80 academic mail surveys. Thus, we estimate that only 1 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 towards 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 “99⁴⁴/₁₀₀% 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 is more than “99⁴⁴/₁₀₀% 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 verified 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. 2007. Significance tests harm progress in forecasting. *Inter. J. Forecasting* **23** (2007) 321-327.
- Armstrong, J. S. 1996. Management folklore and management science: On portfolio planning, escalation bias, and such. *Interfaces* **26**(4) 28–42.
- 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 Association* **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. L. Schulmeister. Quotation and reference accuracy of three nursing journals. *J. Nursing Scholarship* **30**(2) 143 –147.
- Dillman, D. A. 2000, *Mail and Internet Surveys: The Tailored Design Method*, 2nd ed. New York, John Wiley & Sons.
- 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) 1011–12.
- 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 Association.* **263**(10) 1353–54.
- 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. Vucenic, 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. *Social Stud. Sci.* **16** 151–172.
- MacRoberts, M. H., B. R. MacRoberts . 1988. Problems of citation analysis: A critical review. *J. Amer. Soc. Inform. Sci.* **40**(5) 342–249.
- Parsons, H. M. 1974. What happened at Hawthorne?" *Sci.* **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. No. MR000002, DOI: 10.1002/14651858 MR000002.

