

TO CITE OR NOT TO CITE? USE AND ABUSE OF CITATIONS

Robert West and Kerstin Stenius

INTRODUCTION

Research output in the form of papers, books, book chapters and the like is there to be used. Authors need to consider how to make appropriate use of their own past work and the work of others and to ensure that their own work will be used appropriately.

Research output can be used in a number of ways. It can:

- 1 inform subsequent research,
- 2 influence policy decisions,
- 3 influence practitioner behaviour.

A research paper, book, policy document or treatment manual should refer to other writing that is relevant to its message. Citation is the formal vehicle for doing this. It involves explicit reference to a piece of research output which can in principle be anything from an article in a journal to a website. Conventions applying to citation practice regulate the transmission of information. Citation conventions vary from one research field to another; the following text focuses primarily on what might be termed 'cumulative research' in which the goal is to accumulate enduring knowledge and understanding.

Two main types of citation can be identified (see Box 4.1). In this chapter we use the term *referential citation* to refer to the situation where a piece of research output (which may be empirical or conceptual) is being used for what it contributes to the field. The term *critical citation* is used when the citing piece points to what is considered a flaw in some research output.

The citation serves one or more essential functions: it enables the reader to examine the cited work to check the veracity of a statement that it is being used to support, or the correctness of the use of a concept or interpretation of a process. When citing in support of a statement being made in one's own article, it also acknowledges the contribution made by the cited work. Both the *verification function* and the *acknowledgement function* are important. One may also use citations to document how a political debate or historical process or a specific concept has developed and been defined. We can call this the *documentation function*¹.

¹We are grateful to Klaus Mäkelä for this insight

As regards the verification function and the documentation function, the scope for intentional and unintentional distortion of research through making unfounded assertions or misleading statements is enormous. In principle every non-obvious factual claim should be supported in some way: either by direct evidence or by tracing a link through citations and/or inference to that evidence. Similarly every hypothesis, conceptual analysis or statement of a theoretical position that is not advanced for the first time in a given paper should trace the link to its source. Citation offers the reader an opportunity to determine for himself or herself a) whether the original source of a claim was justified and b) whether that claim is being accurately represented in the current piece.

Box 4.1 TYPES OF CITATION AND CITATION FUNCTION

Functions of citations

Verification function: the reader should be able to check the source for its accuracy and the accuracy with which it is reported

Acknowledgement function: the source is given credit for its contribution

Documentation function: the source is identified as the object of the research in its own right

Types of citations

Critical citation: a piece of work is cited because it is believed to mislead the field

Referential citation: a piece of work is cited for what it contributes to the field

As regards the acknowledgement function, it is right and proper that a researcher should receive credit for his or her work and citation is the primary means by which this is achieved. This is not merely a matter of etiquette; employment and promotion of individual researchers are built on reputation and citations play a crucial role in this. The institutions that employ researchers achieve kudos and in many cases funding on the basis of the reputations of their employees. Moreover, charities and government bodies that fund research must receive credit for the work they support. Their own income may well depend on it.

DEVIATIONS FROM IDEAL CITATION PRACTICE

Citation practice often falls far short of the ideal (for a discussion see Reyes 2001). For good practice in the use of citations in systematic reviews there are a number of sources one may use (e.g., Chalmers, Enkin et al. 1993; Cook, Sackett et al. 1995; Bannigan, Droogan et al. 1997; Moher, Cook et al. 1999; Sutton, Jones et al. 1999; Reeves, Koppel et al. 2002). Use of citations in less formal reviews, such as to introduce research reports, is subject to greater variability. The following paragraphs examine common deviations from ideal practice.

SELECTIVE CITATION THROUGH NEED FOR CONCISENESS

A legitimate reason to depart from ideal practice arises from the need for conciseness. Many times in a given field, a large number of studies may be cited in support of a given statement. In the absence of other constraints, the acknowledgement function might dictate that all relevant studies are cited. However, this would be impracticable. This raises the question of which article or articles to cite. There is a case for citing what we might call the *discovery piece*: the first article to record the finding. However, this may be impossible to determine. Moreover, it may not represent the most robust support for the assertion in question. There is a case for citing a review article. This has the advantage of pointing the reader, at least indirectly, to a body of work rather than one or two studies that might be unrepresentative. The disadvantages are a) the increased danger of misrepresentation because of hearsay and b) it does not acknowledge the contribution of the original source.

A possible rule of thumb in determining policy relating to a specific finding is to aim to cite the discovery piece and no more than five other original sources that testify to the generality of the finding, unless there is an authoritative and uncontentious review which can be cited instead. When referring to a conceptual or theoretical exposition, the first major presentation of the current version should be used.

SELECTIVE CITATION IN SUPPORT OF A VIEWPOINT

A common bias in reporting the literature is to select only (or primarily) studies that support a given hypothesis or idea (*viewpoint citation*). This is harder to avoid and to detect than one might imagine. If there were a well defined body of literature that examined a particular hypothesis and numerous high-quality studies conflicting with the hypothesis were ignored in a review, that would amount in the eyes of some to scientific misconduct; a reader who was not familiar with the area would be misled as much as if the author had fabricated data.

Less straightforward is the case where there are doubts about the methodological adequacy of conflicting studies. For example, studies that fail to detect the effect of an intervention may be small or involve inadequate implementation of the intervention.

Unless one is explicitly attempting a comprehensive review where there is the space to explore these issues, the citing author has to make a judgement about how far to go in ignoring weak studies. Given the realistic possibility that the citing author is not wholly disinterested in the matter, it is good practice to alert the reader to conflicting findings and to make a brief comment about the weight that might be attached to these and why.

Even less straightforward is the case where it is extremely difficult to determine what the corpus of findings on the topic is. This can happen for findings which typically do not form the main focus of papers. In the smoking literature, for example, it has been noted, and is widely believed, that depressed smokers are less likely to succeed in attempts to stop than non-depressed smokers. There are certainly studies showing such an association (Glassman, Helzer et al. 1990; Covey 1999). However, often buried in reports of clinical trials and other studies are numerous reports of failures to find such an association and indeed a recent meta-analysis has reported no association (Hitsman, Borrelli et al. 2003); no doubt there are even more instances where the association has been looked for, not found and no report has been made. At the very least, scientific prudence dictates that findings which are susceptible to this kind of phenomenon be cited with suitable caveats.

SELECTIVE CITATION TO ENHANCE REPUTATION

Self-citation or citation of colleagues with a view to enhancing one's own or the colleague's reputation (*reputation citation*) is clearly unacceptable. It distorts science and the process of science and is personally damaging to individuals in less powerful positions or those who do not engage in that practice. One may debate how widespread this practice is but there can be little doubt that self-serving bias runs at some level throughout the scientific literature.

Self-citation can also apply to journals (articles in journals tending to cite articles from the same journal). This may arise for reasons other than reputation citation, some of which may be legitimate, but it can distort the literature. One study found significant difference in self-citation rates among journals of anaesthesiology (Fassoulaki, Paraskeva et al. 2000).

It may be thought that a bias of this kind would be easily detected and an appropriate correction could be applied. However, this is probably optimistic. First of all, it is not unreasonable that one's own name should feature prominently in a reference list given that one's research is presumably to some degree programmatic. A similar principle would hold true for one's close colleagues. It can be difficult therefore to tell when this bias is operating. Secondly, the agencies that count citations (such as the Institute for Scientific Information, see later in the chapter) would only exclude a citation from an index if the same name appeared as first author.

SELECTIVE CITATION FOR CONVENIENCE

Using citations that are easy to find or happen to have come to the attention of the author is not good practice but is probably very common. There may be many ways in which *convenience citation* can distort the literature. Insofar as more accessible articles may not represent the literature it would create a biased impression. Searchable electronic databases could in principle mitigate the problem but they can also lead to their own kind of distortion. It would be expected that they would favour English language articles in journals indexed by the main databases. One would also expect more recent articles to gain preference because of the way that electronic databases sort the results of searches. Convenience citation would also be expected to favour the more popular journals. One might argue that this is no bad thing because it would be the better articles that would in general find their way into these journals; however, this is not necessarily so.

SELECTIVE CITATION BY COUNTRY OF ORIGIN

It goes without saying that a tendency to cite articles simply because they are from one's own country of origin is not good practice. Many researchers are under the impression that this occurs, however. Naturally, the greatest suspicion falls on the USA as the main producer of research output and many non-US researchers can probably recount cases where a US author has cited predominantly or exclusively US references even when more appropriate ones from other countries exist. In fact this bias has been found among both UK and US researchers publishing in major medical journals (Campbell 1990; Grange 1999). Another study found North American journals to cite North American journals to a greater extent than did journals from other regions (Fassoulaki, Paraskeva et al. 2000).

CITING INACCESSIBLE SOURCES

It is quite common for authors to cite conference papers or their abstracts, submitted articles, in-house papers or unpublished reports (the so called *grey literature*). The problem, obviously, with this kind of citation is that it does not fulfil the verification function of citation. Therefore it is generally to be discouraged. There may be cases where it is the only option and important in fulfilling the acknowledgement or documentation role, but if this is not obvious the use should be justified. If that citation is more than a few years old the use becomes increasingly problematic. It is often reasonable to presume that if it is a paper or an abstract and the finding was robust it would have found its way into the peer reviewed literature.

It is becoming common to cite websites. This is reasonable but will pose increasing problems over time as websites move or become inaccessible. In general, for any statement intended to have lasting significance this practice is best avoided until a system is devised for ensuring the longevity of web-based scientific literature. However, in policy analyses or descriptions of historical processes for instance, references to sources such as websites and government documents may be a key part of the research process.

CITING UNEVALUATED SOURCES

When a citation is used to support a substantive statement the implication is that the cited reference reports evidence in support of that statement. Inadequate though it is, peer review is the primary gatekeeper for this kind of report. However, it is commonplace for statements of fact to be supported by citations to book chapters, letters, conference presentations, opinion pieces and other material that has not been peer reviewed. Although in principle the reader can track down the source and make his or her own evaluation, this is often impracticable. The only thing that comes close to a safeguard is the fact that the report has been through a peer review process. Within the social sciences, though, even non-peer reviewed books still remain a main source for new analytical concepts. In some cases, however, the review process for books is as rigorous as the peer review process for journal articles.

CITING WITHOUT READING

There is a great temptation to cite a piece of work on the strength of a report of what it says without going to the original source. Thus if a paper or book chapter that we have access to makes a statement which is relevant to our work and cites another article in support of it, it is tempting to repeat the assertion and the citation without reading the original source material. This is clearly unacceptable because of the risk of misrepresentation. Similarly, having identified an abstract of an article using an electronic database, there is a temptation to cite the article without going to the full text. This is risky practice because one has not taken the opportunity to evaluate the research being cited by reference to the methods and analyses used. As a general principle, authors should not make reference to research output without having read and evaluated that output directly.

OVER-USE OF CITATIONS

Much of the earlier discussion concerned selective use of citations. Quite a common problem is the reverse: providing a long list of citations to support a single statement when fewer would be sufficient. If it is important that the work of the authors of all the various works be acknowledged, or if the intention is to provide a comprehensive review, then a long list of citations is appropriate. Otherwise it can make a paper unwieldy and the rule of thumb of selective citation described earlier could be adopted.

GETTING CITED

All the above should suggest that the process of citation is subject to considerable bias and, while there is a duty on researchers to minimize this, it is unlikely that it will ever be eliminated. This being said, if one is writing an article which one believes is important it would seem reasonable to try to ensure that it is drawn to the attention of its intended audience - and that means being cited. The choice of journal is obviously of relevance (see Chapter 2). However, it may not be the most prestigious

journal that offers the best chance, but rather the best quality *specialist* journal. The most prestigious journals tend to be generalist and as such may not be routinely read by many potential users of the research. Whatever outlet one uses for one's research, it can often be a good idea to take other steps to publicise the findings. Some researchers email or send copies of their papers to colleagues. Others post them on listserv lists. Conference presentations and web-sites are also potentially useful sources of publicity.

CITATION INDEXES

We mentioned earlier that citation is often used as a marker of quality. There is a presumption that the more often an article is cited, in some sense the better it is. This extends to journals, for whom the single most widely used measure of quality is the 'Journal Impact Factor'. The Impact Factor for a journal in a given year is calculated as the average number of citations in that year to articles in the preceding two years. Thus, supposing a journal published 50 articles in 2001 and 2002 and there were 100 citations to these articles in 2003, the journal's Impact Factor for 2003 would be 2.0. Articles from the same journal are included but citations from articles with the same first author as the cited article are excluded.

Researchers are often judged by the citations counts of their articles and by the Impact Factors of the journals in which they publish. Funding decisions in many institutions and departments are based in part on members of those institutions publishing in 'high impact' journals. Unfortunately there are many problems associated with using citation counts as a marker of quality and even more with using Impact Factor (Ophthof 1997; Seglen 1997; Hecht, Hecht et al. 1998; Jones 1999; Semenzato and Agostini 2000). Some researchers have suggested that it may be possible to use them with appropriate caveats and corrections (Rostami-Hodjegan and Tucker 2001; Fassoulaki, Papilas et al. 2002; Braun 2003) while others have argued that they should be abandoned (Bloch and Walter 2001; Ojasoo, Maisonneuve et al. 2002; Walter, Bloch et al. 2003).

As regards citation counts, the various biases in the use of citations discussed earlier should give an indication of the problem with using counts as a marker of quality. In addition, it should be recalled that *critical citation* is quite commonplace; therefore an article might be cited precisely because it is weak or misleading. A recent article examined the association between peer ratings of quality and the numbers of citations between 1997 and 2000 to articles appearing in the journal *Addiction* in 1997 (West and McIlwaine 2002). Although two independent reviewers agreed moderately in their ratings of the papers, the correlation between these ratings and the number of citations was almost zero. One factor that was correlated with citation count was region of origin of the first author of the paper: papers from English speaking countries received more citations than those from continental Europe, which received more than those from the rest of the world. A larger analysis of citations to articles in emergency medicine revealed that citation count of articles was predicted to some extent by the impact factor of the journal in which they appeared and to a more limited extent by quality of the articles (Callaham, Wears et al. 2002). A further study

of citations to papers reporting randomized trials in hepato-biliary disease found a significant association with a positive outcome, but no association with adjudged quality (Kjærgard and Gluud 2002).

Apart from the biases already discussed, the fact that only a small proportion of predominantly US journals is indexed in the Institute of Scientific Information (ISI) databases would lead to a bias, particularly against non-English speaking countries. One study reported that exclusion of core journals in emergency medicine had led citation counts in the field to remain low despite considerable expansion of the field (Gallagher and Barnaby 1998). Another noted that the way to improve the Impact Factors of journals in dermatology was to increase the number of them indexed by the ISI (Jemec 2001). Another bias arises from the fact that some fields, such as biosciences use a lot more citations than others. This will disadvantage authors in low citing fields, typically the social sciences. There is a range of other factors that make citation counts potentially misleading as a marker of quality (Box 4.2).

Box 4.2 WHY CITATION COUNTS ARE OFTEN MISLEADING AS A MARKER OF QUALITY

- Articles are sometimes cited in order to be criticised
- Papers describing important original studies are often neglected in favour of reviews
- There is a bias towards citing articles from one's own country or research group, or that are easily accessible
- Some fields of study generate more citations than others irrespective of how important they are: e.g., fields with high levels of activity and mature fields
- The importance and quality of a piece of work may relate to its policy or clinical implications rather than its use by other researchers
- Other researchers may fail to grasp the importance of a piece of work
- The citation indexes are biased towards English-speaking and US journals

Additional problems arise from the two-year time window used by the Impact Factor score. The publication lag in many journals, particularly social science journals, is at least a year. This means that for an article to be read after it has been published and then used to inform new research would normally take at least two years. Citations made within the Impact Factor window may therefore not have been used to generate new research ideas but more likely be part of a programme involving a community of authors with reciprocal knowledge of each others' research.

CONCLUSIONS

Citations are the primary formal means by which scientific findings are communicated. Ideal citation practice would involve comprehensive and objective use of the whole corpus of published literature. Clearly this is impossible. However, it should still be possible to approximate this ideal by adopting a few guidelines. These recognize that citation serves the dual function of enabling verification of statements and acknowledging contributions.

In the case of formal reviews, the principles are well-documented: the sources searched and the search rules should be clearly specified, as should the inclusion and exclusion criteria for articles; the sources should go beyond ISI databases and include searching reference lists of papers in the search domain. With regard to informal reviews, such as are used to introduce research reports, the principles in Box 4.3 can be applied:

Box 4.3 PRINCIPLES TO BE FOLLOWED WHEN COMPILING AN INFORMAL REVIEW TO INTRODUCE A RESEARCH REPORT

- All non-obvious, substantive claims should be supported by citation or direct evidence
- If there is an authoritative review on a well-supported statement, this may be used in place of original papers
- Where original papers are cited, the discovery paper should be cited together with a small number of other papers that illustrate the generality of the phenomenon
- Authors should resist the propensity to:
 - prefer citations from their own country of origin unless the finding in question is country-specific*
 - prefer citations from themselves and colleagues*
 - limit citations to those that support a contention, when in fact there are others that conflict with it*
 - cite output which is readily retrievable if there are more appropriate references*
 - provide an unnecessarily large number of citations for a single statement*
- Authors should where possible avoid citing inaccessible sources
- When using citations in support of substantive statements one should either use references that have been through some kind of peer review process or provide an appropriate caveat.

Citation counts are widely used as an index of quality. Given that few if any researchers are able to follow all the above principles, together with the fact that many other factors influence the number of times a piece is cited, citation counts are a highly problematic index of quality. Journal Impact Factors are even more problematic. Authors should be aware of this and not be beguiled by their apparent objectivity. Ultimately there is no substitute for peer evaluation of research output, however flawed and subjective this might be.

REFERENCES

- Bannigan, K., Droogan, J. et al. (1997) Systematic reviews: what do they involve? *Nursing Times*, 93(18), 52-53.
- Bloch, S. and Walter, G. (2001) The Impact Factor: time for change. *Australian and New Zealand Journal of Psychiatry*, 35(5), 563-568.
- Braun, T. (2003) The reliability of total citation rankings. *Journal of Chemical Information and Computer Sciences*, 43(1), 45-46.
- Callahan, M., Wears, R. L. et al. (2002) Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Journal of the American Medical Association*, 287(21), 2847-2850.
- Campbell, F. M. (1990) National bias: a comparison of citation practices by health professionals. *Bulletin of the Medical Libraries Association*, 78(4), 376-382.
- Chalmers, I., Enkin, M. et al. (1993) Preparing and updating systematic reviews of randomized controlled trials of health care. *Milbank Quarterly*, 71(3), 411-437.
- Cook, D. J., Sackett, D. L. et al. (1995) Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *Journal of Clinical Epidemiology*, 48(1), 167-171.
- Covey, L. S. (1999) Tobacco cessation among patients with depression. *Primary Care*, 26(3) 691-706.
- Fassoulaki, A., Papilas, K. et al. (2002) Impact factor bias and proposed adjustments for its determination. *Acta Anaesthesiologica Scandinavica*, 46(7), 902-905.
- Fassoulaki, A., Paraskeva, A. et al. (2000) Self-citations in six anaesthesia journals and their significance in determining the impact factor. *British Journal of Anaesthesia*, 84(2), 266-269.
- Gallagher, E. J. and Barnaby, D. P. (1998) Evidence of methodologic bias in the derivation of the Science Citation Index impact factor. *Annals of Emergency Medicine*, 31(1), 83-86.
- Glassman, A. H., Helzer, J. E. et al. (1990) Smoking, smoking cessation, and major depression. *Journal of the American Medical Association*, 264(12), 1546-1549.
- Grange, R. I. (1999) National bias in citations in urology journals: parochialism or availability? *BJU International*, 84(6), 601-603.
- Hecht, F., Hecht, B. K. et al. (1998) The journal impact factor: a misnamed, misleading, misused measure. *Cancer Genetics and Cytogenetics*, 104(2), 77-81.
- Hitsman, B., Borrelli, B. et al. (2003) History of depression and smoking cessation outcome: a meta-analysis. *Journal of Consulting and Clinical Psychology*, 71(4), 657-663.
- Jemec, G. B. (2001) Impact factors of dermatological journals for 1991 - 2000. *BMC Dermatology*, 1(1), 7.
- Jones, A. W. (1999) The impact of Alcohol and Alcoholism among substance abuse journals. *Alcohol and Alcoholism*, 34(1), 25-34.
- Kjaergard, L. L. and Gluud, C. (2002) Citation bias of hepato-biliary randomized clinical trials. *Journal of Clinical Epidemiology*, 55(4), 407-410.
- Moher, D., Cook, D. J. et al. (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-Analyses. *Lancet*, 354(9193), 1896-1900.
- Ojasoo, T., H. Maisonneuve, et al. (2002) [The impact factor of medical journals, a bibliometric indicator to be handled with care]. *Presse Medicale*, 31(17), 775-781.
- Ophof, T. (1997) Sense and nonsense about the impact factor. *Cardiovascular Research*, 33(1), 1-7.
- Reeves, S., Koppel, I. et al. (2002) Twelve tips for undertaking a systematic review. *Medical Teacher*, 24(4), 358-363.
- Reyes, H. (2001) [The references in articles published in biomedical journals]. *Revista Medica de Chile* 129(4), 343-345.
- Rostami-Hodjegan, A. and Tucker, G. T. (2001) Journal impact factors: a 'bioequivalence' issue? *British Journal of Clinical Pharmacology*, 51(2), 111-117.
- Seglen, P. O. (1997) Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079), 498-502.
- Semenzato, G. and Agostini, C. (2000) The impact factor: deeds and misdeeds [editorial]. *Sarcoidosis Vasculitis and Diffuse Lung Diseases*, 17(1), 22-26.
- Sutton, A. J., Jones, D. R. et al. (1999) Systematic reviews and meta-analysis: a structured review of the methodological literature. *Journal of Health Services Research Policy*, 4(1), 49-55.
- Walter, G., Bloch, S. et al. (2003) Counting on citations: a flawed way to measure quality. *Medical Journal of Australia*, 178(6), 280-281.
- West, R. and McIlwaine A. (2002) What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality. *Addiction*, 97(5), 501-504.