

Statistical analyses of digital collections: using a large corpus of systematic reviews to study non-citations

Tove Faber Frandsen, t.faber@videncentret.sdu.dk

Odense University Hospital, Denmark

Jepe Nicolaisen, jep.nic@hum.ku.dk

Royal School of Library and Information Science, University of Copenhagen, Denmark

Libellarium, IX, 2 (2016): 81 – 94.

UDK: 025.12:004:364=111

DOI: <http://dx.doi.org/10.15291/libellarium.v9i2.253>

Research paper

Abstract

Using statistical methods to analyse digital material for patterns makes it possible to detect patterns in big data that we would otherwise not be able to detect. This paper seeks to exemplify this fact by statistically analysing a large corpus of references in systematic reviews. The aim of the analysis is to study the phenomenon of *non-citation*: Situations where just one (or some) document(s) are cited from a pool of otherwise equally citable documents. The study is based on more than 120,000 cited studies, and a total number of non-cited studies of more than 1.6 million. The number of cited studies is found to be much smaller than the number of non-cited. Also, the cited and non-cited studies are found to differ in age. Very recent studies tend to be non-cited whereas the cited studies are rarely of recent age (e.g. within the same year). The greatest differences are found within the first 10 years. After 10 years the cited and non-cited studies tend to be more similar in terms of age. Separating the data set into different sub-disciplines reveals that the sub-disciplines vary in terms of age of cited vs. non-cited references. Some fields may be expanding and the number of published studies is thus growing. Consequently, cited and non-cited studies tend to be younger. Other fields may be more slowly progressing fields that use a greater proportion of the older literature within the field. These field differences manifest themselves in the average age of references.

KEYWORDS: bibliometrics, digital collections, systematic reviews, citation analysis, non-citations

Introduction

The recent production of digital materials more than equals the total amount of information produced so far in the entire history of our species (e.g., Dienes 2012, Hilbert and López 2012, Kitchin 2014b). Clearly, this is a big challenge

to the traditional core competences of information professionals (curation, preservation, organization, and seeking). Yet, it also promises great opportunities. Using statistical methods to analyse digital material for patterns makes it possible to detect patterns in big data that we would otherwise not be able to detect (Kitchin 2014a).

This paper seeks to exemplify this great potential by statistically analyzing a large corpus of references in systematic reviews. The aim of the analysis is to study the phenomenon of non-citation: situations where just a few are cited from a pool of otherwise equally citable documents. Previously, this phenomenon has only been studied by a process of *close reading*. For example, MacRoberts and MacRoberts (1986) analysed 15 randomly selected papers from a single discipline and determined the influences manifested in them. They found that in many cases influence was not captured in references and footnotes. Thus, by close reading, they determined a group of cited and non-cited studies. Yet, their method limited them to just 15 papers, and they also had to admit that their process was quite subjective. Some form of *distant reading* (Moretti 2013) is needed in order to increase the amount of data, and to carry out a more objective evaluation process. Using systematic reviews, we believe to have found a way to increase the amount of data to be analysed as well as a more objective process of evaluation. The research question is thus two-fold: How can we use distant reading to study the non-citation phenomenon, and to what extent are we able to confirm the results of previous studies of non-citations based on close reading?

The next section provides a review of previous studies of the non-citation phenomenon and related citation phenomena. We then outline the new method in detail, and show how it works in practice by applying it to a study of non-citations in the field of healthcare.

Review

According to the so-called *normative citation theory* (Nicolaisen 2007), failure to give credit where credit is due is unusual. Cole & Cole (1972, p. 370), for example, claim that “sometimes [...] a crucial intellectual forebear to a paper is not cited. The omission is rarely due to direct malice on the part of the author but more often to oversight or lack of awareness [...]. We can assume that omitted citations to less influential work are random in nature [...]”. Garfield (1977, 7) agrees stating that “the vast majority of citations are accurate and the vast majority of papers do properly cite the earlier literature”. Garfield, however, admits that this assertion had not been empirically substantiated: “Unfortunately, there has never been a definitive study of this assertion” (Ibid).

The basic assumption of the normative theory of citing was not tested before the 1980s. MacRoberts and MacRoberts wrote a number of articles during the

1980s and 1990s in which they argued that citation analysis is an invalid tool for research evaluation (MacRoberts 1997, MacRoberts and MacRoberts 1984, 1986, 1987a, 1987b, 1988, 1989a, 1989b, 1996). In these articles they challenge the basic assumption of the normative theory of citing: that scientists cite their influences. In their 1986 paper MacRoberts and MacRoberts report the results of a test of this assumption. They had read and analysed fifteen randomly selected papers in the history of genetics, a subject with which they claimed to be familiar, and had found that from zero (paper had no references or footnotes) to 64 percent influence was captured in references and footnotes. After having reconstructed the bibliographies of the fifteen papers, they were able to estimate that the papers required some 719 references at a minimum to cover the influence evident in them, when in fact they contained only 216 - a coverage of thirty percent for the entire sample. In their 1996 paper MacRoberts and MacRoberts claim that this percentage typifies all fields with which they are familiar (botany, zoology, ethology, sociology, and psychology) and conclude that “if one wants to know what influence has gone into a particular bit of research, there is only one way to proceed: head for the lab bench, stick close to the scientist as he works and interacts with colleagues, examine his lab notebooks, pay close attention to what he reads, and consider carefully his cultural milieu” (MacRoberts and MacRoberts 1996, 442).

Terrence A. Brooks published a pair of papers in the mid-1980s, which also challenge the basic assumption of the normative citation theory (Brooks 1985, 1986). These papers report the results of a survey covering 26 authors at the University of Iowa. Brooks had asked the authors to indicate their motivations for giving each reference in their recently published articles by rating seven motives for citing. One of the motives was labelled “persuasiveness”. Brooks claims that the results of his survey show that authors cite for many reasons, giving credit being the least important one. Of the 900 references studied, Brooks found that about 70 percent were motivated by more reasons, and concluded: “No longer can we naively assume that authors cite only noteworthy pieces in a positive manner. Authors are revealed to be advocates of their own points of view who utilize previous literature in a calculated attempt to self-justify” (Brooks 1985, 228). However, as pointed out by White (2004), the results of Brooks’ survey need to be assessed with some caution. This is because respondents almost certainly misunderstood the motive reading “persuasiveness” to denote “citing to help build a case”, and not “citing to utilize previous literature in a calculated attempt to self-justify”.

Zuckerman (1987, 334) asked if persuasion really were the major motivation to cite, would citation distributions look as they do? She answered “plainly not”, referring to data from an article by Eugene Garfield. Garfield (1985, 406) presents a table illustrating the number of citations retrieved by items cited one or more times in the 1975-1979 cumulated Science Citation Index. The table reveals, among other things, that only 6.3 percent of the 10.6 million citations went to documents cited 10 or more times in the five-year period. Zuckerman (1987)

points to the low percentage as evidence against the persuasion hypothesis. According to the persuasion hypothesis, the percentage should be much higher. Zuckerman (1987, 334) refers to Gilbert (1977, 113), one of the “inventors” of the persuasion hypothesis, who states that it is the papers seen as “important and correct” which “are selected because the author hopes that the referenced papers will be regarded as authoritative by the intended audience”. However, as Zuckerman (1987) notes, if one adopts a modest criterion of authoritative papers being equal to those, which have been cited at least ten times in five years (or twice annually), the persuasion hypothesis needs to be radically adjusted.

White (2004) realized that instead of testing the persuasion hypothesis like Zuckerman (1987) had done, by determining the percentage of citations received by authoritative papers, he could test the hypothesis, by determining the percentage of citations received by authoritative authors. Initially he had to determine how to measure the reputation of cited authors. This was done by counting the number of citations the cited authors had received. He then drew a judgment sample that consisted of 28 citing authors from different disciplines (ten from information science, eight from science studies, six from various natural sciences and four from cultural studies in the humanities). Finally, he tabulated the references provided by the 28 citing authors in their publications. The method enabled him to determine the frequencies by which reputable and non-reputable authors appeared in the bibliographies under study. His findings do not support the persuasion hypothesis. Most of the 28 authors cited at all levels over the entire scale of reputation, and they did not exclusively favour high-end names with authoritative reputations. If anything, White’s (2004) findings suggest that authors tend to favour low-end names slightly.

Moed and Garfield (2003, 192) asked “how does the relative frequency at which authors in a research field cite ‘authoritative’ documents in the reference lists in their papers vary with the number of references such papers contain”? They reasoned “if this proportion decreases as reference lists become shorter, it can be concluded that citing authoritative documents is less important than other types of citations, and is not a major motivation to cite” (Ibid). They went on to analyse the references cited in all source items denoted as ‘normal articles’ included in the 2001 edition of the Science Citation Index on CD-ROM. The source papers were arranged by research field, defined in terms of aggregates of journal categories. They limited their study to four fields: molecular biology & biochemistry, physics & astronomy, applied physics & chemistry, and engineering. The cited references were classified in two groups: those published in journals processed for the ISI citation indexes, and those published in non-ISI sources, including monographs, multi-authored books and proceedings volumes. In each research field the distribution of citations among cited items was compiled in each group separately, and the 90th percentile of that distribution was determined. Thus, the ten percent most frequently cited items published in ISI journals, and the ten percent most frequently cited documents in non-ISI sources were identified. These two sets were then combined. The combined set was assumed to represent

the documents perceived in the year 2001 as 'authoritative' in a research field. Source articles were arranged in classes on the basis of the number of references they contained. The percentage of references to 'authoritative' documents was finally calculated per class. The findings of their analysis clearly show that authors in all four fields cite proportionally fewer 'authoritative' documents as their bibliographies become shorter. In other words, when the authors display selective referencing behaviour, references to 'authoritative' documents are found to drop radically. Moed and Garfield (2003, 195) therefore concluded "In this sense, persuasion is not the major motivation to cite".

Method

From the review above it is evident that some types of citation behaviour have been studied statistically by analyzing large corpuses of references and citations. Yet, the specific citation behaviour known as *non-citation* has so far only been studied by close reading, which, as noted in the introduction, for various reasons limits the generalisability of the obtained results. We believe to have found both a way to increase the amount of data to be analysed, and a more objective process of evaluation, which make it possible to draw stronger conclusions regarding the non-citation phenomenon.

Using systematic reviews from the field of healthcare we are able to identify studies addressing the same research question. "Systematic reviews seek to collate all evidence that fits pre-specified eligibility criteria in order to address a specific research question" (Higgins and Green 2011). Importantly, systematic reviews aim to minimize bias by using explicit, systematic methods. Cochrane reviews are usually held among the best systematic reviews (Collier, Heilig, Schilling, Williams, and Dellavalle 2006, Moseley, Elkins, Herbert, Maher, and Sherrington 2009). Consequently, we decided to use Cochrane reviews for our analyses. Cochrane Reviews are prepared by authors who register titles with one of the 53 Cochrane Review Groups.¹ A Cochrane review contains a list of included studies. These studies address the same research question. The included studies of any given systematic review may therefore be examined to determine which of the preceding studies were cited in that specific study. To give an example:

Three studies (A, B, C) are included in a given Cochrane review (X); consequently we know that they are addressing the same research question.

1 Each Cochrane Review Group focuses on a specific topic area and is led by a Coordinating Editor(s) and an editorial team including a Managing Editor and Trials Search Coordinator. The Cochrane Review Groups provide authors with methodological and editorial support to prepare Cochrane Reviews, and manage the editorial process, including peer review (www.cochrane.org)

References in Cochrane review X	Publication year
A	2005
B	1999
C	1995

Thus, study A can cite study B and study C, whereas study B can only cite study C and study C is unable to cite any of the other two studies. By matching these three references to references in Web of Science, we are able to detect whether or not these studies are citing each other.

We retrieved 5,843 systematic reviews from 53 Cochrane groups (withdrawn reviews were excluded). Some did not contain any included studies resulting in 5,042 systematic reviews. In those reviews we were able to match the included studies to 60,495 references in Web of Science. These approximately 60,000 studies can cite the previous studies addressing the same question but obviously not the ones published after. We look at studies that can cite similar previous studies and this results in more than 1.5 million incidences of a given study citing or not citing a preceding study. This means that in our sample every study can on average potentially cite about 27 previous studies. We only included a pair of potentially citing and cited documents if the cited / non-cited document is from the same publication year or older. We included pairs of publications with the same publication year although we know that citing within the same year is rare. However, it does happen (due to preprint, early view etc.).

The data collected consists of the following information:

- Publication year of the citing or potentially citing study
- Publication year of the cited or potentially cited study
- Whether or not the study is actually being cited

We analysed the data focusing on the distribution of cited and non-cited publications, both in general as well as between sub-disciplines. The publication year of the cited or potentially cited study was in each case subtracted from the publication year of the citing or potentially citing study, such that we only look at the age difference between the citing and cited study.

Results

Table 1 provides an overview of how many cited and non-cited studies we were able to match by the age difference between the cited and citing studies. Since the age difference distribution has a very long tail, we limit the table to pairs with age differences of 25 years or less.

The total number of cited studies is just above 120,000 whereas the number of non-cited studies is more than 1.6 million. Consequently, the cited studies are

outnumbered by the non-cited by a factor 13 which is what we would expect knowing that a considerable part of all publications are never cited even within the health sciences (see e.g. Ranasinghe et al. 2015, Weale, Bailey, and Lear 2004). As we would expect pairs with a zero age difference 0 exhibit very few cited studies and many more non-cited studies, since only rarely will a study be able to cite another study published in the same year.

Table 1. Overview of the age difference between cited and non-cited studies (only years 0 to 25 are shown)

	Cited studies	Non-cited studies
Total	121605	1610450
0	3408	202996
1	12632	177770
2	16396	155454
3	15577	139417
4	13573	122807
5	11508	108011
6	9314	94698
7	7766	81204
8	6126	70804
9	5162	61167
10	4067	52463
11	3168	44741
12	2520	38170
13	2098	33488
14	1616	29838
15	1310	25513
16	1041	22751
17	780	19830
18	694	17035
19	510	14670
20	403	12846
21	364	11067
22	265	9682
23	295	8701
24	200	7542
25	146	6676

Figure 1 provides an overview of the cited as well as the non-cited studies by age. We can see that the uncited studies outweigh the cited. The non-cited studies are voluminous in the most recent years indicating publications are less likely to cite a relevant, similar study if that study is published recently. Again, publication lags may also play a role.

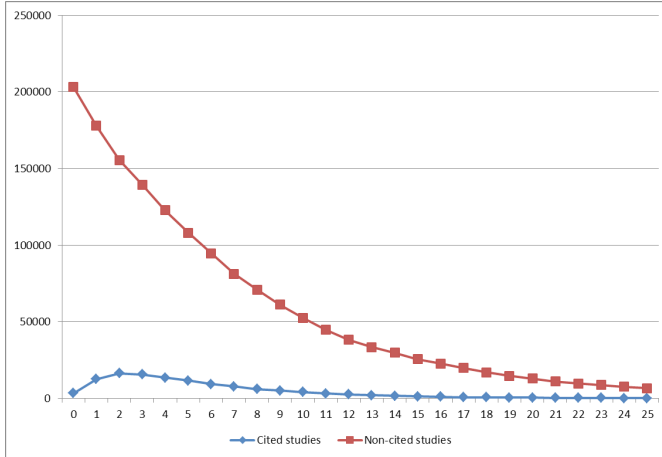


Figure 1. The number of cited and non-cited studies by age difference (only years 0-25 are shown)

Figure 2 provide us with an overview of the distribution of the percentages of cited and non-cited studies by age difference. As we can see, 13 per cent of non-cited studies are published the same year as the citing study. These results are consistent with what we would expect as we would expect the shares within the first year to be high. The cited studies are very few in year 0 which is also what we would expect. Already in year 1 and 2 the shares are considerably increased. By year 9 and 10 the cited and non-cited studies show similar shares and thus it is within the first 10 years of publication we find the greatest differences.

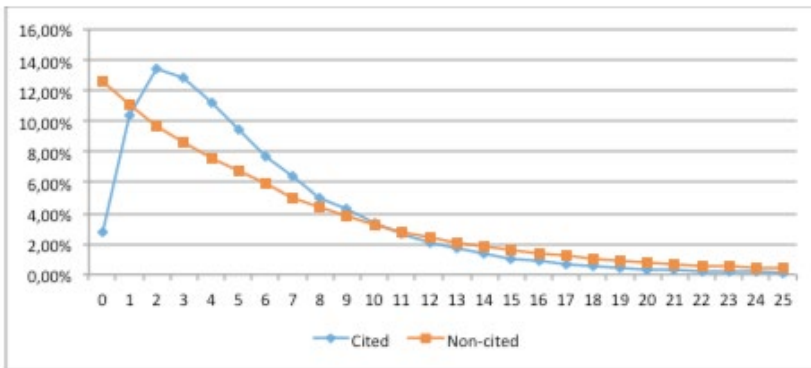


Figure 2. Distribution of the percentages of cited and non-cited studies by age difference (only 0-25 years are shown)

Figure 3 shows the distribution in cumulative distribution of the cited and non-cited studies by age difference. In the figure we can see that non-cited studies a year or less old makes up about 25 per cent of the non-cited studies whereas this is only a little over 10 per cent for the cited studies. In year 4 both cited and non-cited studies have reached 50 per cent.

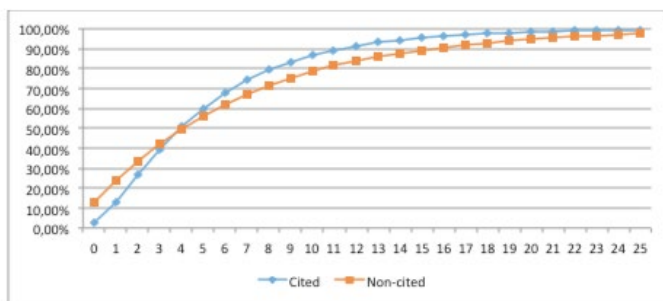


Figure 3. Cumulative distribution functions for cited and non-cited studies by age difference (only 0-25 years are shown)

The distributions depicted in figures 1-3 may not necessarily be valid for all disciplines or sub-areas. To examine this we separate the data into the 53 Cochrane groups from which the data originate. Due to space limitations we cannot show figures for all groups, but will present some examples.

The first example is the Cochrane Dementia and Cognitive Improvement Group². The majority of the cited studies are found in the literature dating back 5 years. For the non-cited studies the tendency is even clearer as the non-cited studies are plenty in the literature from the last 3 years. This implies that studies within this area tend to be relatively young which may be caused by an increase in production.

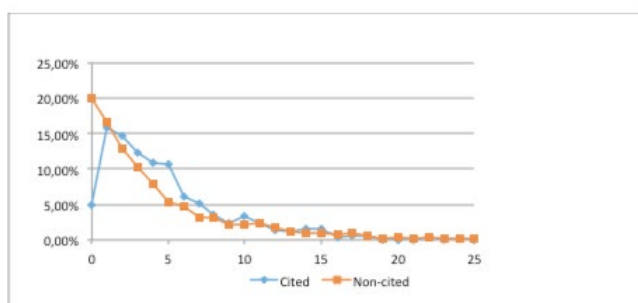


Figure 4. Distribution of the shares of cited and non-cited studies in the dementia group (only year 0-25 are shown)

2 The Cochrane Dementia and Cognitive Improvement Group is based at Oxford University in the Radcliffe Department of Medicine. The group was established in 1995 and now manages a portfolio of over 150 intervention reviews and around 20 reviews focused on diagnostic test accuracy (<http://dementia.cochrane.org/background>).

The next example is the Methodology Review group³. Figure 5 indicates that studies within this area are slightly fewer during the first year whereas the number of studies that are 2-4 years old are greater than in the case of the dementia group.

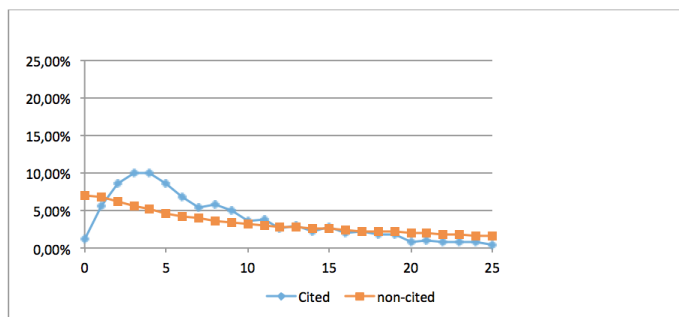


Figure 5. Distribution of the shares of cited and non-cited studies in the methods group (only year 0-25 are shown)

Based on these figures, the methods group seems to be a more slowly progressing field that uses a greater proportion of the older literature within the field whereas the dementia group is a more fast-moving group.

Discussion and conclusion

This study exemplifies the great potential of digital collections for statistical analyses. By statistically analyzing a large corpus of references in systematic reviews we are able to draw at least three conclusions regarding the phenomenon of non-citation:

1. The number of non-citations far outweighs the number of citations.
2. Citations and non-citations differ in age.
3. There are great differences between fields.

These conclusions would be difficult or even impossible to reach by other methods (i.e., close reading).

The data for this study consists of 120,000+ cited studies and more than 1.6 million non-cited studies. It is thus obvious that the number of cited studies is

³ The Cochrane Methodology Review Group is the Cochrane Review Group responsible for Cochrane methodology reviews. These are a special type of Cochrane review, examining the evidence on methodological aspects of systematic reviews, randomised trials and other evaluations of health and social care (<http://methodology.cochrane.org/>).

much smaller than the number of the non-cited. Apart from the difference in size pools of cited and non-cited studies differ in age. Very recent studies tend to be non-cited whereas the cited studies are rarely very recent (e.g. within the same year). The greatest differences are found within the first 10 years. After 10 years the cited and non-cited studies tend to be more similar in terms of age. Also, we have separated the data set into different sub-disciplines, and we find that the various sub-disciplines vary in terms of the age of cited and non-cited references. Some fields may be expanding, and the number of published studies is thus growing fast. Consequently, cited and non-cited studies tend to be younger. Other fields may be more slowly progressing fields that use a greater proportion of the older literature within the field. These field differences manifest themselves in the average age of references.

Our results confirm the results of previous studies of the non-citation phenomenon: Only a small fraction of the literature that could or should have been cited end up being cited. Traditionally, sceptics of citation analysis have argued that this makes research evaluation based on citation analysis an invalid evaluation method (MacRoberts and MacRoberts 1996, Seglen 1997). Conversely, proponents have defended citation analysis by arguing that as long as citation analyses are based on many reference lists, results are valid (e.g., Narin 1987, Nederhof and Van Raan 1987, Small 1987, White 2001). The sceptics have countered this by claiming that this would only be true if bias were distributed randomly, but biased citing, they claim, is not random (MacRoberts 1997). For instance, it is claimed that authors cite works by established authorities, so as to gain credibility by association (Gilbert 1977, Latour 1987). This hypothesis is known as *the persuasion by name-dropping hypothesis* (White 2004). As we saw in the review section, proponents of citation analysis have challenged the hypothesis by pointing to empirical studies of citation distributions that show little or no signs of such biased citing (e.g., Zuckerman 1987, Moed and Garfield 2003, White 2004). These studies reveal that authors do not tend to favour highly cited names or highly cited publications. Instead, authors generally cite the entire scale of citation reputation. However, these studies do not test the essence of the persuasion by name-dropping hypothesis. To test the hypothesis, we need to investigate whether authors that have a choice between citing equally citable authors or documents tend to choose citing the highly cited ones. As we have just shown, cohorts of research addressing the same research question may be identified using systematic reviews. The reviewed studies of any given systematic review may then be examined to determine which of the preceding studies were cited in later publications. On the basis of that analysis it should be possible to match pairs of cited and uncited studies of the same age and then to trace their number of citations at the time of citation/non-citation. In a forthcoming article (Frandsen and Nicolaisen, In Press), we present the results of a large-scale test of the persuasion by name-dropping using this approach on a dataset similar to the one we study in the present paper. Our results seem to suggest a more careful interpretation than simply name-dropping.

References

- Brooks, T. A. 1985. "Private acts and public objects: An investigation of citer motivations." *Journal of the American Society for Information Science* 36, 4: 223-229. <https://doi.org/10.1002/asi.4630360402>
- Brooks, T. A. 1986. "Evidence of complex citer motivations." *Journal of the American Society for Information Science* 37, 1: 34-36. <https://doi.org/10.1002/asi.4630370106>
- Cole, J. R., and S. Cole. 1972. "The Ortega hypothesis." *Science* 178, 4059: 368-375. <https://doi.org/10.1126/science.178.4059.368>
- Collier, A., L. Heilig, L. Schilling, H. Williams, and R. P. Dellavalle. 2006. "Cochrane skin group systematic reviews are more methodologically rigorous than other systematic reviews in dermatology." *British Journal of Dermatology* 155, 6: 1230-1235. <https://doi.org/10.1111/j.1365-2133.2006.07496.x>
- Dienes, I. 2012. "A meta study of 26 "how much information" studies: Sine qua nons and solutions." *International Journal of Communication* 6: 874–906.
- Frandsen, T. F., and J. Nicolaisen. In Press. "Citation behavior: A large-scale test of the persuasion by name-dropping hypothesis." *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.23746>
- Garfield, E. 1977. "To cite or not to cite: A note of annoyance." *Current Contents* 35, August 29: 5–8.
- Gilbert, G. N. 1977. "Referencing as persuasion." *Social Studies of Science* 7: 113-122. <https://doi.org/10.1177/030631277700700112>
- Higgins J. P. T., and S. Green. 2011. *Cochrane handbook for systematic reviews of interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org.
- Hilbert, M. and P. López. 2012. "How to measure the world's technological capacity to communicate, store and compute information? Part I-II: Results and scope." *International Journal of Communication* 6: 936-955; 956-979.
- Kitchin, R. 2014a. "Big data, new epistemologies and paradigm shifts." *Big Data %26 Society* 1: 1-12.
- Kitchin, R. 2014b. *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage. <https://doi.org/10.4135/9781473909472>

- Latour, B. 1987. *Science in action*. Cambridge, MA: Harvard University Press.
- MacRoberts, M. 1997. "Rejoinder." *Journal of the American Society for Information Science*, 48, 10: 963.
- MacRoberts, M. H., and B. R. MacRoberts. 1984. "The negational reference: Or the art of dissembling." *Social Studies of Science*, 14: 91-94. <https://doi.org/10.1177/030631284014001006>
- MacRoberts, M. H., and B. R. MacRoberts. 1986. "Quantitative measures of communication in science: A Study of the formal level." *Social Studies of Science* 16: 151-172. <https://doi.org/10.1177/030631286016001008>
- MacRoberts, M. H., and B. R. MacRoberts. 1987a. "Another test of the normative theory of citing." *Journal of the American Society for Information Science* 38: 305-306. [https://doi.org/10.1002/\(SICI\)1097-4571\(198707\)38:4<305::AID-ASI11>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-4571(198707)38:4<305::AID-ASI11>3.0.CO;2-I)
- MacRoberts, M. H., and B. R. MacRoberts. 1988. "Author motivation for not citing influences: A methodological note." *Journal of the American Society for Information Science* 39: 432-433. [https://doi.org/10.1002/\(SICI\)1097-4571\(198811\)39:6<432::AID-ASI8>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-4571(198811)39:6<432::AID-ASI8>3.0.CO;2-2)
- MacRoberts, M. H., and B. R. MacRoberts. 1989a. "Problems of citation analysis: A critical review." *Journal of the American Society for Information Science* 40: 342-349. [https://doi.org/10.1002/\(SICI\)1097-4571\(198909\)40:5<342::AID-ASI7>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U)
- MacRoberts, M. H., and B. R. MacRoberts. 1989b. "Citation analysis and the science policy arena." *Trends in Biochemical Science* 14: 8-10. [https://doi.org/10.1016/0968-0004\(89\)90077-7](https://doi.org/10.1016/0968-0004(89)90077-7)
- MacRoberts, M. H., and B. R. MacRoberts. 1996. "Problems of citation analysis." *Scientometrics* 36: 435-444. <https://doi.org/10.1007/BF02129604>
- MacRoberts, M. H., and MacRoberts. 1987b. "Testing the Ortega hypothesis: Facts and artefacts." *Scientometrics* 12: 293-295. <https://doi.org/10.1007/BF02016665>
- Moed, H. F., and E. Garfield. 2003. "Basic scientists cite proportionally fewer "authoritative" references as their bibliographies become shorter." *Proceedings of the 9th International Conference on Scientometrics and Informetrics*: 190-196.
- Moretti, F. 2013. *Distant Reading*. London: Verso
- Moseley, A. M., M. R. Elkins, R. D Herbert, C. G. Maher, and C. Sherrington. 2009. "Cochrane reviews used more rigorous methods than non-Cochrane reviews: Survey of systematic reviews in physiotherapy." *Journal of Clinical Epidemiology* 62, 10: 1021-1030. <https://doi.org/10.1016/j.jclinepi.2008.09.018>

- Narin, F. 1987. "To believe or not to believe." *Scientometrics* 12, 5-6: 343-344.
- Nederhof, A. J., and A. F. J. Van Raan. 1987. "Citation theory and the Ortega hypothesis." *Scientometrics* 12, 5-6: 325-328.
- Nicolaisen, J. 2007. "Citation analysis." *Annual review of Information Science and Technology* 41: 609-641. <https://doi.org/10.1002/aris.2007.1440410120>
- Ranasinghe, I., A. Shojaei, B. Bikdeli, A. Gupta, R. Chen, J. S. Ross, F. A. Masoudi, J. A. Spertus, B. K. Nallamothu, and H. M. Krumholz. 2015. "Poorly cited articles in peer-reviewed cardiovascular journals from 1997 to 2007: Analysis of 5-year citation rates." *Circulation* 131, 20: 1755-1762. <https://doi.org/10.1161/CIRCULATIONAHA.114.015080>
- Seglen, P. O. 1997. "Why the impact factor of journals should not be used for evaluating research." *British Medical Journal* 314, 7079: 498-502. <https://doi.org/10.1136/bmj.314.7079.497>
- Small, H. 1987. "The significance of bibliographic references." *Scientometrics* 12, 5-6: 339-341.
- Weale, A. R., M. Bailey, and P. A. Lear. 2004. "The level of non-citation of articles within a journal as a measure of quality: A comparison to the impact factor." *BMC Medical Research Methodology* 4: 14. <https://doi.org/10.1186/1471-2288-4-14>
- White, H. D. 2001. "Authors as citers over time." *Journal of the American Society for Information Science and Technology* 52, 2: 87-108. [https://doi.org/10.1002/1097-4571\(2000\)9999:9999<AID-ASI1542>3.0.CO;2-T](https://doi.org/10.1002/1097-4571(2000)9999:9999<AID-ASI1542>3.0.CO;2-T)
- White, H. D. 2004b. "Reward, persuasion, and the Sokal hoax: A study in citation identities." *Scientometrics* 60, 1: 93-120. <https://doi.org/10.1023/B:SCIE.0000027313.91401.9b>