Does Research With Statistics Have More Impact? The Citation Rank Advantage of Structural Equation Modelling[1]

Mike Thelwall, Paul Wilson
Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK.

Statistics are essential to many areas of research and individual statistical techniques may change the ways in which problems are addressed as well as the types of problems that can be tackled. Hence, specific techniques may tend to generate high impact findings within science. This article estimates the citation advantage of a technique by calculating the average citation rank of articles using it in the issue of the journal in which they were published. Applied to Structural Equation Modelling (SEM) and four related techniques in three broad fields, the results show citation advantages that vary by technique and broad field. For example SEM seems to be more influential in all broad fields than the four simpler methods, with one exception, and hence seems to be particularly worth adding to statistical curricula. In contrast, Pearson correlation apparently has the highest average impact in medicine but the least in psychology. In conclusion, the results suggest that the importance of a statistical technique may vary by discipline and that even simple techniques can help to generate high impact research in some contexts.

Introduction

Academic research typically has a topic and a method. Whilst some researchers may use a vague general approach, such as "the historical method" (Shafer & Bennett, 1969), others may employ specific named methods, such as virtual ethnography, the statistical t-test or the faecal egg count procedure. Within a discipline there may be a set of methods that are commonly taught to new researchers, although this set may change over time and vary between institutions and fields. In some research fields the same issue could be tacked with a wide variety of different but reasonable approaches, including qualitative and quantitative, and the investigator can use their judgement in selecting an overall approach (e.g., Creswell, & Clark, 2007; Easterbrook, Singer, Storey, & Damian, 2008). Some of these methods may make it possible to solve important problems that would otherwise be unresolved, whilst others may make more routine contributions, such as being an alternative to an existing technique, or may even tend to produce weak or erroneous findings. In this context, it would be useful to evaluate the general contribution of a method to research in order to help to prioritise or justify its teaching within a discipline and to motivate researchers to learn the methods that are apparently the most influential.

The value of a method may be obvious to practitioners, particularly if it is the only choice for solving key problems. Nevertheless, these practitioners may not be impartial because methods can be part of the identity and time investment of a researcher (Becher & Trowler, 2001) and so it is not clear how the general contributions of methods should be evaluated. In practice, methods are probably best judged by outsiders, such as journal referees, although this is not always a guarantee of the validity of the results (e.g., Tachibana, 1980), and perhaps also by new researchers selecting methods for their problems. Thus, the simple prevalence of a method in the academic literature is probably a good indicator of its value. Nevertheless, inappropriate methods may become popular because of their simplicity rather than their power (Wilson, 2014), methods may be routinely misused (Fabrigar, Wegener, MacCallum, & Strahan, 1999), and appropriate but

[1]This is a preprint of an article to be published in the Journal of the Association for Information Science and Technology

complex methods may be avoided by researchers because they lack the skills or resources to learn or apply them. Moreover, techniques may even be widely used because they are standard in a field despite severe criticisms by authoritative figures, as is the case for significance testing in psychology (Hubbard, Parsa, & Luthy, 1997). Hence other indicators of method value are also needed.

This article introduces a simple indicator for comparing the value of a set of methods against each other and against articles that do not use them. This indicator is an average normalised rank for articles mentioning the methods, compared to similar articles. The approach introduced here is simple enough to be practical and is unobtrusive in the sense of not calling for new peer judgements. The techniques are demonstrated for the case of structural equation modelling (SEM) (Kline, 2011). This is a multivariate technique that allows the researcher to hypothesise the existence of a set of relationships within a group of variables, for example by drawing a picture of the connections between them, and then test whether this set of relationships is consistent with empirical data. SEM was chosen because it is a relatively complex statistical technique (Anderson & Gerbing, 1988) and is widely used (Baumgartner & Homburg, 1996; MacCallum, & Austin, 2000; Steiger, 2001). Moreover, there have been concerns with its use in practice due to a lack of mastery by its users (Marcoulides & Saunders, 2006; Steiger, 2001) and a proliferation of interpretation errors (Schreiber, Nora, Stage, Barlow, & King, 2006) and so, despite its strengths, it might tend to produce low quality or erroneous research

Evaluating Research Methods

As argued above, counting the number of peer-reviewed publications that mention a research method gives a reasonable first assessment of its utility. This is because each publication reflects the uptake of the method by a researcher, in terms of their belief that it will help them to solve their research problem, as well as referee judgements that the completed research is valid and significant enough to be published. This number of articles and similar publications could be estimated by searches in a full-text database with a wide coverage, such as Google Scholar, Google Books or publishers' full-text digital libraries. An alternative is to search a database that indexes titles, abstracts and keywords. This has the disadvantage that methods are not always mentioned in these parts of an article but the advantage that it will presumably identify articles for which a method is particularly important. Citation indexes have search functions that make them particularly useful for identifying mentions of methods. They also have clearly defined coverage of publications which can help in the interpretation of the results.

An example of a methods prevalence study is an investigation of the statistical techniques of 623 articles published in 2010 from a sample of 8 journals with a high impact factor in the Clinical Psychology Web of Science category (Sesé & Palmer, 2012). The methods in these articles were classified into a set of intuitive categories by a human coder, presumably by reading them. The most common category was Correlation, but Factor Analysis and PCA was 12th and SEM was 13th.

To estimate the contribution of a method in terms of the power that it gives to research, it seems reasonable to compare the citation counts of articles that use the method within the citation counts of similar articles that don't. This is like the problem of assessing the citation impact of the publications of an academic department, which may be scattered across years, fields and journals. Methods for departmental citation impact normalisation have been extensively discussed. For example, article citations could be field-normalised by dividing them by the average number of citations received by documents of the same type, published in the same year and field, although other variants are also possible (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). Field normalisation is not ideal for analysing methods, however, because a method is likely to be disproportionately used by specialisms within fields and hence field normalisation would be unfair if these specialisms did not have the same citation norms as the field as a whole. Hence, it seems more appropriate to normalise by journal because journals can focus on individual specialist areas (e.g., scientometrics).

Normalising by journal is an imperfect solution because there are also generalist and multidisciplinary journals. Nevertheless, to at least partially correct for specialism and publication date, the article citations should be compared to citations to other articles published in the same journal and year. For a more fine-grained analysis the comparison could be restricted to a single volume or issue. Moreover, since citation distributions are typically skewed, it is advantageous to consider citation ranks rather than citation means (Leydesdorff, Bornmann, Mutz, & Opthof, 2011). This method may still be unfair because, even within a specialism, quantitative methods may tend to be more (or less) cited than qualitative methods (e.g., Antonakis, Bastardoz, Liu, & Schriesheim, 2014; Swygart-Hobaugh, 2004) although there is not always a difference (Stremersch, Verniers, & Verhoef, 2007) and so the advantage of a specific method may be due to its type rather than its specific nature. Comparing a set of methods rather than a single method can help to reduce this possibility by revealing the advantage of one method in comparison to others.

Some previous studies have used a variety of approaches to assess the contributions of methods to research. Statistical modelling has been used to identify factors associating with higher citation counts for articles published in *The Leadership Quarterly*, finding that some methods, including SEM, associated with higher citation counts for quantitative research (Antonakis, Bastardoz, Liu, & Schriesheim, 2014). Other statistical related-findings include that articles in the *Strategic Management Journal* with higher statistical power (amongst other properties) tended to be more cited (Bergh & Perry, 2006), and that almost half of the highest cited psychiatric articles focused on metrics (Mazhari, 2013). Some small scale studies have found empirical research to be more highly cited than other types, such as theoretical research (Bergh & Perry, 2006; Johnston, Piatti, & Torgler, 2013; Van Dalen & Henkens, 2001), and empirical studies seem to be the most likely to use statistics. More generally, articles in the *Strategic Management Journal* that used a mixed method approach tended to be more cited (Molina-Azorin, 2012).

Research designs have also been compared for their citation impacts within medicine (meta-analyses, randomized controlled trials, cohort studies, case-control studies, case reports, nonsystematic reviews, decision analyses or cost effectiveness analyses), on the basis that there may be a "hierarchy of evidence" between them. Meta-analyses tended to attract the most citations but the relative order of the designs changed somewhat between 1991 and 2001 (Patsopoulos, Analatos, & Ioannidis, 2005). This study identified articles for the sample with initial keyword searches of titles in the Web of Science, (e.g., for randomized controlled trials: TI=random* AND TI=trial) followed by manual screening of the matches based on article titles and abstracts. In addition to comparing raw citation counts, the study also used logistic regression to assess whether research design and various other factors (author affiliation country, journal impact factor) were significant predictors of an article being highly cited. Another investigation, with a smaller sample (137) restricted to clinical orthopaedic research, confirmed that different research designs could have different average citation impacts (Bhandari, Busse, Devereaux, et al., 2007), but a regression analysis of emergency medicine publications derived from a particular meeting did not find the type of research design to be significant for citation counts, except that the presence of a control group was significant (Callaham, Wears, & Weber, 2002).

A limitation of using citation counts to indicate the impact of academic publications is that articles can have other types of impact, such as societal impact, educational impact or impact through commercial applications. In response to this issue the fields of patent metrics (Breitzman & Mogee, 2002), webometrics (Vaughan & Shaw, 2003) and altmetrics (Priem, Piwowar, & Hemminger, 2012) have developed metrics for other types of impact of research. The newest field, altmetrics, specialises in developing metrics for the impact of academic articles from the social web, including tweet citations (Eysenbach, 2011), Mendeley readers (Li & Thelwall, 2012), blog citations (Shema, Bar-Ilan, & Thelwall, 2014) and many others (Thelwall, Haustein, Larivière, & Sugimoto, 2013; Costas, Zahedi, & Wouters, in press). Of these, Mendeley readers seem to be the most promising because Twitter citations seem to be typically quite trivial (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013), despite more substantial uses in previous years (Priem & Costello, 2010). Mendeley readers also have the important advantage of being relatively numerous, at least in some fields (Maflahi & Thelwall, in press; Zahedi, Costas, & Wouters, 2014) and appear to be positive indications that an article has been read, although this is not always true. Perhaps the most important limitation of Mendeley reader counts is that the users of Mendeley are "a small and biased sample" of the users of academic articles (Mohammadi, Thelwall, Haustein, & Larivière, in press).

Metrics for the impact advantage of specific methods

Suppose that a set of articles mentioning a method has been obtained as well as all of the citations to articles from the same journal and year (or volume or issue). There are many different ways with which to assess the citation impact of the method:

- *Create a statistical model for the number of citations to papers (e.g.,* Kulkarni, Busse, & Shams, 2007; *Didegah & Thelwall, 2013) and include the nature of the method as a factor, testing for its size and statistical significance*. This has the disadvantage that a single model will be needed to span a heterogeneous, non-random collection of articles from many different fields and from journals of different sizes, which will reduce the power of the test and undermine its accuracy, although it is possible (e.g., for a methodologically similar issue: Peng, & Zhu, 2012).
- *Compare the average impact of articles using the method with the average impact of the remaining articles*. If the former is divided by the latter then this simple indicator would be greater that 1 when articles using the method tended to have a higher citation impact. This has the disadvantage of being skewed towards large journals and for being relatively insensitive since the averaging will cross fields with different average levels of citation.
- *Calculate the normalised score for each article separately (e.g. article citations divided by issue citations) and average these scores*. This has the disadvantage that, due to skewed citation distributions, it has a tendency to be less than 1, especially for small samples.
- *Calculate the normalised rank of each article in its reference set (i.e., journal and year, volume or, issue) by (rank-1)/(articles-1), which is between 0 and 1, and average these figures*. Switching from means to ranks prevents skewed citation distributions from affecting the results (Leydesdorff et al., 2011). This indicator is called the *normalised citation rank*.

A disadvantage that all of the above indicators suffer from, and which seems to be unavoidable, is that they will be less powerful for collections of articles with fewer citations. This could occur for fields with low citation norms and for recent articles that have not had time to accrue many citations. To illustrate the power issue with a simple example, suppose that 3 articles are being compared. The minimum number of citations needed to rank them completely is 3, assuming that they receive 2, 1 and 0 citations. Hence if they collectively receive less than 3 citations then their ranks will have ties. More generally, the more citations collectively received by a set of articles, the more likely they are to be ranked without ties, assuming that they all have a different rate of attracting citations. Thus, the more citations received by a collection of articles, the more likely they are to be ranked “correctly” by their citations. In practice, there will be many ties for any set of citation counts due to the skewed nature of this data, but there will tend to be fewer ties when there are more citations.

Research Questions

The method evaluation technique will be demonstrated by applying it to SEM and four related techniques: Confirmatory Factor Analysis (CFA), which is a special case of SEM, Exploratory Factor Analysis (EFA), which is an exploratory version of CFA, Principal Components Analysis (PCA), which has similarities with EFA but a different theoretical approach (Fabrigar, Wegener, MacCallum, & Strahan, 1999), and Pearson Correlation (Cor), which is a simple bivariate technique to test for a linear relationship between variables. These four techniques were chosen for having goals and methods that are related to the main version of SEM. Other techniques, such as Spearman correlation, could also have been chosen for being related to the non-parametric version of SEM but the four chosen here seem to be the most closely related. Although path analysis is a type of SEM, it is excluded because the phrase *path analysis* is often used in other contexts and hence it will be difficult to accurately identify articles using it. Citations are used as the primary data source because this is the standard for impact assessment, and Mendeley readers are used as a secondary data source to help guard against citation analysis results being due to patterns of citation rather than patterns of use of articles. The following research questions are designed to assess the impact of SEM and related techniques.

- Do articles using SEM or related techniques tend to have above average impact for their specialism?
- Do articles using SEM or related techniques tend to have above average numbers of readers for their specialism?
- Are there disciplinary differences in the uptake and impact of SEM and related techniques?

Methods

The study is based on publications indexed in Scopus. This was chosen in preference to Google Scholar because the latter does not allow discipline-specific searches. The Web of Science would probably have given similar results and could have covered a longer time period but Scopus has wider coverage of the social sciences (Erfanmanesh & Didegah, 2013; Li, Burnham, Lemley, & Britton, 2010; Minasny, Hartemink, McBratney, & Jang, 2013). The years 1996-2012 were chosen because 1996 is the start of extended coverage for Scopus (Li et al., 2010; Scopus, 2011) and 2013 is too recent (in mid-2014) to give enough citations to analyse.

The medicine, psychology and social science subject areas were selected as, according to Scopus searches, the three largest users of SEM. Articles were searched for in Scopus using year-specific searches in the following forms in July 2014, where X was replaced by MEDI (medicine), PSYC (psychology) or SOCI (social science) and a separate query was used for each year examined.

- TITLE-ABS-KEY("Pearson Correlation") AND SUBJAREA(X) AND NOT TITLE-ABS-KEY("factor analysis")
- TITLE-ABS-KEY("Principal components analysis") AND SUBJAREA(X) AND NOT TITLE-ABS-KEY("factor analysis")
- TITLE-ABS-KEY("exploratory factor analysis") AND SUBJAREA(X) AND NOT TITLE-ABS-KEY("structural equation modeling")
- TITLE-ABS-KEY("confirmatory factor analysis") AND SUBJAREA(PSYC) AND NOT TITLE-ABS-KEY("structural equation modeling")
- TITLE-ABS-KEY("Structural equation modelling") OR TITLE-ABS-KEY("structural equation model") OR TITLE-ABS-KEY("LISREL") AND SUBJAREA(X)

To ensure that each article matching one of the above did not also match a more complex method, records that matched a query lower down the order in the above list were removed. For example, any record that matched the Pearson correlation query that also matched the PCA query was removed (with a computer program to automatically identify matches) from the set of results for the Pearson query. The methods could also be referred to in other ways that the queries may have missed, such as "Pearson product moment" for Pearson correlation or acronyms for the other methods. The acronyms were not used because of alternative possible meanings (e.g., Scanning Electron Microscope for SEM, and Proportional Counter Array for PCA) and the term Pearson product moment was relatively rare (5% of the main term) but, in retrospect could have also been included. Pearson correlation could also have been referred to without the name of the method's creator but the term correlation alone was not included because it is ambiguous and could match Spearman correlation and could also be used in a general, non-statistical sense.

Each record obtained as above was checked and, for those with a volume and issue, the remaining articles in the issue were downloaded with a Scopus query for the journal name, year, volume and issue, as in the following example.

SRCTITLE("Academic Emergency Medicine") AND YEAR(1994) AND VOLUME(1) AND ISSUE(3)

To assess whether articles using a specific technique tended to attract more (or less citations) than normal for their specialism, the normalised citation rank was calculated by *(rank-1)/(articles-1)* for each article. Issues were used rather than volumes as a practical step to limit the amount of data that needed to be gathered. The averages of the normalised ranks were calculated for each year and 95% confidence limits were also calculated, assuming a normal distribution (which should not skew the results much since the values are all bounded between 0 and 1). The years 1996-2000 and 2001-2005 were grouped in order to give large enough numbers for small confidence limits because of the low numbers of matches before 2006.

Following previous practice (Maflahi & Thelwall, in press), Mendeley readers were collected with Webometric Analyst by querying the free Mendeley API (Applications Programming Interface) for articles using the article title, first author last name, and publication year. In cases where multiple articles were returned, the record with the largest number of readers was retained. In cases where no record was found, the article was assumed to have zero Mendeley readers although in some cases an article may have had readers but the article details were recorded differently between Mendeley and Scopus. Mendeley information was collected for all articles identified from Scopus, including all articles in the same issues as the articles matching the five statistical methods queries.

Results

Figures 1-3 show the number of Scopus articles for each technique in each of the three broad fields. Although the articles show an increasing trend, an unknown proportion of this trend is due to increases in the amount of journals covered by Scopus and so the graphs should not be taken as evidence of increasing use of the techniques. Nevertheless, there are clearly different levels of uptake for the five methods between broad fields. For example, whilst SEM is the most mentioned method in psychology and social science, it is only the third most popular method in medicine. In contrast, Pearson correlation is the least mentioned of the five methods in both psychology and social sciences but is the third most popular method in medicine. Although differences in styles of writing abstracts may account for some of this difference – for example, if medical articles tend to use structured abstracts that give more details of methods – this seems unlikely to account the differences in ranking between methods found here.

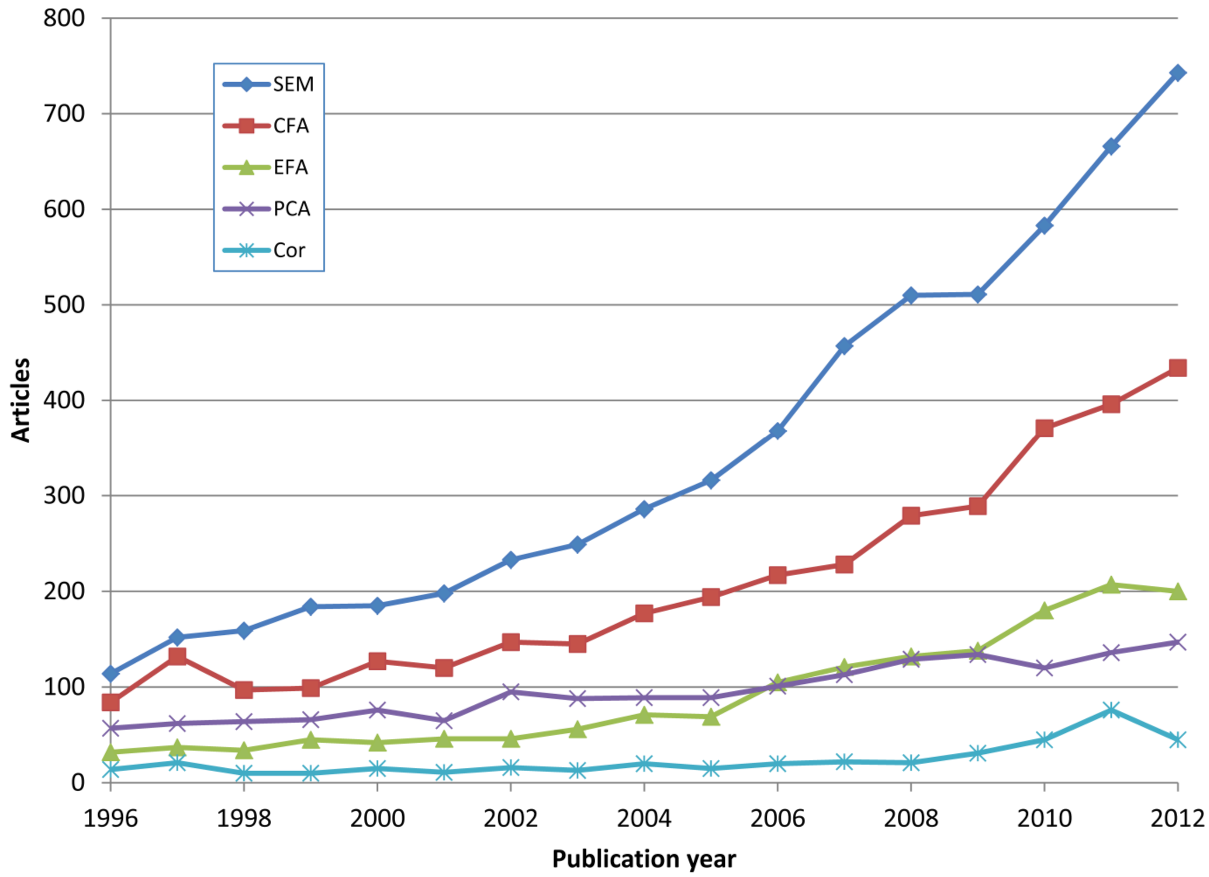

Figure 1. The number of psychology articles retrieved from Scopus for the five methods.

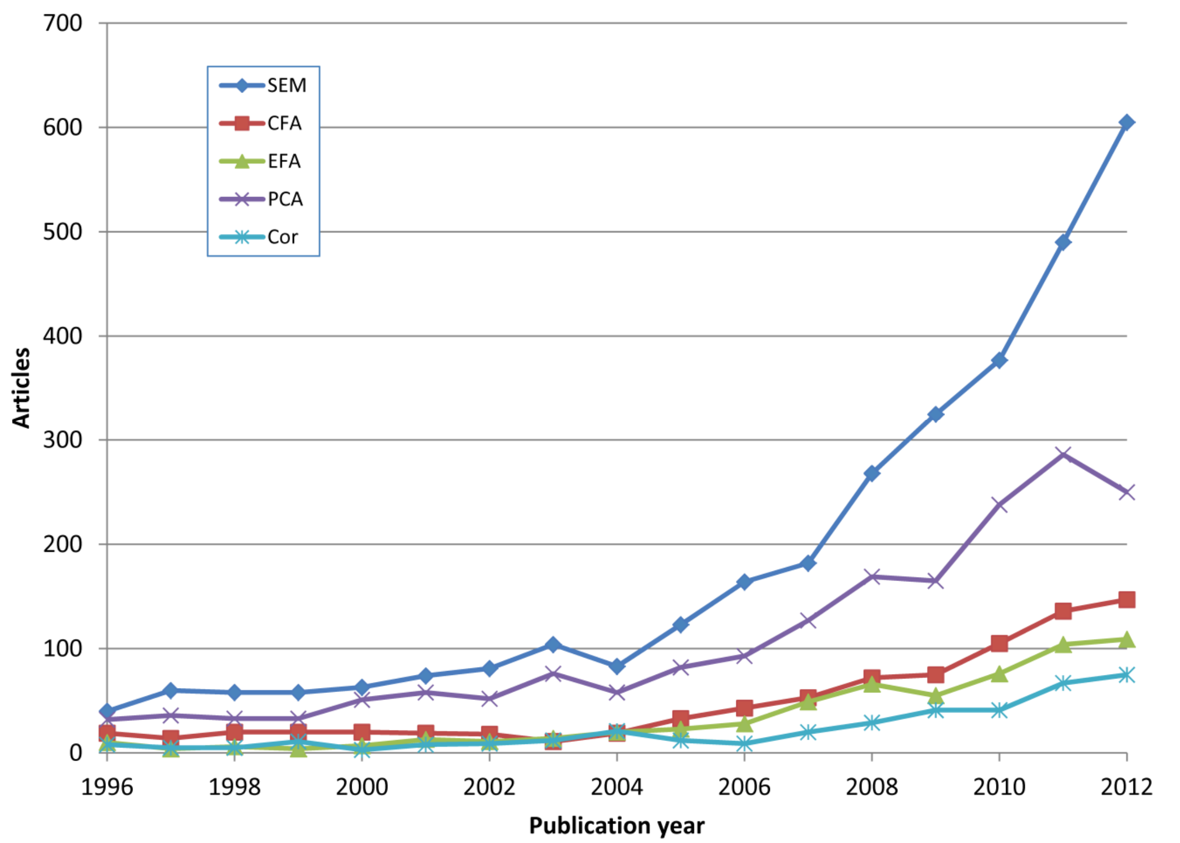

Figure 2. The number of social science articles retrieved from Scopus for the five methods.

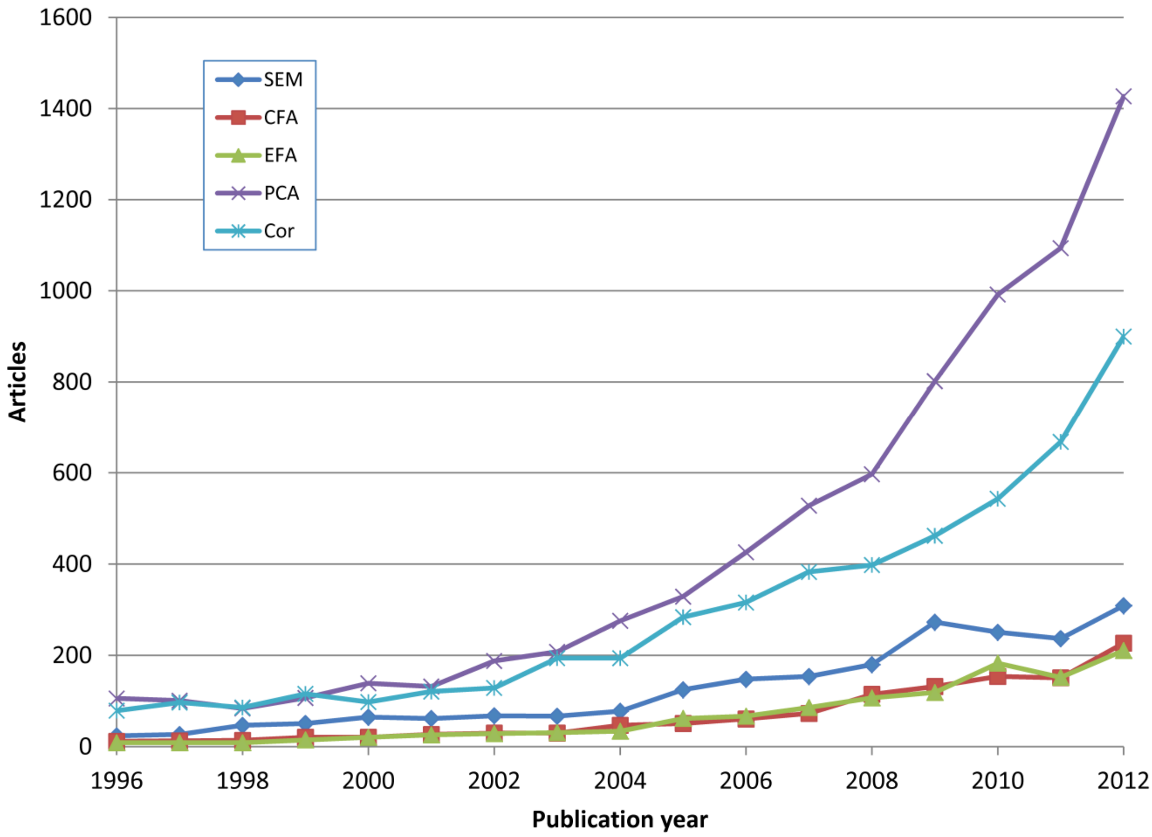

Figure 3. The number of medicine articles retrieved from Scopus for the five methods.

Citation ranks. Figures 4-6 suggest that, in general, articles using any of the five methods tend to be slightly more highly cited than other articles in their specialism. This is because the points on these graphs tend to be above 0.5 (i.e., the average for the issues in which they are published), indicating that the mean normalised citation rank is above average. The main exception is Pearson correlation within psychology. The confidence intervals for the points on the lines in the figures (not shown) are too wide to give clear evidence that one method tends to be more highly ranked than another during any particular year, however. Nevertheless, from Table 1 it is clear that SEM has the highest citation rank advantage in all broad fields and that the citation rank advantage of individual methods varies between broad fields.

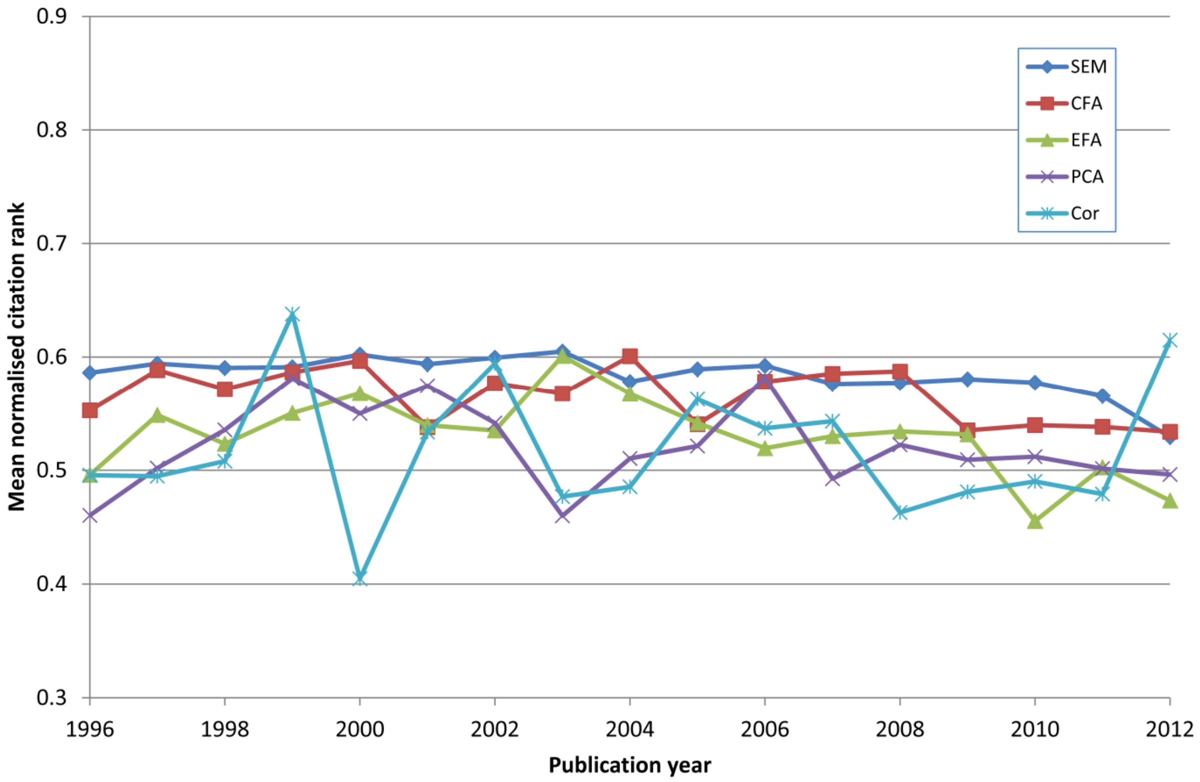

Figure 4. The mean Scopus citation rank of psychology articles for the five methods in their issue.

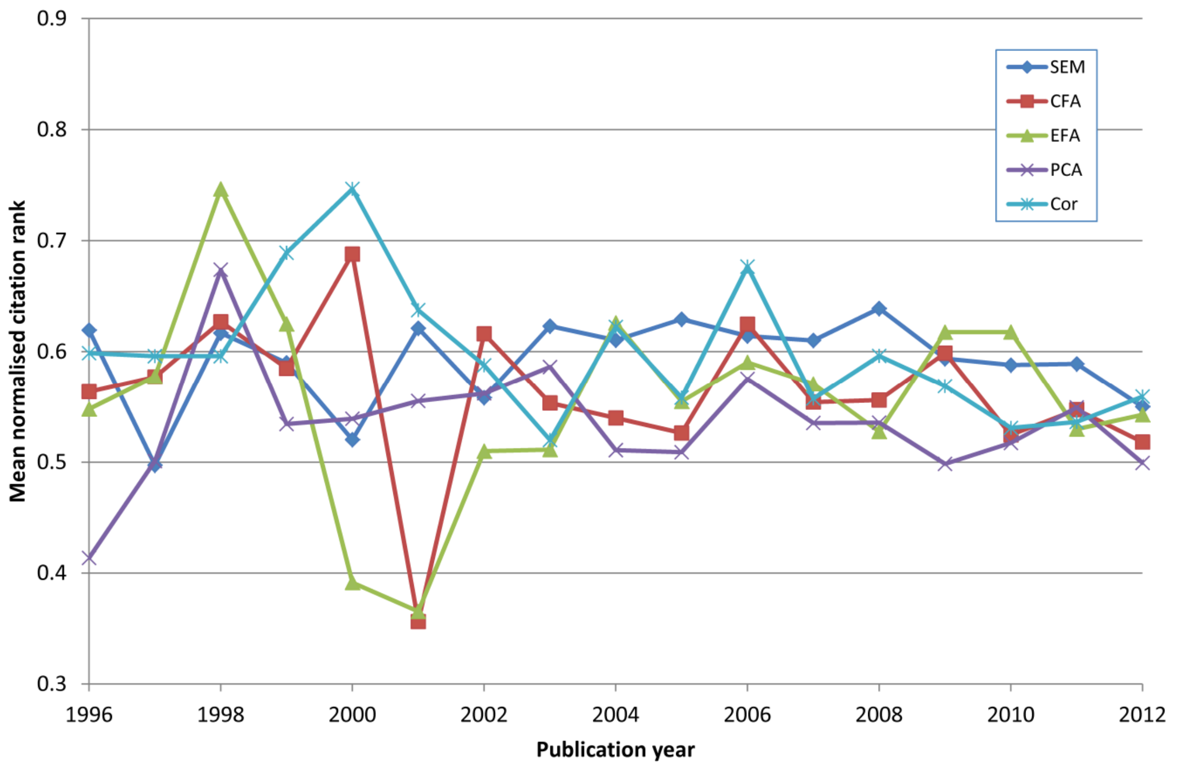

Figure 5. The mean Scopus citation rank of social science articles for the five methods in their issue.

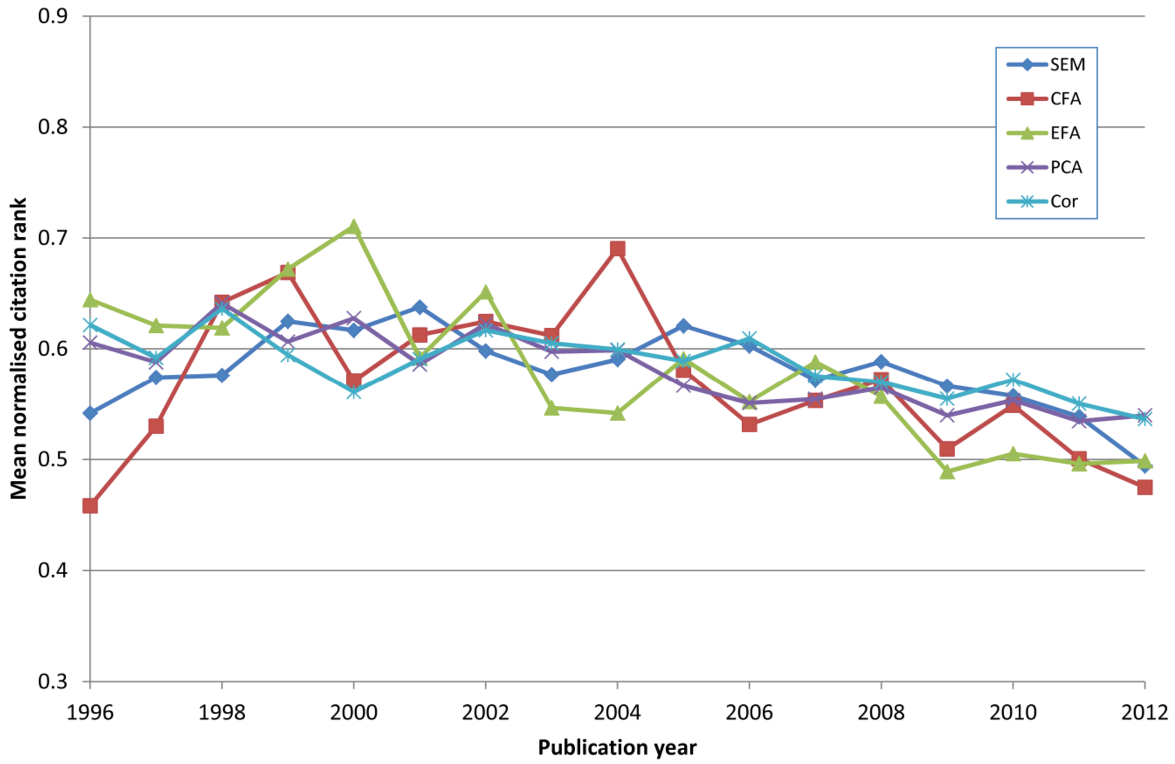

Figure 6. The mean Scopus citation rank of medicine articles for the five methods in their issue.

Table 1. 95% confidence intervals for normalised Scopus citation ranks in issues for all data.

	Psychology	Social science	Medicine
SEM	(0.57, 0.58)	(0.58, 0.60)	(0.55, 0.58)
CFA	(0.55, 0.57)	(0.53, 0.57)	(0.52, 0.56)
EFA	(0.50, 0.53)	(0.53, 0.58)	(0.52, 0.55)
PCA	(0.50, 0.53)	(0.52, 0.54)	(0.55, 0.56)
Cor	(0.49, 0.54)	(0.54, 0.60)	(0.57, 0.59)

Readership ranks. Figures 7-9 and Table 2 broadly conform to the patterns in figures 3-6 and Table 1, except that the readership advantage is not as large as the citation advantage. This may be due to the smaller numbers of readers than citers for this data. The lower readership is consistent with the statistics-based papers being more cited but not necessarily more read because a proportion of Mendeley readers are not citing authors (e.g., students), and so Mendeley readership partly reflects citing authors and partly reflects others. Moreover, Mendeley is not just used for research but is also used for education and professional activities (Mohammadi, 2014). The lower apparent readership advantage could also be due to the lower number of registered Mendeley readers than citations, however, since the lower numbers reduce the power of the statistics.

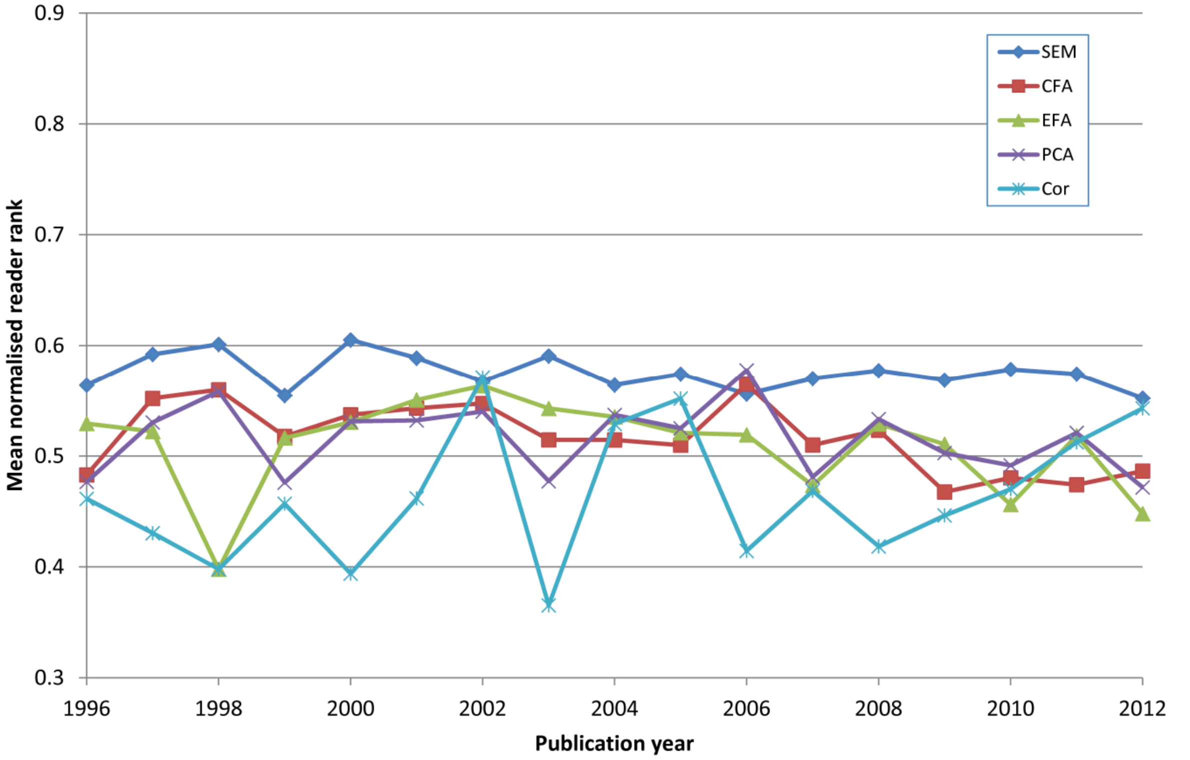

Figure 7. The mean Mendeley readership rank of psychology articles for the five methods in their issue.

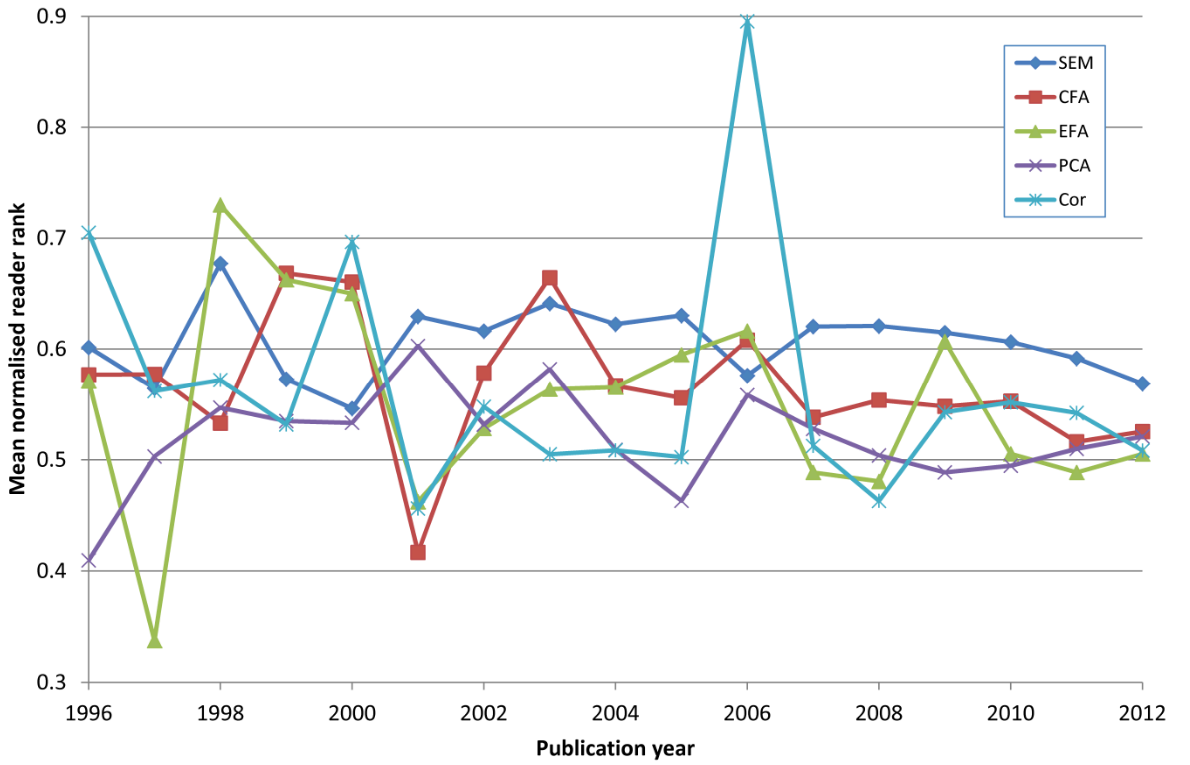

Figure 8. The mean Mendeley readership rank of social science articles for the five methods in their issue.

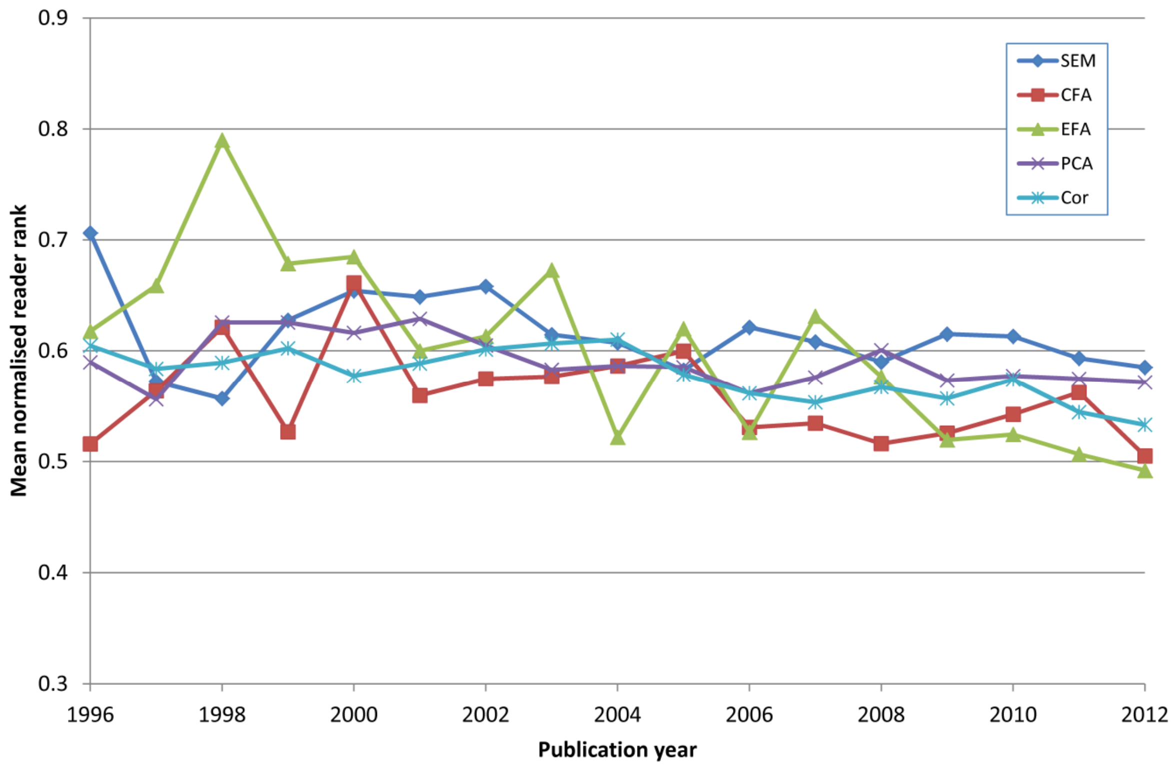

Figure 9. The mean Mendeley readership rank of medicine articles for the five methods in their issue.

Table 2. Confidence intervals for normalised Mendeley readership ranks in issues for all data.

	Psychology	Social science	Medicine
SEM	(0.56, 0.58)	(0.59, 0.61)	(0.59, 0.62)
CFA	(0.50, 0.52)	(0.53, 0.57)	(0.52, 0.56)
EFA	(0.50, 0.52)	(0.50, 0.55)	(0.53, 0.57)
PCA	(0.50, 0.53)	(0.50, 0.53)	(0.57, 0.59)
Cor	(0.45, 0.51)	(0.51, 0.56)	(0.56, 0.58)

The declared occupations of the Mendeley article readers were also compared to the declared occupations of the Mendeley readers of the other articles in the same issues to assess whether papers using statistical techniques attracted different readers. These were compared by averaging the percentages of readers for each method and then comparing the averages, correcting for the fact that there were 5.3% more declared reader occupations for the selected articles than average for the issues that they were in. Although there were some differences in the percentages of readers of each type of article between the articles mentioning one of the five techniques and the remaining articles, these exceeded 1% in only two cases and are hence minor (Table 3). These figures are also broadly consistent with other studies of Mendeley (e.g., Mohammadi, 2014). It is perhaps relevant that bachelor's degree students had a tendency to avoid the articles mentioning the four more advanced statistical techniques and master's degree students tended to avoid all five methods, especially CFA. Perhaps more surprisingly, however, PHD students also tended to avoid four out of the five methods. In contrast, professors were 2% more likely to read SEM articles than other articles in the same issue. These observations are consistent with more senior academics being slightly more likely to read articles mentioning the statistical methods investigated here.

Table 3. Average percentages of declared reader types in Mendeley for articles mentioning each technique.

Mean	SEM	CFA	EFA	PCA	Cor
PhD Student	27%-	27%	25%-	27%-	21%-
Master Student	13%-	13%--	14%-	11%-	12%-
Doctoral Student	7%	8%+	7%	4%	5%
Postdoc	3%	3%	4%+	8%	7%
Bachelor Student	4%-	4%-	4%-	4%-	4%
Researcher at an Academic Institution	3%	3%	4%	5%+	6%
Professor	5%++	2%	3%+	2%	3%
Assistant Professor	3%	3%+	2%	2%+	3%
Postgrad	2%	2%	3%+	2%	4%
Other Professional	1%-	2%	2%-	3%	6%
Researcher at a Non- Academic Institution	1%	1%	2%	3%	4%
Associate Professor	1%	2%	2%	2%	2%+
Lecturer	1%	1%	1%	1%	2%+
Senior Lecturer	0%	0%	1%	1%+	1%
Librarian	1%	1%	1%	1%	1%
Total reported Mendeley occupations	**71%**	**73%**	**76%**	**75%**	**81%**

+, ++, -, --: 1%, 2% higher and 1%, 2% lower than the average percentage for the issue.

Discussion

The most important limitation of this study is that the higher citation ranks for articles using the specified methods may be due to them tending to be in fields with higher citation norms. Although the method used is designed to correct for this by comparing articles only with other articles in the same journal (and issue), it did not distinguish between specialist and generalist journals and in any case, articles from the same subject are known to scatter unevenly across journals (Bradford, 1934; Nicolaisen & Hjørland, 2007) and so it would not be possible to ensure that only articles within the same specialism were compared. Hence, whilst average normalised citation ranks above 0.5 are consistent with the methods having a higher than average impact within their specialism, they do not prove this, even when the citation results agree with the readership results.

Another limitation is that articles may employ the methods analysed here without mentioning them in their abstracts. This seems particularly likely to be true for the simpler methods, such as Pearson correlation, that may not be the central method of a paper, and less likely for SEM.

The papers matching the searches were analysed in more detail to give insights into why each of the five statistical methods were useful in each of the three areas. For this, a list of the 20 words that occurred disproportionately often for one method compared to the other three was compiled for each method and field (15 lists of 20 words in total). From these lists, some patterns emerged.

Articles mentioning SEM unsurprisingly tended to disproportionately often discuss types of relationships between variables (e.g., direct, indirect, mediated). Within medicine SEM was used as a specific technique for monozygotic vs. dizygotic twin studies, but it did not have a specific topic association in the other areas. In contrast, PCA was strongly associated with specific and different methods in each discipline. In psychology it was used to analyse event-related potential evidence from subjects' brains. In medicine, it was often used in nuclear magnetic resonance spectroscopy, for example to detect metabolites. In social sciences, it was used to determine the multiple origins of water samples. Both *EFA* and *CFA* tended to be used to develop or analyse questionnaires or psychometric tests and co-occurred with words relating to validly assessment. Pearson correlations were disproportionately often used in conjunction with other simple techniques (t-tests, ANOVA, regression) and p-values were often explicitly mentioned. In medicine the technique was relatively often found in studies with patients undergoing treatments, and in psychology it was often used for a depression rating scale. Overall, it is probably not possible to disentangle the impact of specific statistical techniques from the purposes that they are used for. This is particularly true for PCA.

A possible reason for the apparent citation advantage of Pearson correlations in medicine is that medical studies may require considerable effort to obtain sufficient data to analyse (e.g., if each data point is based upon measurements of an individual patient, with their permission and ethical approval) and that, in this context, the use of any statistics suggests a study with a large population and hence substantial financial backing. In contrast, areas of study in which larger data sets are easier to obtain might need more complex analyses to generate important findings. Alternatively, medical research may need a higher standard of proof for its findings and SEM may not meet these standards. This is because, in contrast to most other standard statistical methods, a positive result only shows that the SEM model is consistent with the data and does not give evidence to reject the alternative hypothesis that the model is incorrect. The differing statistical needs for health-related research is perhaps recognized by the existence of a separate research field for it, known as medical statistics or biostatistics, although many general statistical methods are used to solve medical problems (Armitage, Berry, & Matthews, 2008).

Although higher citation counts tend to positively correlate with higher quality for academic articles (Franceschet & Costantini, 2011), there are field and subfield differences in citation norms. In this context, although the results are consistent with certain methods tending to be mentioned in higher impact articles in particular broad fields, this could be because they tend to be used for problem topics that generate a higher level of impact, even within their specialism.

Conclusions

The results suggest that SEM may be a particularly influential technique in the sense that articles mentioning it tend to be more highly cited than average for their journal and issue, and tend to be more highly cited than the other related techniques examined, with the exception of Pearson correlation in medicine. Although the apparent citation advantage of SEM could be due to specialisms using SEM tending to be more highly cited than other similar research, the findings should at least give confidence to those using and teaching this technique, and to those calling for more advanced statistical training for researchers (e.g., in psychology: Aiken, West, & Millsap, 2008). PCA is perhaps a special case because it tends to be used for a range of specific methods rather than being employed as a general purpose statistical tool. Moreover, EFA and CFA seem to be particularly relevant for validating and analysing questionnaires and psychometric tests.

The apparent citation advantage of Pearson correlations in medicine, in contrast to other fields, gives evidence that the most influential statistical techniques may vary by field and so those teaching statistics and those conducting research should be sensitive to field norms in the use of statistical methods. For example, whilst more complex techniques are needed to disentangle direct and indirect affects in some studies, Pearson correlations, in conjunction with other simple methods, seem to be adequate for many medical investigations involving patient treatments. It is also possible that the type of data influences this conclusion, however, because the Pearson correlation requires normally distributed data whereas SEM can be applied to data with any type of distribution using bootstrapping. Of course, for any particular research study the choice of statistical method should be driven by the type of data that can be gathered and the analysis required and not by the average impact of particular methods. Nevertheless, the findings suggest that it is reasonable to equip PhD students and other researchers with complex statistical techniques so that they are able to tackle topics or types of analyses that may be particularly useful.

More generally, the techniques introduced here could be used to assess the value of any research method (or perhaps even theoretical contributions, which are sometimes examined: e.g., Hoffman & Novak, 2009), as long as it tended to be mentioned in article abstracts or keywords. This may help to filter out poor research methods and highlight the particularly important ones, although the limitations discussed above mean that the results should be interpreted cautiously. It may also be that future studies of many methods may also identify general rules, such as more complex statistical methods tending to generate higher citation impacts than do simpler methods, except in specific identifiable contexts.

References

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, 63(1), 32-50.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411.

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C. A. (2014). What makes articles highly cited? *The Leadership Quarterly*, 25(1), 152-179.

Armitage, P., Berry, G., & Matthews, J. N. (2008). Statistical methods in medical research. New York: John Wiley & Sons.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: a review. *International Journal of Research in Marketing*, 13(2), 139-161.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. New York: McGraw-Hill International.

Bergh, D. D., & Perry, J. (2006). Some predictors of SMJ article impact. *Strategic Management Journal*, 27(1), 81-100.

Bhandari, M., Busse, J., Devereaux, P. J., Montori, V. M., Swiontkowski, M., Tornetta III, P., Einhorn, T.A., Khera, V., & Schemitsch, E. H. (2007). Factors associated with citation rates in the orthopedic literature. Canadian Journal of Surgery, 50(2), 119.

Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, 26, 85-86.

Breitzman, A. F., & Mogee, M. E. (2002). The many applications of patent analysis. *Journal of Information Science*, 28(3), 187-205.

Callaham, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA, 287(21), 2847-2850.

Creswell, J. W., & Clark, V. L. P. (2007). Designing and conducting mixed methods research. London: Sage.

Costas, R., Zahedi, Z., & Wouters, P. (in press). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*. arXiv preprint arXiv:1401.4321.

Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, 7(4), 861-873.

Easterbrook, S., Singer, J., Storey, M. A., & Damian, D. (2008). Selecting empirical methods for software engineering research. In F. Shull (ed.). Guide to advanced empirical software engineering (pp. 285-311). London: Springer.

Erfanmanesh, M. A., & Didegah, F. (2013). A Comparison of Web of Science and Scopus for Iranian Publications and Citation Impact. *International Journal of Information Science & Management*, 11(1), 11-27.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological methods, 4(3), 272.

Franceschet, M., & Costantini, A. (2011). The first Italian research assessment exercise: A bibliometric perspective. Journal of Informetrics, 5(2), 275-291.

Hoffman, D. L., & Novak, T. P. (2009). Flow online: lessons learned and future prospects. Journal of Interactive Marketing, 23(1), 23-34.

Hubbard, R., Parsa, R. A., & Luthy, M. R. (1997). The Spread of Statistical Significance Testing in Psychology The Case of the Journal of Applied Psychology, 1917-1994. Theory & Psychology, 7(4), 545-554.

Johnston, D. W., Piatti, M., & Torgler, B. (2013). Citation success over time: theory or empirics? *Scientometrics*, 95(3), 1023-1029.

Kline, R. B. (2011). Principles and practice of structural equation modeling. New York: Guilford Press.

Kulkarni, A. V., Busse, J. W., & Shams, I. (2007). Characteristics associated with citation rate of the medical literature. PloS ONE, 2(5), e403.

Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables on citation analysis one more time: Principles for comparing sets of documents. *Journal of the American Society for Information Science and Technology*, 62(7), 1370-1381.

Li, J., Burnham, J. F., Lemley, T., & Britton, R. M. (2010). Citation Analysis: Comparison of Web of Science, Scopus, SciFinder, and Google Scholar. *Journal of Electronic Resources in Medical Libraries*, 7(3), 196-217.

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461-471.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51(1), 201-226.

Maflahi, N. & Thelwall, M. (in press). When are readers as good as citers for bibliometrics? Scopus vs. Mendeley for LIS journals. Journal of the Association for Information Science and Technology.

Mazhari, S. (2013). The 100 top-cited articles published in psychiatric journals. *Journal of Psychiatric Practice*, 19(4), 327-338.

Marcoulides, G. A., & Saunders, C. (2006). PLS: A silver bullet? *Management Information Systems Quarterly*, 30(2), iii-ix.

Minasny, B., Hartemink, A. E., McBratney, A., & Jang, H. J. (2013). Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar. PeerJ, 1, e183.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (in press). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*.

Mohammadi, E., (2014). Mendeley Altmetrics. PhD Thesis: University of Wolverhampton.

Molina-Azorin, J. F. (2012). Mixed methods research in strategic management: Impact and applications. *Organizational Research Methods*, 15(1), 33-56.

Nicolaisen, J., & Hjørland, B. (2007). Practical potentials of Bradford's law: A critical examination of the received view. *Journal of Documentation*, 63(3), 359-377.

Patsopoulos, N. A., Analatos, A. A., & Ioannidis, J. P. (2005). Relative citation impact of various study designs in the health sciences. JAMA, 293(19), 2362-2366.

Peng, T. Q., & Zhu, J. J. (2012). Where you publish matters most: A multilevel analysis of factors affecting citations of internet studies. *Journal of the American Society for Information Science and Technology*, 63(9), 1789-1803.

Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, 47(1). https://asist.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/201_Final_Submission.pdf

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. arXiv preprint arXiv:1203.4745.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338.

Scopus (2011). Content coverage guide. http://www.scopus.fecyt.es/SiteAssets/Pages/info_editores/sccg0510.pdf

Sesé, A., & Palmer, A. (2012). The Current Use of Statistics in Clinical and Health Psychology Under Review. Clínica y Salud, 23(1), 97-108.

Shafer, R. J., & Bennett, D. H. (1969). A guide to historical method. Homewood, Illinois: Dorsey Press.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. Journal of the Association for Information Science and Technology, 65(5), 1018–1027.

Steiger, J. H. (2001). Driving fast in reverse: The relationship between software development, theory, and education in structural equation modelling. *Journal of the American Statistical Association*, 96(453), 331-338.

Stremersch, S., Verniers, I., & Verhoef, P. C. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing*, 71(3), 171-193.

Swygart-Hobaugh, A. J. (2004). A citation analysis of the quantitative/qualitative methods debate's reflection in sociology research: Implications for library collection development. *Library Collections, Acquisitions, and Technical Services*, 28(2), 180-195.

Tachibana, T. (1980). Persistent erroneous interpretation of negative data and assessment of statistical power. *Perceptual and Motor Skills*, 51(1), 37-38.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. PloS ONE, 8(5), e64841.

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, (17), paper 1. http://cybermetrics.cindoc.csic.es/articles/v17i1p1.pdf

Van Dalen, H. P., & Henkens, K. (2001). What makes a scientific article influential? The case of demographers. Scientometrics, 50(3), 455-482.

Vaughan, L., & Shaw, D. (2003). Bibliographic and web citations: what is the difference?. Journal of the American Society for Information Science and Technology, 54(14), 1313-1322.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37-47.

Wilson, P. (2014). The misuse of the Vuong test for non-nested models to test for zero-inflation. In Kneib, T. ed. Proceedings of the 29th International Workshop on Statistical Modelling (IWSM2014), University of Göttingen http://cybermetrics.wlv.ac.uk/paperdata/misusevuong.pdf.

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, 101(2), 1491-1513.