

Guideline References and Academic Citations as Evidence of the Clinical Value of Health Research¹

Mike Thelwall, Nabeil Maflahi

This article introduces a new source of evidence of the value of medical-related research: citations from clinical guidelines. These give evidence that research findings have been used to inform the day-to-day practice of medical staff. In order to identify whether citations from guidelines can give different information from that of traditional citation counts, this article assesses the extent to which references in clinical guidelines tend to be highly cited in the academic literature and highly read in Mendeley. Using evidence from the UK, references associated with the UK's National Institute of Health and Clinical Excellence (NICE) guidelines tended to be substantially more cited than comparable articles, unless they had been published in the most recent three years. Citation counts also seemed to be stronger indicators than Mendeley readership altmetrics. Hence, whilst presence in guidelines may be particularly useful to highlight the contributions of recently-published articles, for older articles, citation counts may already be enough to recognise their contributions to health in society.

Introduction

In order to identify and disseminate best practice for treating patients, many governments periodically pool together medical experts to review existing evidence and then formulate official guidelines for diagnosis and treatment (Shekelle, Woolf, Eccles, & Grimshaw, 1999; Woolf, Grol, Hutchinson, Eccles, & Grimshaw, 1999). In turn, physicians, surgeons, nurses and other medical practitioners are then expected to keep up-to-date with relevant guidelines but to use their professional judgement when applying them, improving patient outcomes (Grimshaw & Russell, 1993). Being cited in a clinical guideline is direct evidence that a study has had a societal impact by guiding medical practice. Although guidelines are not universally followed (Fleming, Mackley, Camacho, et al., 2014; Haymart, 2010), they can be an authoritative source of information in their own right (Clark, 2011) and may also directly influence working protocols for health professionals in particular contexts (Yue, Tabloski, Dowal, Puella, Nandan, & Inouye, 2014). Guidelines also have the advantage that they are public documents that can be evaluated post hoc by researchers seeking to identify areas for improvement (Deader, Tiboni, Malone, & Fairhurst, 2012; Tabassum & Batty, 2013). Given the importance of guidelines, when evaluating medical researchers it would be appropriate to include information about the guidelines that they have been cited in (e.g., Ovseiko, Oancea, & Buchan, 2012). Whilst an individual researcher could add information about citations from guidelines to their CV, when evaluating departments or other large groups it would be more appropriate to use some form of guideline citation count.

Guidelines have already started to attract attention from medical research funders as potential indicators of societal impact (Kryl, Allen, Dolby, Sherbon, & Viney, 2012). In theory, presence in guideline reference lists could serve as evidence that a research funder's money has been useful in ways other than scientific knowledge building. Assessing the value of medical funding is of continuing importance (e.g., Glover, Buxton, Guthrie, Hanney, Pollitt

¹ This is a preprint of an article to be published in the Journal of the Association for Information Science and Technology © copyright 2014 John Wiley & Sons, Inc.

& Grant, 2014). Unfortunately, however, guidelines do not always contain references, and, when present, these references do not explicitly indicate their value to the guideline (Kryl, Allen, Dolby, Sherbon, & Viney, 2012). Nevertheless, if research funders push for references to be systematically added to, and labelled in, clinical guidelines then they will form a rich and systematic source of impact evidence (Kryl, Allen, Dolby, Sherbon, & Viney, 2012). Of course there are likely to be limitations, such as citations of review articles rather than the original studies, biases in the guidelines selection committees (e.g., towards their own research), and citing follow up studies rather than the original ground breaking research. A further problem is that procedures for updating guidelines may not be standardised, even within an organisation such as NICE (McFarlane, Barnes, Sanabria, Alonso-Coello, & Alderson, 2013).

Since clinical guidelines have not previously been used for systematic assessments of groups of researchers, this article assesses their potential for this, and reports a proof-of-concept study with the UK's NICE guidelines (www.nice.org.uk). These are a systematic collection prepared by groups of experts brought together by NICE (e.g., Mayberry, Lobo, Ford, & Thomas, 2013) and, whilst they do not incorporate references, they are associated with Clinical Knowledge Summaries (CKSs), which do. Although these documents form a small collection in comparison to the National Guideline Clearinghouse (www.guideline.gov), the latter is not a single coherent collection and hence would give less clear results.

Background

A brief review of patent citation analysis can give insights into guideline citation analysis since both genres reflect transitions from theory to practice. A citation to an academic article from a patent may reflect the use of academic knowledge in an invention that is judged by the assignee to be of potential value, although citations added by examiners do not necessarily point to knowledge of the patent inventor (Alcacer & Gittelman, 2006). Patent citations are not a universal source of evidence, however, because patents are common in some areas, such as pharmaceuticals, but are rare in others, such as psychology (Coupe, 2003). Moreover, most patents do not represent commercialisation of research in the sense that patenting and technology transfer in general seem to fail most of the time (Valdivia, 2013) and because of this, citations to patents have been proposed as indicators of the value of the patents (Hall, Jaffe, & Trajtenberg, 2005; Trajtenberg, 1990). Even within a field, patent citations offer only a partial picture of technology transfer, with other types being hidden, and traditional citations may better reflect the value of academic research (Roach & Cohen, 2013). In addition, universities have in the past promoted patenting through technology transfer offices, science parks and other methods (Minguillo & Thelwall, 2012), which seems to have increased the number of patents that do not reflect genuine commercial products (Henderson, Jaffe, & Trajtenberg, 1998). In summary, whilst patent citations may seem to be useful evidence of commercial value, in practice they are probably not, except perhaps in certain narrow contexts.

One study has analysed citations to the references in clinical guidelines, finding that more highly cited references associated with higher quality guidelines (Andersen, 2013) but it did not assess whether references in guidelines could be a source of impact evidence for the scholars that published the articles. Another study showed that the references in UK cancer guidelines had a bias towards UK-based sources of evidence, tended to draw disproportionately on funded research, and revealed the geographic origins of the research

informing the guidelines (Lewison, 2007; Lewison, & Sullivan, 2008). Finally, clinical guidelines can themselves be discussed and cited in an attempt to publicise their recommendations to practitioners (Stenke & Hussey, 2014).

Research questions

It is not possible to fully evaluate the use of guideline citations for researcher evaluation until more countries systematically add references to their guidelines so this article employs two questions that can start to assess their value.

- Are articles that are cited in clinical guidelines more highly cited in the academic literature than other similar articles? A positive answer would suggest that guideline citations might not be needed for evaluating researchers because the extra value of the articles would already be reflected to some extent by traditional academic citations.
- Do articles that are cited in clinical guidelines attract (a) more Mendeley readers and (b) more Mendeley readers with an occupation classed as "Other Professional" in comparison to similar articles? A positive answer to (a) would suggest that Mendeley readership could at least partly reflect the impact of articles cited in clinical guidelines. A positive answer to (b) would suggest the same for the other professional readers' category to check because it is the only non-academic reader category and so would presumably contain all non-academic medical practitioners using Mendeley.

In the research questions academic citations are logical to check because they are widely used. Mendeley readers are also worth checking because they are a particularly promising source of altmetrics (Priem, Taraborelli, Groth, & Neylon, 2010) and the only one that categorises users into professional groups (Li, Thelwall, & Giustini, 2012; Mohammadi, Thelwall, Haustein, & Larivière, in press; Zahedi, Costas, & Wouters, 2013).

Methods

NICE CKSs were downloaded in July 2014 from the alphabetical clinical topic list on the home page of cks.nice.org.uk. To check for unlisted guidelines, Google and Bing were searched with the following three queries and their results checked against the existing list.

- `site:cks.nice.org.uk "free full-text"`
- `site:cks.nice.org.uk "references"`
- `site:cks.nice.org.uk "last revised in"`

This process resulted in a list of 327 CKSs, excluding one that was missing from the site at the time of downloading. A program was written to parse the references in the guidelines (now added to Webometric Analyst) to extract the authors, publication year, title, journal name, volume and issue from each one. Only articles published in journals were considered because these have a natural comparison set of other articles in the same journal and year and possibly also in the same volume or issue. This gave 6128 references.

To check the relative academic citation level for clinical guidelines-cited articles, each article was ranked in its journal issue using the formula $(rank-1)/(articles-1)$, which is called here the *normalised citation rank*. This gives a number between 0 and 1, with 0.5 meaning that the article has received the issue median number of citations (but not vice versa due to ties). If articles cited in clinical guidelines tend to have rank scores greater than 0.5 then this suggests that academic citations reflect their applied values to some extent. Scopus was

used for the citations in preference to the Web of Science since its wider coverage could encompass more applied research and hence more citations to the guideline-cited articles. Issues were found in Scopus for 4516 of the references (66%). The missing articles seemed to be in journals not indexed by Scopus at the time that the article was published. Moreover, in some cases the names of the journals in the references did not match the name of the journal in Scopus due to abbreviation, miss-spelling or alternative name formats (e.g., The BMJ, BMJ, The British Medical Journal).

The above was repeated using Mendeley reader counts extracted from the Mendeley Applications Programming Interface (API) by Webometric Analyst (Mendeley tab). Mendeley reports the total number of readers for an article and the percentage of readers in the top three occupation categories for the article. When one of these categories included the 'other professional' category its percentage was multiplied by the number of article readers and rounded to get the number of 'other professional' readers. When this class not was mentioned, it was assumed that there were no such readers, although there may have been a small number.

Results

The earliest cited article was from 1933 and the median age of the cited articles was 2004, with a clear skew towards relatively recent research, although the peak was seven years ago in 2007 (Figure 1). The distribution reflects the guidelines being compiled and updated at different times and so some are more recent than others. The median last updated date of the CKSs from which the references were taken was 2012, although the CKS service began in March-April 2007 (Smart, 2007) and replaced a previous computer-based service, PRODIGY, which had begun in 1998 (Wilson, Purves, & Smith, 1999). The 2007 reference peak in Figure 1 possibly reflects a degree of conservatism in the CKS references from the foundation date, although summaries on new topics have been created since then.

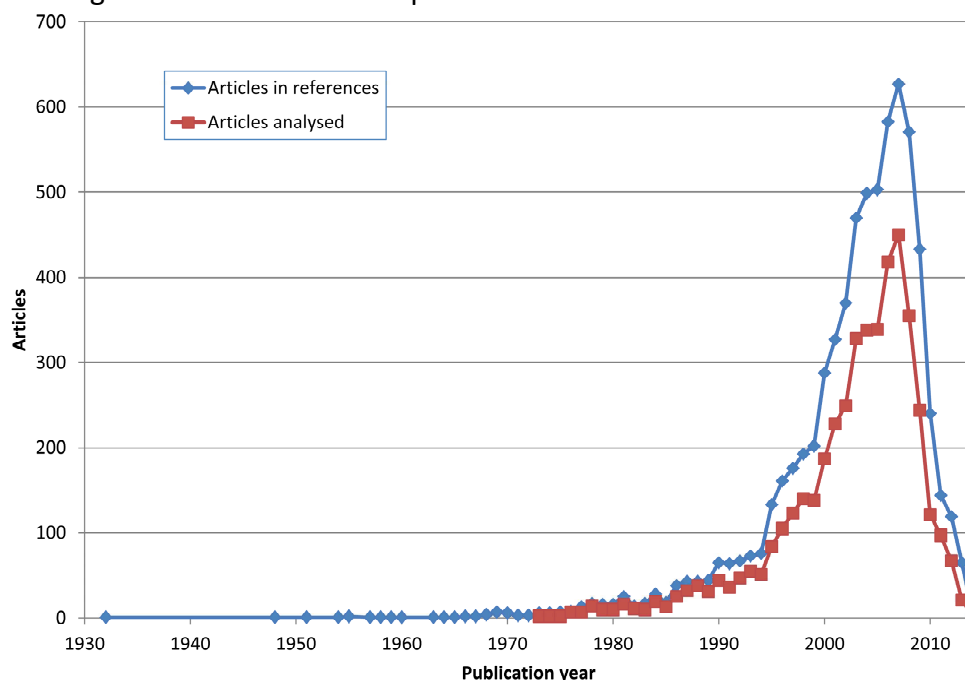


Figure 1. Journal articles extracted from NICE CKSs and journal articles analysed by year.

There is clear evidence that CKS references that are journal articles are much more highly cited than other journal articles from the same issue. For all the years combined the mean normalised citation rank is 0.80, with a 95% confidence interval of (0.80, 0.81). This citation advantage is stable over time, except for the most recent articles, which presumably will show the same pattern after more time has elapsed to attract citations (Figure 2).

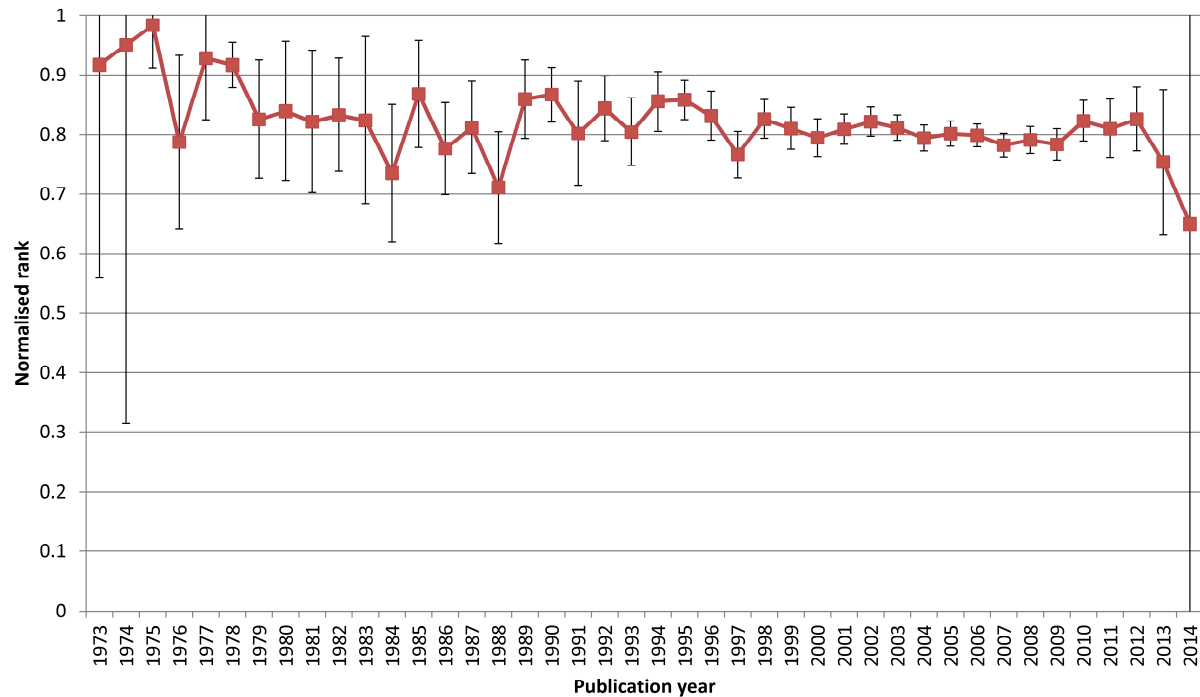


Figure 2. Mean normalised citation ranks for articles in NICE CKS references by year. Error bars indicate 95% confidence intervals for the mean.

Mendeley reader counts and Mendeley Other Professional reader counts also point to NICE CKS references being more cited than comparable articles, but their rank advantage is less pronounced (figures 3 and 4). The reason for the lower rank advantage is that there are many fewer Mendeley readers than there are citers (the overall means were 142.5 for citations, 17.1 for readers and 1.74 for Other Professionals). Mendeley readers were more numerous than citers for the most recent full year, however (in 2013 the means were 13.2 for citations, 23.62 for readers and 1.57 for Other Professionals) so they may be useful impact indicators for very recent articles (much longer time periods have been suggested for another discipline: Maflahi & Thelwall, in press).

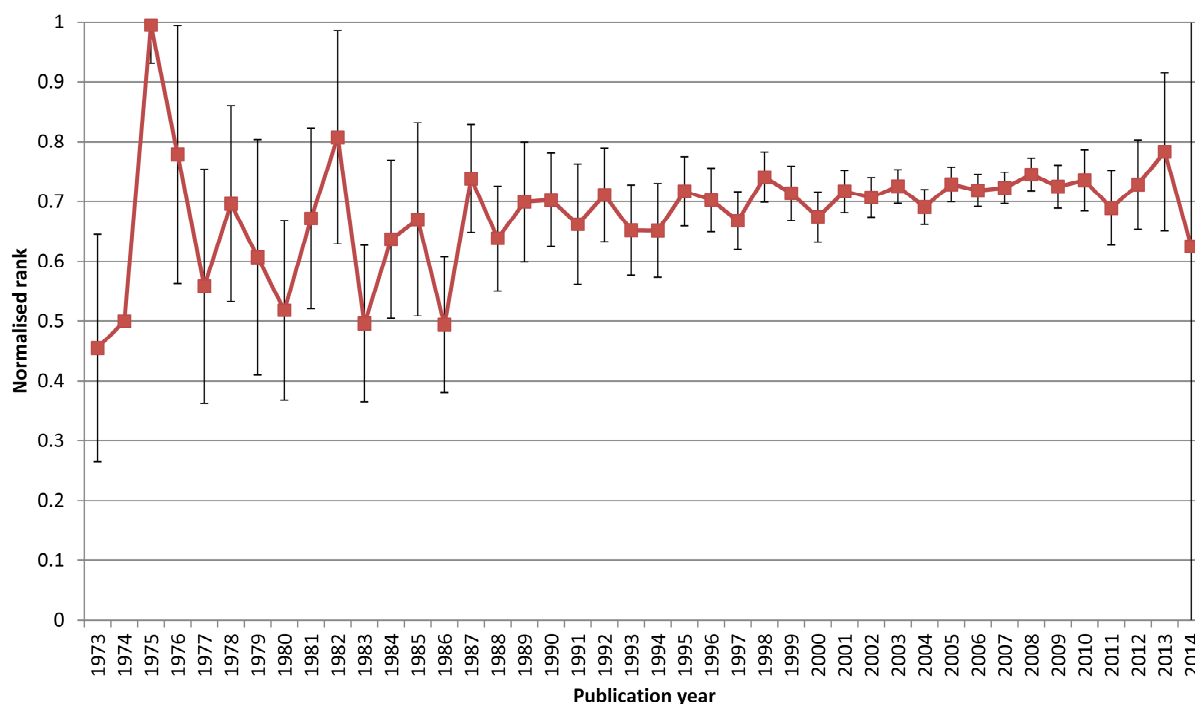


Figure 3. Mean normalised Mendeley reader ranks for articles in NICE CKS references by year. Error bars indicate 95% confidence intervals for the mean.

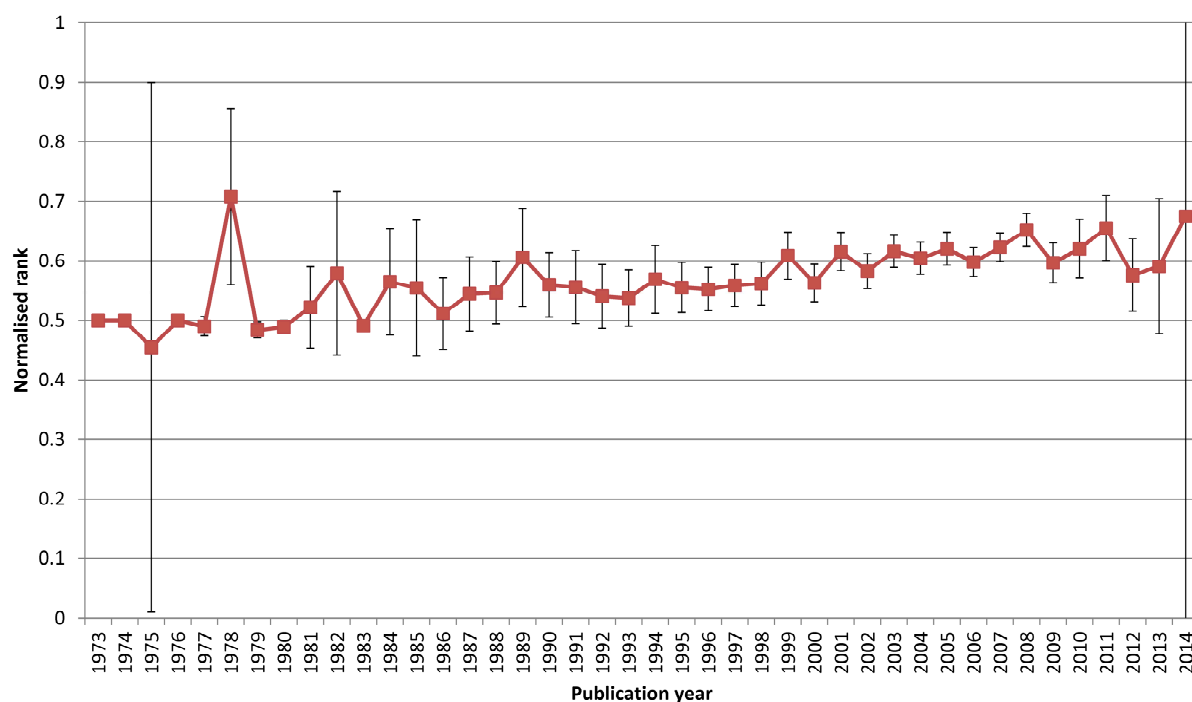


Figure 4. Mean normalised Mendeley Other Professional reader ranks for articles in NICE CKS references by year. Error bars indicate 95% confidence intervals for the mean.

Discussion

The results show that CKS references tend to be much more highly cited than comparable articles in the sense that they tend to fall within the top 20% most cited articles in their journal issue. This suggests that their applied value is reflected through increased citations

but other explanations are also possible. In particular, CKS references could tend to be published in low impact journals for their specialism, and hence are highly ranked in their issue, although this seems to be unlikely. This is hard to check directly in a fair way because there are many very low-ranked journals, including those outside of Scopus and WoS. As an indirect and partial check, Spearman correlations were calculated between the normalised citation rank of journals and the mean number of citations per article in the journal issue. This correlation was calculated separately for articles published in the peak year, 2007, and the two previous years. If a moderate proportion of CKS references were published in low impact journals then, other factors being equal, a negative correlation could be expected because articles in lower impact issues would tend to attract more citations than articles in higher impact issues. All three correlations were small and insignificant, however (2007: $\rho = -0.014$, $p = 0.773$, $n = 450$; 2006: $\rho = 0.023$, $p = 0.639$, $n = 424$; 2005: $\rho = -0.029$, $p = 0.590$, $n = 351$), suggesting that this is not a factor and so CKS references probably do not tend to be published in low impact journals for their specialism.

The results are limited to the UK NICE CKS references but it seems likely that a similar pattern would be evident for any carefully compiled set of references from clinical guidelines.

It is not clear whether the CKS references tend to be more highly cited because they are more valuable or because they have the extra publicity of being in CKS guidelines. Both may be true but, from the perspective of research evaluation, this probably does not matter. It is also possible that more highly cited articles are more likely to be added to clinical guidelines because their citations make them easier to find (e.g., more highly ranked in systems that use citations for ranking, and easier to find through the citations in other documents).

To give additional context to Figure 2, only nine articles were the single least cited in the issue in which they were published (see Table 1), but 1097 were the single most cited. Four of the nine lowest-ranked articles had attracted no Scopus citations, although all except one of these had attracted a few Google Scholar citations (13 for *The management of acute bronchitis in children*, 3 for *Accuracy of cardiovascular risk estimation* and 1 for *Hormonal breast cancer agents: implications for the primary care provider*). The articles all seem to be valuable and hence it is plausible that they would be undervalued by citation counts. Thus it is possible that *some* articles cited in clinical guidelines are undervalued by citation counts, even though it may not be common. Nevertheless, only one of the articles is a traditional primary research report and some are clearly practitioner-oriented, suggesting that such literature may be particularly undervalued by citations.

Table 1. The 9 articles with the fewest articles in their issue out of the 4993 articles for which issues were found.

Guideline	Last revised	Referenced article title [comment]	Journal	Year	Issue cit.	Issue size	Article cit.
Herpes simplex - oral	Sept. 2012	A risk-benefit evaluation of aciclovir for the treatment and prophylaxis of herpes simplex virus infections [review article]	Drug Safety 23(2)	2000	446	6	15
CVD risk assessment and management	Dec. 2008	Accuracy of cardiovascular risk estimation [letter]	Clinical Chemistry 49(4)	2003	1584	42	0
Achilles tendinopathy	April 2010	Achilles tendinopathy [practitioner-oriented review]	Critical Reviews in Physical and Rehabilitation Medicine 16(2)	2004	26	4	3
Tamoxifen - managing adverse effects	Feb. 2009	Hormonal breast cancer agents: implications for the primary care provider [review article]	Journal of the American Academy of Nurse Practitioners 18(11)	2006	26	7	0
Cough - acute with chest signs in children	Oct. 2012	The management of acute bronchitis in children [expert review article]	Expert Opinion on Pharmacotherapy 8(4)	2007	161	12	0
Halitosis	Jan. 2010	Effects of sodium lauryl sulphate (SLS), present in dentifrice, on volatile sulphur compound (VSC) formation in morning bad breath [article - randomized controlled trial]	Journal of the International Academy of Periodontology 10(4)	2008	11	3	2
Itch - widespread	Feb. 2010	Novel agents for intractable itch [informal practitioner review article]	Skin Therapy Letter 13(1)	2008	20	2	5
Meibomian cyst	May 2010	The eyelids: some common disorders seen in everyday practice [review article]	Geriatrics 64(4)	2009	20	4	1
Hypothyroidism	February 2011	Hypothyroidism in pregnancy: do guidelines alter practice? [editorial]	Thyroid 20(3)	2010	228	20	0

Conclusions

The positive results undermine the need for specific guideline metrics because, if articles are evaluated through their citations then articles in CKS references will tend to be more highly cited and hence will tend to be rewarded even without singling out their applied value. Nevertheless, it is not clear whether the reward from enhanced citations would be enough or whether an additional guideline-specific metric would also be needed. Perhaps more importantly, although the citation rank advantage of these articles is lower for articles from the previous three years (after 2011 in Figure 2 the articles have lower citation rank advantages), this may well be due to the smaller number of citations available for more recent articles, which reduces the discriminatory power of the citation rank advantage

statistic. Explicit guideline metrics may therefore be needed if recent articles are to be assessed, as is often the case.

The data from Mendeley gave a less clear indication of the value of CKS references than did the citation data. This is probably because there were many fewer readers than citers for most years, making the rank-based statistics less powerful. Although the Other Professional category in Mendeley could include medical professionals that use clinical guidelines but that do not publish research, few of these were found for the CKS references. Mendeley statistics are therefore not currently useful for assessing the professional impact or academic impact of medical guideline references, except perhaps for very recent articles.

In terms of future work, it would be useful to assess other guidelines to confirm that the pattern found is not unique to NICE and also to compare the references between guidelines to assess the extent to which the same references are used. For example, it may be that some references are unique and irreplaceable but others are more a matter of choice. Presumably the unique references would be the most highly cited. It would also be useful to attempt a finer-grained assessment of references in guidelines that give quality indicators to their references (Restrepo, 2010) to see if these can help to identify particularly important citations.

References

- Alcacer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics*, 88(4), 774-779.
- Andersen, J. P. (2013). Association between quality of clinical practice guidelines and citations given to their references. arXiv preprint arXiv:1301.5782.
- Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International Review of Psychiatry*, 23(4), 318-327.
- Coupe, T. (2003). Science is golden: Academic R&D and university patents. *The Journal of Technology Transfer*, 28(1), 31-46.
- Deader, R., Tiboni, S. G., Malone, P. S., & Fairhurst, J. (2012). Will the implementation of the 2007 National Institute for Health and Clinical Excellence (NICE) guidelines on childhood urinary tract infection (UTI) in the UK miss significant urinary tract pathology?. *BJU International*, 110(3), 454-458.
- Fleming, S. T., Mackley, H. B., Camacho, F., Seiber, E. E., Gusani, N. J., et al. (2014). Clinical, Sociodemographic, and Service Provider Determinants of Guideline Concordant Colorectal Cancer Care for Appalachian Residents. *The Journal of Rural Health*, 30(1), 27-39.
- Glover, M., Buxton, M., Guthrie, S., Hanney, S., Pollitt, A., & Grant, J. (2014). Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. *BMC Medicine*, 12(1), 99.
- Grimshaw, J. M., & Russell, I. T. (1993). Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *The Lancet*, 342(8883), 1317-1322.
- Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 36(1), 16-38.
- Haymart, M. R. (2010). The role of clinical guidelines in patient care: thyroid hormone replacement in women of reproductive age. *Thyroid*, 20(3), 301-307.

- Henderson, R., Jaffe, A. B., & Trajtenberg, M. (1998). Universities as a source of commercial technology: a detailed analysis of university patenting, 1965–1988. *Review of Economics and Statistics*, 80(1), 119-127.
- Kryl, D., Allen, L., Dolby, K., Sherbon, B., & Viney, I. (2012). Tracking the impact of research on policy and practice: investigating the feasibility of using citations in clinical guidelines for research evaluation. *BMJ Open*, 2(2), e000897. doi:10.1136/bmjopen-2012-000897
- Lewison, G., & Sullivan, R. (2008). The impact of cancer research: how publications influence UK cancer clinical guidelines. *British Journal of Cancer*, 98(12), 1944-1950.
- Lewison, G. (2007). The references on UK cancer clinical guidelines. *Proceedings of 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2007)*. Madrid: CSIC (pp.489-498).
- Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement, *Scientometrics*, 91(2), 461-471.
- 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*.
- Mayberry, J. F., Lobo, A., Ford, A. C., & Thomas, A. (2013). NICE clinical guideline (CG152): the management of Crohn's disease in adults, children and young people. *Alimentary Pharmacology & Therapeutics*, 37(2), 195-203.
- McFarlane, E., Barnes, S., Sanabria, A. J., Alonso-Coello, P., & Alderson, P. (2013). Updated recommendations: an assessment of NICE clinical guidelines. *Implementation Science*, 9, 72.
- Minguillo, D., & Thelwall, M. (2012). Mapping the network structure of science parks: An exploratory study of cross-sectoral interactions reflected on the web. *ASLIB Proceedings*, 64(4), 332-357.
- 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*.
- Ovseiko, P. V., Oancea, A., & Buchan, A. M. (2012). Assessing research impact in academic clinical medicine: a study using Research Excellence Framework pilot impact indicators. *BMC Health Services Research*, 12(1), 478.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. <http://altmetrics.org/manifesto/>
- Restrepo, R. D. (2010). AARC Clinical Practice Guidelines: from “reference-based” to “evidence-based”. *Respiratory Care*, 55(6), 787-788.
- Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science*, 59(2), 504-525.
- Shekelle, P. G., Woolf, S. H., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: developing guidelines. *BMJ*, 318(7183), 593.
- Smart, S. (2007). Clinical confidence. *Nursing Standard*, 21(29), 21-21.
- Stenke, E., & Hussey, S. (2014). Ulcerative colitis: management in adults, children and young people (NICE Clinical Guideline CG166). *Archives of Disease in Childhood-Education & Practice Edition*, edpract-2013-305512.
- Tabassum, F., & Batty, G. D. (2013). Are Current UK National Institute for Health and Clinical Excellence (NICE) Obesity Risk Guidelines Useful? Cross-Sectional Associations with

- Cardiovascular Disease Risk Factors in a Large, Representative English Population. *PLoS ONE*, 8(7), e67764.
- Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The RAND Journal of Economics*, 21(1), 172-187.
- Valdivia, W. D. (2013). University Start-Ups: Critical for Improving Technology Transfer. Center for Technology Innovation at Brookings. Washington, DC: Brookings Institution.
- Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Potential benefits, limitations, and harms of clinical guidelines. *BMJ*, 318(7182), 527-530.
- Yue, J., Tabloski, P., Dowal, S. L., Puella, M. R., Nandan, R., & Inouye, S. K. (2014). NICE to HELP: Operationalizing National Institute for Health and Clinical Excellence guidelines to improve clinical practice. *Journal of the American Geriatrics Society*, 62(4), 754-761.
- Wilson, R. G., Purves, I. N., & Smith, D. (1999). Utilisation of computerised clinical guidance in general practice consultations. *Studies in Health Technology and Informatics*, 77, 229-233.
- Zahedi, Z., Costas, R., & Wouters, P. F. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? *PLoS ALM Workshop*. <https://openaccess.leidenuniv.nl/handle/1887/23579>