

A cautionary bibliometric tale of two cities

G. E. Derrick · H. Sturk · A. S. Haynes · S. Chapman ·
W. D. Hall

Received: 24 June 2009 / Published online: 12 January 2010
© Akadémiai Kiadó, Budapest, Hungary 2010

Abstract Reliability of citation searches is a cornerstone of bibliometric research. The authors compare simultaneous search returns at two sites to demonstrate discrepancies that can occur as a result of differences in institutional subscriptions to the Web of Science and Web of Knowledge. Such discrepancies may have significant implications for the reliability of bibliometric research in general, but also for the calculation of individual and group indices used for promotion and funding decisions. The authors caution care when describing the methods used in bibliometric analysis and when evaluating researchers from different institutions. In both situations a description of the specific databases used would enable greater reliability.

Keywords Web of Science · Web of Knowledge · Institutional subscriptions · Evaluative bibliometrics

With the rise of the evaluation culture within universities, scientists are increasingly using online databases to track citations and calculate their own indices, e.g. the h-index (Hirsh 2005). Researchers take for granted that the publication information returned from a citation database search is reliable, replicable and comparable across different sites; however, this is not the case with WoK as we shall demonstrate. The following cautionary tale emphasizes the importance of researchers considering the sometimes significant variations in databases searched by the ISI Web of Knowledge (WoK) platform, depending on the subscription level of their host institution. These discrepancies may have significant

Electronic supplementary material The online version of this article (doi:[10.1007/s11192-009-0118-7](https://doi.org/10.1007/s11192-009-0118-7)) contains supplementary material, which is available to authorized users.

G. E. Derrick (✉) · A. S. Haynes · S. Chapman
Sydney School of Public Health, Edward Ford Building,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: gderrick@health.usyd.edu.au

H. Sturk · W. D. Hall
School of Population Health, University of Queensland, Brisbane, Australia

implications not only for the reliability of bibliometric research, but also for institutional block research funding and individual academic advancement decisions, both of which can be informed by bibliometric indices such as the h-index and career citation counts.

Bibliometric research rarely, if ever, specifies the name of the institution where the research was performed nor lists which databases are included in the Institute of Scientific Information subscriptions of those institutions. Similarly, researchers listing their h-index are not in the habit of qualifying their index by adding, 'as calculated from databases a-n'. Nevertheless, differences in institutional subscriptions can produce significantly different returns from searches, highlighting the need to specify the database subscription details of the institution where the bibliometric research was undertaken.

When researching the publication output of Australia-based researchers in the alcohol studies field, we noticed significant differences in data returned from identical Web of Knowledge searches conducted simultaneously at the University of Sydney (USYD) and the University of Queensland (UQ). Analyses by USYD and UQ were run concurrently at 3 p.m. on Wednesday 1 April 2009 and repeated on Tuesday 12 May 2009 at 4.30 p.m. WoK was used to search for papers published using the search string (Topic = (alcohol NOT cannabis) AND Year Published = (1999–2008) AND Address = (Australia)). The results were then refined using the ISI Analyze Results function to identify the authors who had 10 or more publications within this time period. On both occasions the searches yielded different results between the two institutions.

Supplementary table compares search returns from USYD and UQ. The USYD search returned 4,444 papers and UQ 4,125 papers (a 7.2% difference). The USYD search returned 143 authors and UQ 129 (a 9.8% difference). This difference in author returns had a significant impact on the final lists. For example, the author Pitson SM was at position 40 on the USYD list of returns with 20 publications (well above our threshold of 10); however, this author was not returned at all by the UQ search. Likewise, author Vadas MA who held position 44 on the USYD list with 19 publications did not appear in the UQ returns. To ensure that this phenomenon was not restricted to alcohol related research, we also ran the following search string (Topic = (Injury) AND Year Published = (1999–2008) AND Address = (Australia)). A similar difference was observed: the USYD search returned 8,811 papers compared to UQ's 8,315 (a 5.6% difference), and 382 authors compared to UQ's 366 authors (a 4.1% difference).

Thomson Reuters advised us that the reason for differing results was that WoK is not a distinct database but a search platform that searches a varying number of databases simultaneously, depending on an institution's ISI subscription. USYD and UQ have different subscriptions and thus different results were 'inevitable' when conducting the same search. A comparison of the different WoK subscriptions of the universities is presented in Supplementary table.

Thomson Reuters suggested that Web of Science (WoS) would deliver more consistent results. However, after repeating the search on WoS, there was still a difference in the number of publications received for the 'alcohol NOT cannabis' search: 3,017 at UQ and 2,986 at USYD (a 1% difference). This difference was caused by UQ having additional access to the Conference Proceedings Citation Indexes (Science and Social Sciences) through WoS.

The ability to specify and search selected databases is not available when using WoK, but the WoS platform does allow this. Using WoS's more controlled search environment AND knowing the subscription details of both universities allowed us to remove the additional indexes from the USYD search and to obtain the same search results for both institutions (Table 1).

Table 1 Full subscription available from Thomson ISI Web of Knowledge compared to the subscription of the University of Sydney and the University of Queensland

Database name	USYD	UQ
Backfiles		
Biological Abstracts®		
BIOSIS Previews®	Y	Y
CAB Abstracts®	Y	Y
Chinese Science Citation Database		
Current Contents Connect®	Y	Y
Derwent Innovations IndexSM		Y
Essential Science IndicatorsSM	Y	Y
FSTA-Food Science and Technology Abstracts®		Y
Global Health from CABI		
InCites™		
Inspec®		
ISIHighlycited.com		Y
Journal Citation Reports®	Y	Y
Journal Use Reports™		
MEDLINE	Y	
ResearcherID		
Scientific Webplus		
TR Links		
Web of Science®	Y	Y
Zoological Record	Y	

The question remains: to what extent are other researchers aware of the importance of determining if there are different institutional subscriptions to WoK and WoS, especially considering that the default setting in WoS is for a search to be conducted in all available databases (Jasco 2005)? We urge researchers to consider the implications. Other papers have drawn attention to miscalculations due to spelling mistakes of author, address or journal names within references (Jasco 2005; Garfield 1990; Moed and Vriens 1989; Oscalluch et al. 2009); however, few papers have mentioned discrepancies in findings between institutions. We investigated this further with a brief search of PubMed using the search term ‘h-index’ and restricting the search to original journal articles published in English. The top 10 results showed 6 papers evaluating different institutions and/or researchers using citation analysis that included calculating the h-index from WoS (Sypsa and Hatzakis 2009; Zhang 2009; Lee et al. 2009; Sorensen 2009; Fuller et al. 2009; Thompson et al. 2009). None of these papers specified their institutional database subscriptions; however, 3 papers enhanced their WoS search with Medline, Scopus and/or Google Scholar. One paper (Hagen 2008) noted a disparity in publication results between ISI and Scopus (a frequently described phenomenon) but no paper, to our knowledge, has investigated the implications of different ISI subscriptions on the calculation of individual indices such as the h-index.

As we have shown, institutional subscriptions involving different combinations of databases can cause significant differences in the number of papers returned by WoS and WoK. Therefore, caution should be exercised when conducting and reviewing bibliometric research using either WoK or WoS where the individual databases and indexes are not explicitly mentioned. As it is conventional for authors to specify which search platforms

were used in their bibliometric research e.g. WoS, WoK, Scopus or Google Scholar, we would recommend that editors require authors submitting bibliometric research to additionally specify the exact database, or combination of databases which are included in the institutional subscription to WoK or WoS used to conduct the study.

Finally, with the rise of citation metrics as a method of ranking researchers, institutions and countries, reliable and replicable results are essential. This study has shown that there is a potential for significant differences arising in the citation counts per researcher. As researchers commonly self-calculate such indices for inclusion in grant and promotional applications, such discrepancies could contribute to unfair or undeserving outcomes. One must be aware of organisational subscription differences and the implications for comparable results, and be able to normalise this before comparing h-indices (or other related indices) as this may either advantage or disadvantage a researcher and an organization, depending on where the search is conducted.

References

- Fuller, C. D., Choi, M., & Thomas, C. R. Jr. (2009). Bibliometric analysis of radiation oncology departmental scholarly publication productivity at domestic residency training institutions. *Journal of American College of Radiology*, *6*(2), 112–118.
- Garfield, E. (1990). Journal editors awaken to the impact of citation errors. How we control them at ISI. *Current Contents*, *41*, 5–13.
- Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLoS One*, *3*(12), e4021.
- Hirsh, J. (2005). An index to quantify an individual's scientific research output. *PNAS*, *102*(46), 16569–16572.
- Jacso, P. (2005). As we may search—comparison of major features of the *Web of Science*, *Scopus* and *Google Scholar* citation-based and citation-enhanced databases. *Current Science*, *89*(9), 1537–1547.
- Lee, J., Kraus, K. L., Couldwell, W. T. (2009). Use of the h index in neurosurgery. *Journal of Neurosurgery*, *111*(2), 387–392.
- Moed, H. F., & Vriens, M. (1989). Possible inaccuracies occurring in citation analysis. *Journal of Information Science*, *15*(2), 95–102.
- Osca-Lluch, J., Molla, C. C., & Ortega, M. P. (2009). Consequences of the error in bibliographical references. *Psicothema*, *21*(2), 300–303.
- Sorensen, A. A. (2009). Alzheimer's disease research: Scientific productivity and impact of the top 100 investigators in the field. *Journal of Alzheimers Disease*, *16*(3), 451–465.
- Sypsa, V., & Hatzakis, A. (2009). Assessing the impact of biomedical research in academic institutions of disparate sizes. *BMC Research Methodology*, *9*(1), 33.
- Thompson, D. F., Callen, E. C., & Nahata, M. C. (2009). Publication metrics and record of pharmacy practice chairs. *Annals of Pharmacotherapy*, *43*(2), 268–275.
- Zhang, C. T. (2009). The e-index, complementing the h-index for excess citations. *PLoS One*, *4*(5), e5429.