https://doi.org/10.1016/j.biocon.2017.06.028

Calver, M.C., Goldman, B., Hutchings, P.A. and Kingsford, R.T. (2017) Why discrepancies in searching the conservation biology literature matter. Biological Conservation, 213. pp. 19-26.

Biological Conservation

Why discrepancies in searching the conservation biology literature matter

MICHAEL C. CALVER[1]

BARRY GOLDMAN[2]

PATRICIA A. HUTCHINGS[2]

RICHARD T. KINGSFORD[3]

1 Environment and Conservation Cluster, School of Veterinary and Life Sciences, Murdoch University, Murdoch, Western Australia 6150, Australia m.calver@murdoch.edu.au

2 Australian Museum Research Institute, Australian Museum, 1 William Street, Sydney, New South Wales, 2010, Australia

3 Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, UNSW Australia, Sydney, New South Wales, 2052, Australia

Correspondence to M C Calver - Telephone: +61 9 360 2277, Fax +61 9 360 6303

Email – m.calver@murdoch.edu.au

Running title: Literature searches by conservation biologists

Word count:

ABSTRACT

Conservation biologists seek as much information as possible for evidence-based conservation actions, so they have a special concern for variations in literature retrieval. We assessed the significance for biological conservation of differences in literature retrieval across databases by comparing five simple subject searches in Scopus, Web of Science (WoS) (comparing two different subscriptions), Web of Science (Core Collection) (WosCC) (comparing two different subscriptions) and Google Scholar (GS). The efficiency of a search (the number of references retrieved by a database as a percentage of the total number retrieved across all databases) ranged from 5% to 92%. Different subscriptions to WoS and WoSCC returned different numbers of references. Additionally, we asked 114 conservation biologists which databases they used, their awareness of differing search options within databases and their awareness of different subscription options. The four most widely used databases were GS (88%), WoS (59%), WoSCC (58%) and Scopus (27%). Most respondents (≥65%) were unsure about specific features in databases, although 66% knew of the service GS Citations, and 76% agreed that GS retrieved grey literature effectively. Respondents' publication history did not influence their responses. Researchers seeking comprehensive literature reviews should consult multiple databases, with online searches using GS important for locating books, book chapters and grey literature. Comparative evaluations of publication outputs of researchers or departments are susceptible to variations in content between databases and different subscriptions of the same database, so researchers should justify the databases used and, if applicable, the subscriptions. Students value convenience over thoroughness in literature searches, so relevant education is needed.

Keywords: citation counts; cited reference search; secondary documents; grey literature

1. Introduction

Conservation biologists, managers, policy-makers, administrators and funding agencies routinely search literature databases for scientific publications on specific topics, often for meta-analyses (Barral *et al.,* 2015; Doerr *et al.,* 2015; Hall *et al*. 2016), evaluating researchers' track records (Hodge and Lacasse, 2011), tracing networks of collaboration (Liu *et al.,* 2011; Ji *et al.,* 2014), prioritising subscriptions (Garfield, 2005) and testing impacts of hypotheses on fields of study (Kumar *et al.,* 2013, 2015). Databases offer fast, cheap information retrieval and research metrics compared to searching hard copy or using peer review (Hodge and Lacasse, 2011; Buela-Casal and Zych, 2012), although there can be daunting logistic issues for large scale evaluations or hypothesis testing (d'Angelo *et al.,* 2011). Nevertheless, online literature searches and bibliometrics – quantitative evaluations of research literature – are now established firmly as tools for many disciplines, including biological conservation.

Despite this growing popularity, little attention is paid other than by bibliometric specialists to errors and idiosyncrasies in individual databases that affect data retrieval and conclusions (Leydesdorff, 2007; Franceschini *et al.,* 2016), or to difficulties in detecting grey literature (unpublished reports, internal documents and theses (Calver and King, 2000)) and books. For example, the Web of Science database only searches for terms in titles of papers published before 1990, but expands this to titles, keywords and abstracts for subsequent papers, an idiosyncrasy not found easily (Pautasso, 2014). This has led some to conclude mistakenly that literature in various fields expanded markedly since the early 1990s (e.g., Leuzinger and Hättenschwiler, 2013; Borrett *et al.,* 2014), but that is partly a simple artefact of the increased search retrieval rate of Web of Science post-1990 (Pautasso, 2014). Contrastingly, Google Scholar scans the full text of papers pre- and post-1990, delivering less bias (Pautasso, 2014), which is ironic given the scathing evaluation of Google Scholar by Jacsó (2008a). Other idiosyncrasies exist: the Scopus database has only complete citation data for papers published since 1996 (although there is a project to extend the coverage earlier that has already made extensive gains) (Elsevier, 2015); Web of Science searches are sensitive to the year range of an institution’s subscription and to the number and year range of subsidiary databases included (Thomson Reuters, 2016); and different databases may vary markedly in their literature retrieval of the same search term (Meho and Yang, 2007; Jacsó, 2005; 2011). Finally, although grey literature is a repository of vital data, it is often covered poorly in the major databases (Corlett, 2011), despite its importance in systematic reviews (Haddaway and Bayliss 2015).

Limited research on these topics in the field of biological conservation identified poor coverage of relevant regional or non-journal literature in some databases (Stergiou and Tsikliras, 2006; Calver *et al.,* 2011, 2013a,b), and incomplete research profiles for individual researchers if only one database is used for assessment (Calver *et al.,* 2013c). Assessments of subtler but potentially important topics such as variations in literature retrieval using different subscriptions or search options within the same database and conservation biologists' awareness of the limitations of different databases are yet to be made, although there are examples from other disciplines such as informetrics (Jacsó, 2006), neurology (García-Pérez, 2011), the sciences in general (Franceschini *et al.,* 2015a) and manufacturing (Franceschini *et al.,* 2015b).

We assessed the significance for biological conservation of differences in literature retrieval by comparing five simple subject searches in the widely used databases Scopus (main search and a secondary documents search), Web of Science, Web of Science (Core Collection) and Google Scholar. We predicted that: (i) the four databases would each recover unique references; (ii) given Scopus’ broader coverage of regional journals and Google Scholar’s coverage of books and book chapters, Scopus and Google Scholar would retrieve more references overall than Web of Science and Web of Science (Core Collection) when the search term involved regional rather than international literature; (iii) Scopus secondary documents and Google Scholar would retrieve grey literature and books absent from standard Scopus, Web of Science and Web of Science (Core Collection) searches; (iv) alternative subscriptions to Web of Science and Web of Science (Core Collection) would be substantially different, and (v) conclusions relating to the previous four predictions would be unchanged, irrespective of whether the search concerned a biological or sociological aspect of conservation.

Additionally, we canvassed a sample of conservation biologists to determine which databases they used, their awareness of differing search options within databases and their appreciation of the significance of different subscription options to well-known databases. We predicted that awareness of search options and the significance of different subscriptions would be low, with implications for the conclusions drawn from literature searches. Finally, we developed some recommendations as to how conservation researchers and practitioners may avoid identified biases in the future.

2. Methods

2.1 Selection of databases

We chose four widely used databases to test our predictions: Scopus, Web of Science, Web of Science (Core Collection) and Google Scholar. All except Google Scholar require a subscription. These databases are the subject of several comparative studies (e.g. Meho and Yang, 2007; Harzing and Alakangas, 2016), and are often used to evaluate researchers or fields of study, as well as undertake meta-analyses (e.g. Harzing and van der Wal, 2008; Jacsó, 2010; Côté *et al.,* 2013).

Scopus is an Elsevier database, established in 2004 (Jacsó, 2005), covering many conventional journals, trade journals (intended for trade or professional readers, often not peer-reviewed and often without an editorial board) and conference proceedings (but only full papers, not abstracts). Originally, books and book chapters (excepting those within a named series) were excluded because of the range of publishers and languages and the diversity of citation styles adopted by authors (especially when chapters in edited books were involved). However, since mid-2013 Scopus includes books from over 30 publishers (Elsevier, 2014). Scopus also offers a 'secondary documents' search for retrieving documents not included in Scopus but cited by documents in the database (see Calver *et al.,* 2013b for an application or the online tutorial at http://help.scopus.com/flare/Content/tutorials/sc_CitRefSearch.html?swfTarget=label03). The free access SCImago bibliometric site (http://www.scimagojr.com), based on Scopus data, lists 45 journals covering the topic of biological conservation.

Web of Science (WoS), known until January 2014 as Web of Knowledge (WoK), is published by Thomson Reuters. It covers books, journals and conference proceedings. Although known as a single database, it actually comprises several distinct specialist subsidiary databases, each of which can be searched individually. Institutional subscriptions vary in their inclusion of subsidiary databases and in years covered. Subsidiary databases can be searched simultaneously by selecting the 'search all databases' tab on the search page (Testa, 2006). Coverage of subsidiary databases is in a dropdown menu on the search page, sometimes accompanied by warnings if the subscription is not up to date.

Web of Science Core Collection (WoSCC) is the well-known specialist database within WoS that was called Web of Science with Conference Proceedings prior to January 2014. It covers journals, conference proceedings and books, with a bias to the sciences (Jacsó, 2011). For inclusion, publications must meet Thomson Reuters' rigorous selection criteria (Testa, 2006). Sometimes this leads to gaps in coverage of regional literature, of publications from the social sciences and humanities, and of publications lacking at least an English abstract (Harzing and van der Wal, 2008). Thomson Reuters' Journal Citation Reports, which list bibliometric data for the journals covered in WoSCC, list 49 journals in their Biodiversity Conservation category, similar to "biological conservation" used by Scopus.

Strictly, Google Scholar (GS) is a search engine, not a database (Franceschini *et al*., 2016). GS uses web-crawling algorithms to gather publication details and covers journals, books, book chapters, conference proceedings, grey literature, theses and blogs. Harzing and van der Wal (2008) see its free availability and wide searching as major advantages, while detractors highlight poor specification of the scope of GS searches and errors in data retrieval (Jacsó, 2008b; 2009; 2010). There is also the possibility of fraudulent manipulation (Labbé, 2010; López-Cózar *et al.,* 2012). Franceschini *et al.,* (2016, p. 174) concluded "that most consider GS simply as a search engine, certainly not a serious bibliometric database." We refer to GS as a database for simplicity, but acknowledge its uniqueness and academic limitations.

2.2 Selection of search terms and searching procedures

Initially, we ran four simple searches in each database, based on the following key words: "dugong" & "Australia", "waterbirds" & "Australia", "polychaetes" & "Australia", and "koala" & "Australia". The different organisms were chosen to give taxonomic and environmental diversity with conservation relevance, while "Australia" was included to assess the significance of regional literature. In Scopus, we ran a standard search but additionally examined the secondary documents (those that are not included in Scopus but are cited by documents in Scopus, see Calver *et al*., 2013b). The WoSCC and WoS databases were searched at the University of Sydney (New South Wales, Australia) as well as Murdoch University (Perth, Western Australia) to test for effects of different subscriptions. WoSCC at the University of Sydney extended back to 1900 compared to 1974 at Murdoch University; while the WoS subscription at the University of Sydney included more component databases than Murdoch University's subscription. GS searches were conducted using Publish or Perish (PoP) freeware for automating searches in GS and outputting the results in .csv files for analysis in Excel (http://www.harzing.com/pop.htm). Each PoP search returned many hits (> 10,000), so a subset of the 1,000 most highly cited was selected for comparing the outputs of the selected databases. Searches were completed between April and December 2014, with all records dated after 31st December 2013 discarded to ensure comparability of the date range of searches. We did not specify a starting date for searches because we deliberately wanted to accentuate the differences caused by differing date ranges in databases and subscriptions.

The above searches were taxonomically focused and included a regional term, so in August 2016 we complemented them with a further search for “wildlife tourism” that addressed the social context of conservation and had no regional search term. Given the later date of this search the subscription to WoS and WoSCC at Murdoch University had changed to be much closer to the University of Sydney subscription, so the “wildlife tourism” search was done only within WoS (Sydney), WoSCC (Sydney), Scopus and GS.

2.3 Survey of conservation biologists

Conservation biologists’ use of literature databases in general and awareness of features of the widely used Scopus, GS, WoS and WoSCC were assessed via an online survey (Online Appendix 1). First, respondents were asked to indicate their awareness and use of 18 databases, with opportunities to indicate others that they used. They then indicated whether 15 statements about features of Scopus, GS, WoS and WoSCC were True, False, or if they were unsure (questions, with answers, are in Online Table A1). Other questions sought demographic data about the respondents, including information on their publication history and whether or not English was their first language, both points that might influence their use of databases.

The survey ‘population’ was derived from two sources. The primary source was from the Society for Conservation Biology, Oceania Section (SCB Oceania), which includes all society members, many lapsed members and other ad hoc addresses. Biologists were notified of the survey and could respond to the questionnaire using appropriate links provided by the Society's listserver; (RTK has access as an officer of SCB Oceania). The second source was derived by searching, in November 2015, the Scopus database for all publications with a 'source title' of 'biological conservation' in the year 2014. This retrieved 362 entries, mainly to papers published in *Biological Conservation*, but also to chapters from books with 'biological conservation' in the title. We emailed an invitation to complete the survey to the 795 authors of these publications whose contact emails were provided in their publications.

2.4 Search retrieval comparisons and data analysis

2.4.1 Comparison of outputs from different searches

The search results from the four primary database sources were initially saved as Excel spread sheets. Column headings and data types were standardised, then imported into an Access® database. Detailed descriptions of the processing of these data are given in Online Appendix 2. We expressed the commonality between searches in different databases as 'efficiency', the number of references retrieved by a database for a search as a percentage of the total number retrieved.

2.4.2 Survey responses

Respondents' knowledge and use of the different databases mentioned were tabulated, and other databases they mentioned listed. GS, WoS, WoSCC and Scopus were most known and used (ranging from 27% of all responses for Scopus to 88% for GS), so we also used Generalized Linear Models (GLMs) to determine if respondents' publication histories (10 or less, 11-30, 31-50 or > 50 peer reviewed publications) and first language (English or not) predicted their use or not of GS, WoS, WoSCC and Scopus. We decided against using age as a predictor because the ages of respondents correlated significantly with the midpoints of the publication intervals (Spearman rank correlation = 0.64, $p < 0.05$). Significance values for the comparisons were set using the sequential Bonferroni correction (Quinn and Keough, 2002, p. 50), given the four databases involved.

In addition to describing responses to the 15 statements about features of Scopus, GS, WoS and WoSCC, we also used GLMs to determine if respondents' publication history and first language predicted their response to each statement. Most respondents chose one category (True, False or Unsure) for each item, complicating attempts to predict responses from publication history or first language because of small or empty cells in the response variable if the three categories were used. Therefore we used a binary dependent variable with options of Unsure or True/False combined. This assessed whether the predictors influenced respondents' confidence in assessing each item, rather than whether or not they answered correctly. Significance values for the comparisons were set using the sequential Bonferroni correction, given the 15 statements involved. All analyses used Statistica Version 7 (Statsoft, 2006).

3. Results

3.1 Outputs from different searches

3.1.1 Distinctiveness of the retrieval by the different databases

Our prediction that each of the four databases (excluding Scopus secondary documents) would retrieve numerous unique references was affirmed, with the highest efficiency observed being 92% for the "waterbirds" & "Australia" GS search. Only five of the other 27 searches had >50% efficiency (Table 1). Representing the publications visually, only 35 results were in common across the databases for waterbirds, 71 for dugongs, 67 for polychaetes, 249 for koalas and 53 for wildlife tourism (Figure 1).

3.1.2 Scopus and GS versus WoS and WoSCC

As predicted, GS retrieved more references than Web of Science and Web of Science (Core Collection) for all four searches (> 10,000 each time, with only the top 1,000 by citations shown) (Table 1, Figure 1). Even the top 1000 figure was greater than any other database for all searches except koalas, where WoS (Sydney) retrieved 1436 references. Although we expected a similar result for Scopus, Scopus did not retrieve more references than WoS and WoSCC (both Murdoch and Sydney), except for the ‘waterbirds’ search. To test if this result would change if the focus was only on the most highly cited papers, we looked at the 20 most highly cited papers for the search term "dugong" & "Australia" in GS, Scopus, WoS (Sydney) and WoS (Murdoch). GS had 16 unique entries in its top 20, Scopus 10, WoS (Sydney) 1 and WoS (Murdoch) 0. Nine of the 16 unique entries in GS were for books or book chapters, reflecting the poor coverage until recent years of the book literature in Scopus and WoS.

3.1.3 Retrieval of grey literature

The large numbers of references retrieved in each search by GS included books, book chapters, theses, reports and papers in minor journals, not covered in other databases. Similarly, references retrieved in Scopus secondary documents searches included examples of all these categories of references, as well as mis-citations of references actually included in Scopus. While secondary documents searches were not as efficient as GS in retrieving grey literature and book literature, they did broaden the range of literature retrieved. There were far more secondary documents for the koala search (920) than any other search, with the closest being wildlife tourism (510) (Table 1).

3.1.4 Effects of subscription specification

As predicted, the Sydney subscriptions to WoS and WoSCC returned more references than the Murdoch subscriptions (Table 1). This reflects the increased chronological coverage in Sydney and the increased range of subsidiary databases included in the Sydney WoS subscription. However, the differences were not marked, especially for WoSCC.

3.1.5 Duplicate returns

All databases searched suffered from the multiple listing of the same reference in at least some of the searches. For Scopus, the percentage of duplicates ranged from 0.0% to 1.7%, for GS 1.4 – 3.3%, for WoS (all locations) 2.1 – 9.1%, and for WoSCC (all locations) 0.0 – 0.8% (Table 2)..

3.1.6 Effects of search topic and inclusion of regional terms

The pattern of low overlap between the references retrieved by the different databases appeared irrespective of whether or not the search term included a regional term, or whether it was taxonomically or socially focused. The efficiency for the wildlife tourism searches (9% - 90%) was very similar to the range from the taxonomic search terms (5% - 92%) (Table 1).

3.2 Conservation biologists' survey responses

Twenty-seven respondents (24%) were from SCB (Oceania) and 87 (76%) were authors who had published in *Biological Conservation*. Response rates, defined as the number of people responding divided by the number approached (less any requests returned as undeliverable) were 6% and 12% respectively. Mean ages were similar between both groups of respondents ($t_{(112)} = 1.64$, $p = 0.86$), as were the relative proportions of men and women (Fisher exact test, $p = 0.08$). However, the authors from *Biological Conservation* were more likely to have published many papers than SCB Oceania respondents (Fisher exact test, $p = 0.01$). Given the small sample size for respondents from SCB (Oceania), all respondents were combined for further analyses.

Language may have been a problem in responses from authors whose first language was not English, so this potential bias was assessed for authors from *Biological Conservation*, using the country domain in the email address as an indication (admittedly inexact) of whether or not the first language was likely to be English. Using the categories of Australia, United States of America, United Kingdom, other English speaking countries and all non-English speaking countries combined there was no association between the country of respondents and the distribution of these countries in the survey invitations (chi-squared, 4 d.f. = 4.1, p = 0.39); responses were unrelated to country of residence (and presumably first language).

Two thirds (66.7%) of respondents were male. We did not know the gender of people invited to participate, so we cannot tell if the gender-bias in the responses simply followed the gender ratio amongst all invitees. The mean age for male respondents was 45.8 (range 27-67) and for females 40.2 (range 28-74). Respondents came from 29 countries, mainly the United States (31), Australia (21) and the United Kingdom (7). Nearly two-thirds (63%) spoke English as a first language. Publication histories were evenly spread with 30% having fewer than 11 peer-reviewed publications, 30% 11-30, 12% 31-50, and 29% over 50.

All respondents knew of GS with 88% using it often, far higher than the next most popular database (WoS, 59%). WoSCC (58%) and Scopus (27%) were the next most well known and used (Table 3). Publication history did not influence respondents' use of any of these four databases, but respondents with English as a first language were significantly more likely to use GS (Wald statistic 6.6, df = 1, p = 0.01, odds ratio 5.5). Respondents reported using 31 databases other than the ones listed in our survey (although they interpreted databases very loosely – some respondents included their colleagues or library catalogues). Researchgate (9 respondents) and PubMed (4 respondents) were the most common.

Respondents were generally unsure about the 15 questionnaire statements regarding GS, WoS, WoSCC or Scopus (65% or more unsure), except for two items regarding GS: 66% knew of the service Google Scholar Citations, and 76% agreed that GS was effective at retrieving grey literature (Table 4). Publication history did not influence respondents' surety for any statement. Respondents with English as a first language were significantly more likely to be unsure that WoSCC offered users the option of a unique ID (Wald statistic 8.6, df = 1, $p < 0.01$, odds ratio 4.3), and more likely to be unsure that WoS offered the opportunity to correct an error in the database (Wald statistic 8.4, df = 1, $p < 0.01$, odds ratio 11.2). Non-significant results are reported in Online Table A2.

4. Discussion

4.1 Searches

In common with other studies we found that Scopus, WoS, WoSCC and GS returned quite different results from the same searches (Meho and Yang, 2007; Sarkozy *et al.,* 2015; Harzing and Alakangas, 2016). The average efficiency across all databases for a search term was greatest in koalas (an endemic species) (46%). Average efficiencies were lowest for waterbirds (22%) and dugongs (28%). The outcome is likely a combination of the effects of Scopus' incomplete records prior to 1996 at the time of our searches, journal selectivity in WoSCC and inclusion of substantial grey literature in GS. Using the social search term "wildlife tourism" rather than a taxonomic term and excluding the regional search term "Australia" did not change the conclusion of different results from different databases, nor did restricting the search to the most highly cited references in each database.

Mongeon and Paul-Huis (2016) concluded that Scopus and WoS shared heavy biases to the natural and biomedical sciences, as well as engineering, and to publications in English. However, coverage still varied strongly between them. Our finding that GS retrieved a much broader range of literature than Scopus, WoS or WoSCC is also more widely supported. Meho and Yang (2007) found GS excellent for searching conference proceedings, Hilbert *et al.,* (2015) found GS retrieved references from a wider range of journals than WoS or Scopus, and Harzing and van der Wal (2008) recommended GS for searching books, book chapters, conference proceedings and publications in languages other than English. We found that the secondary documents function in Scopus returned many references in addition to a main Scopus search, so it may have a similar value to GS in locating publications outside the mainstream journal literature that nevertheless document details of biology or management important to conservation practitioners (Calver *et al*., 2011).

The increased search range of GS comes at a cost in ease of analysing search results. Adriaanse and Rensleigh (2013) reported a high degree of duplication in the output of searches in GS, while Meho and Yang (2007, p. 205) noted that the time to "clean" their literature searches took twice as long for Scopus than for WoS, while GS took 30 times as long as WoS – 100, 200 and "a grueling 3 000 hours" respectively. If Publish or Perish (PoP) is used for searching in GS, the free, web-based utility CleanPoP (Baneyx 2008) imports the comma-separated values (.csv) output from PoP and, after questioning regarding target authors and incomplete publications, deletes questionable records and combines duplicate entries. However, this approach is most suitable when the search is for an author, not a subject, and Calver *et al*. (2013) reported that some legitimate papers identified in the original PoP output may disappear after running CleanPoP. Although we did not keep records, our subjective assessment is that we also invested more time in cleaning GS files, although our problems were more with formatting than duplication. This may be because we used only the top 1000 references by citations and hence lost a long "tail" of infrequently cited or uncited references that might simply be mis-cited duplicates of more highly cited entries, or publications of questionable relevance to the search. If the aim is to find as many relevant papers as possible, then the cost of recording reference details in GS may be more than offset by the value of finding a key reference. However, in fields with a rich literature the task may be overwhelming. Secondary documents searches in Scopus are equally messy, including multiple entries for the same publication that must be identified and aggregated (Calver, 2015).

Whether the differences we found between our search results from WoS and WoSCC based on the different subscriptions in Murdoch and Sydney are substantial enough to cause concern will depend on the purpose of the study. People working collaboratively across institutions might need to be aware of the differences. We summarise strengths, weaknesses and idiosyncrasies of the four databases in Online Table A3.

4.2 Conservation biologists

The respondents to our survey used diverse techniques to locate relevant literature. We found no statistically significant association between their publication achievements and their tendency to use a particular database, nor did publication frequency predict their confidence in responding to particular statements about individual databases, so there is no evidence that differences in literature searching techniques are associated with publication success.

GS was the best-known and most widely used database, which is to be expected given GS's recognised value as a search engine (Franceschini *et al.,* 2016). Speed and convenience are the primary drivers for students using online searches (Markland, 2005), and the same may be true for conservation professionals, especially given the preference of libraries for electronic subscriptions. If searches use the freeware Publish or Perish results can be downloaded and sorted by citations to identify highly cited papers, which may be an indication of reliability (but with problems of its own, including the low ranking of recent publications and a likely bias to reviews, which often attract higher citations (Calver and Bradley 2010).

The extensive use of GS offers significant benefits and possible risks. The greatest benefit is the increased likelihood of finding regional literature or grey literature (Stergiou and Tsikliras, 2006), as well as books and book chapters (Calver *et al.,* 2013b,c). WoS has only included book citation details since 2011, covering the previous five years (Testa, 2012). Scopus decided originally not to list books and book chapters (excepting books in a named series) (Calver *et al.,* 2013b), but changed this policy (Elsevier, 2014). Scopus and WoSCC offer specialist searches ('secondary documents' function and 'cited reference search' respectively) that find books and book chapters (Calver *et al.,* 2013b, Calver, 2015), but they only find sources that have been cited by items in the respective databases. Additionally, van Dijck (2013) highlights GS's ability to find exact text within a document.

There are also significant disadvantages in using GS: it returns many hits for general subject searches and ranks on 'popularity' (based on linkages from other sites online). Users often only consider the top 10 items displayed on the first output page, whose appearance does not reflect scholarly relevance (van Dijck, 2013). Researchers can thus miss important references (Markland, 2005), with potential bias towards popular or highly cited publications (Evans, 2008;Bar-Ilan, 2008). We also were inevitably biased in selecting the top 1000 GS hits by citations for our analyses.

While our respondents knew of and used GS, they were largely unsure about whether or not GS can be manipulated fraudulently and whether or not users can request corrections. GS is susceptible to fraud. Labbé (2010) created a fictitious researcher and elevated him to high levels of citations, while López-Cózar *et al.,* (2012) raised their citation counts by placing fraudulent documents online to be detected by web-crawlers. Additionally, van Dijck (2013) refers to techniques to increase the online links or click records for specific documents to elevate their position in online searches. Users can delete an incorrect entry from their own profiles, but they cannot change the content of an individual entry in GS.

Such uncertainties were also reflected in understanding of WoS, WoSCC and Scopus. Respondents were unsure about options for unique researcher IDs, advanced literature search techniques, ranges of coverage and the opportunity to make corrections. Most of these points relate to the presentation of an individual's personal profile in the databases, with potential for career advancement (i.e. promotions or grant applications). Uncertainty about the range and period of coverage of a database can affect literature reviews for meta-analysis, or comparative studies of the research outputs of individuals or departments.

4.3 Implications

Given the varying results of searches in the different databases, we conclude that researchers seeking comprehensive reviews of the literature should consult multiple databases, not just one (Bar-Ilan, 2008, 2010; Walters, 2011; Tripathi and Garg, 2014). Online searches using GS are important to locate books, book chapters and grey literature. Subscription databases such as WoS and Scopus may be inadequate on their own. There are valuable protocols for systematic literature searches, which detail methods to find and screen literature before selection of studies for detailed examination or meta-analysis (Moher *et al.,* 2014, Stewart *et al.,* 2013; Barral *et al.,* 2015).

We are particularly concerned about database searches used for comparative evaluations of researchers or departments (Calver *et al.,* 2013b), or tracing patterns of collaboration among researchers or the historical development of fields of research (Borrett *et al.,* 2014; Ji *et al.,* 2014; Boix *et al.,* 2015). Demonstrated biases in the databases mean that it is important for researchers to document and justify the databases they use, including the subscription (i.e. WoS and WoSCC). Substantial duplicate records (over 9% in some of our searches) mandate caution when using bibliographic metrics for comparing citation rates or defining fields of interest. Universal use of a comprehensive DOI (Digital Object Identifier) system will facilitate identification and aggregation of duplicates, and alleviate many future problems in bibliometric analyses.

Finally, there are implications for education. If students value convenience over thoroughness in their literature searches and miss key papers (Markland, 2005; van Dijck, 2013), then teaching techniques for online searching should be part of the curriculum for students of biological conservation, as has been advocated more broadly

(Ettinger, 2008; Exner, 2014). Students searching in subscription databases such as WoS or Scopus will, at least, locate peer-reviewed literature where they can be confident in their findings. Online searches will find dubious sources as well as valuable ones, so skills in identifying authoritative publications are important (van Dijck, 2013).

5. Acknowledgements

We thank Chris Dickman for access to library facilities at the University of Sydney, and Cissy Ballen and Justin McCann for completing some literature searches. Murdoch University Human Research Ethics Permit 2014/157 covered the work.

6. References

Adriaanse, L.S., Rensleigh, C., 2013. Web of Science, Scopus and Google Scholar: a content comprehensiveness comparison. Electron. Libr. 31, 727-744.

Baneyx, A., 2008. "Publish or Perish" as citation metrics used to analyze scientific output in the humanities: International case studies in economics, geography, social sciences, philosophy, and history. Arch. Immunol. Ther. Exp. (Warsz) 56**,** 363-371.

Bar-Ilan, J., 2008. Which h-index? - A comparison of WoS, Scopus and Google Scholar. Scientometrics 74, 257-271.

Bar-Ilan, J., 2010. Citations to the "Introduction to informetrics" indexed by WOS, Scopus and Google Scholar. Scientometrics 82, 495-506.

Barral, M.P., Rey Benayas, J.M., Meli, P., Maceira, N.O., 2015. Quantifying the impacts of ecological restoration on biodiversity and ecosystem services in agroecosystems: A global meta-analysis. Agr. Ecosyst. Environ. 202, 223-231.

Boix, M., Montastruc, L., Azzaro-Pantel, C., Domenech, S., 2015. Optimization methods applied to the design of eco-industrial parks: A literature review. J. Clean. Prod. 87, 303-317.

Borrett, S.R., Moody, J., Edelmann, A., 2014. The rise of Network Ecology: Maps of the topic diversity and scientific collaboration. Ecol. Model. 293, 111-127.

Buela-Casal, G., Zych, I., 2012. What do the scientists think about the impact factor? Scientometrics 92, 281-292.

Calver, M., 2015. The importance of authors ensuring referencing and page proofs are correct. Pac. Conserv. Biol. 21, 173-174.

Calver, M.C. , King, D. R., 2000. Why publication matters in conservation biology. Pac. Conserv. Biol. 6, 2-8.

Calver, M.C., Lilith, M. and Dickman, C. R., 2013a. A 'perverse incentive' from bibliometrics: could National Research Assessment Exercises (NRAEs) restrict literature availability for nature conservation? Scientometrics 95, 243-255.

Calver, M.C., Fontaine, J.B., Linke, T.E., 2013b. Publication models in a changing environment: Bibliometric analysis of books and book chapters using publications by Surrey Beatty & Sons. Pac. Conserv. Biol. 19, 394-408.

Calver, M.C., Beatty, S.J., Bryant, K.A., Dickman, C.R., Ebner, B.C., Morgan, D.L., 2013c. Users beware: Implications of database errors when assessing the individual research records of ecologists and conservation biologists. Pac. Conserv. Biol. 19, 320-330.

Calver, M.C., Bradley, J.S., 2010. Patterns of citations in open access and non-open access conservation biology journal papers and book chapters. Conserv. Biol. 24, 872-880.

Calver, M., Wardell-Johnson, G., Bradley, S., Taplin, R., 2011. What makes a journal international? A case study using conservation biology journals. Scientometrics 85, 387-400.

Corlett, R., 2011. Trouble with the gray literature. Biotropica 43, 3-5.

Côté, I.M., Curtis, P.S., Rothstein, H.R., Stewart, G.B., 2013. Gathering data: Searching literature and selection criteria. In: Koricheva, J., Gurevitch, J, Mengersen, K. (Eds.), Handbook of Meta-analysis in Ecology and Evolution. Princeton University Press, Princeton, New Jersey, pp. 37-51.

D'Angelo, C. A., Giuffrida, C., Abramo, G., 2011. A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. J. Assoc.

Inf. Sci. Technol. 62, 257-269.

Doerr, E.D., Dorrough, J., Davies, M.J., Doerr, V.A.J., McIntyre, S., 2015. Maximizing the value of systematic reviews in ecology when data or resources are limited. Austral Ecol. 40, 1-11.

Elsevier, 2014. Scopus content: book expansion project update. http://blog.scopus.com/posts/scopus-content-book-expansion-project-update. Accessed April 17th, 2017.

Elsevier, 2015. Breaking the 1996 barrier: Scopus adds nearly 4 million pre-1996 articles and more than 83 million references. https://blog.scopus.com/posts/breaking-the-1996-barrier-scopus-adds-nearly-4-million-pre-1996-articles-and-more-than-83. Accessed April 17th, 2017.

Ettinger, D., 2008. The triumph of expediency: The impact of google scholar on library instruction. J. Libr. Adm. 46, 65-72.

Evans, J.A., 2008. Electronic publication and the narrowing of science and scholarship. Science 321, 395-399.

Exner, N., 2014. Research information literacy: addressing original researchers' needs. J. Acad. Librariansh. 40, 460-466.

Franceschini, F., Maisano, D., Mastrogiacomo, L., 2015a. Errors in DOI indexing by bibliometric databases. Scientometrics 102, 2181-2186.

Franceschini, F., Maisano, D., Mastrogiacomo, L., 2015b. Influence of omitted citations on the bibliometric statistics of the major manufacturing journals. Scientometrics 103, 1083-1122.

Franceschini, F., Maisano, D., Mastrogiacomo, L., 2016. The museum of errors/horrors in Scopus. J. Informetr. 10, 174-182.

García-Pérez, M.A., 2011. Strange attractors in the Web of Science database. J. Informetr. 5, 214-218.

Garfield, E., 2005. The agony and the ecstasy – the history and meaning of the journal impact factor. Paper presented at the International Congress on Peer Review and Biomedical Publication Chicago, September 16, 2005.

Haddaway, N.R., Bayliss, H.R., 2015. Shades of grey: Two forms of grey literature important for reviews in conservation. Biol. Conserv. 191, 827-829.

Hall, C. M., Bryant, K. A., Haskard, K., Major, T., Bruce, S., Calver, M. C., 2016. Factors determining the home ranges of pet cats: A meta-analysis. Biol. Conserv. 203, 313-320.

Harzing, A.W., Alakangas, S., 2016. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics 106, 787-804.

Harzing, A. W. K., van der Wal, R., 2008. Google Scholar as a new source for citation analysis. Ethics Sci. Environ. Polit. 8**,** 61-73.

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., Reinhardt, W., Stock, W.G., 2015. Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. Online Inform. Rev. 39, 255-264.

Hodge, D. R., Lacasse, J. R., 2011. Ranking disciplinary journals with the Google Scholar h-index: A new tool for constructing cases for tenure, promotion, and other professional decisions. J. Soc. Work Educ. 47, 579-596.

Jacsó, 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. Curr. Sci. 89, 1537-1547.

Jacsó, P., 2006. Deflated, inflated and phantom citation counts. Online Inform. Rev. 30, 297-309.

Jacsó, P., 2008a. Google Scholar's ghost authors. Libr. J. 134, 26-27.

Jacsó, P., 2008b. Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for F. W. Lancaster. Libr. Trends 56, 784-815.

Jacsó, P., 2009. Calculating the h-index and other bibliometric and scientometric indicators from Google Scholar with the Publish or Perish software. Online Inform. Rev. 33, 1189-1200.

Jacsó, P., 2010. Savvy searching pragmatic issues in calculating and comparing the quantity and quality of research through rating and ranking of researchers based on peer reviews and bibliometric indicators from Web of Science, Scopus and Google Scholar. Online Inform. Rev. 34, 972-982.

Jacsó, P., 2011. The h-index, h-core citation rate and the bibliometric profile of the Web of Science database in three configurations. Online Inform. Rev. 35, 821-833.

Ji, Q., Pang, X., Zhao, X., 2014. A bibliometric analysis of research on Antarctica during 1993-2012. Scientometrics 101, 1925-1939.

Kumar, L., Khormi, H.M., 2013. Landscape of ecological research in Australia: A bibliometric analysis of trends in research output and hotspots of research from 1991 to 2010. Austral Ecol. 38, 599-608.

Kumar, L., Khormi, H.M., Leis, K., Taylor, S., 2015. Ecological research in Australia: Identifying links versus gaps between hotspots of ecological research and biodiversity. Austral Ecol. 40, 581-590.

Labbé, C., 2010. Ike Antkare one of the great stars in the scientific firmament. *ISSI Newsletter* 6 (1), 48-52. Available from: http://hal.inria.fr/docs/00/71/35/64/PDF/TechReportV2.pdf. (Accessed April 17th, 2017)..

Leydesdorff, L., 2007. Caveats for the use of citation indicators in research and journal evaluations. J. Assoc. Inf. Sci. Technol. 59, 278-287.

Leuzinger, S., Hättenschwiler, S., 2013. Beyond global change: Lessons from 25 years of CO_2 research. Oecologia 171, 639-651.

Liu, X., Zhang, L., Hong, S., 2011. Global biodiversity research during 1900-2009: A bibliometric analysis. Biodivers. Cons. 20, 807-826.

López-Cózar, E., Robinson-García, N. and Torres Salinas, D., 2012. Manipulating Google

Scholar Citations and Google Scholar Metrics: simple, easy and tempting. EC3 Working Papers **6:** 29 May, 2012. Available from: http://arxiv.org/abs/1212.0638. (Accessed April 17th, 2017.).

Markland, M., 2005. Does the student's love of the search engine mean that high quality online academic resources are being missed? Perform. Meas. & Metrics 6, 19-31.

Meho, L. I., Yang, K., 2007. Impact of data sources on citations counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. J. Assoc. Inf. Sci. Technol. 58, 2105–2125.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J.A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J.J., Devereaux, P.J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P.C., Grimshaw, J., Guyatt, G., Higgins, J., Ioannidis, J.P.A., Kleijnen, J., Lang, T., Magrini, N., McNamee, D., Moja, L., Mulrow, C., Napoli, M., Oxman, A., Pham, B., Rennie, D., Sampson, M., Schulz, K.F., Shekelle, P.G., Tovey, D., Tugwell, P., 2014. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Rev. Esp. Nut. Hum. Diet. 18, 172-181.

Mongeon, P., Paul-Hus, A., 2016. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 106, 213-228.

Pautasso, M., 2014. The jump in network ecology research between 1990 and 1991 is a Web of Science artefact. Ecol. Model. 286, 11-12.

Quinn, G.R., Keough, M.J., 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge.

Sarkozy, A., Slyman, A., Wu, W., 2015. Capturing citation activity in three health sciences departments: a comparison study of Scopus and Web of Science. Med. Ref. Serv. Q. 34, 190-201.

Statsoft, Inc., 2006. STATISTICA (data analysis software system), version 7.1. www.statsoft.com.

Stergiou, K.I.S., Tsikliras, A.C., 2006. Underrepresentation of regional ecological research output by bibliometric indices. Ethics Sci. Environ. Polit. 6, 15-17.

Stewart, G.B., Côté, I.M., Curtis, P.S., Rothstein, H.R., Curtis, P.S., 2013. First steps in beginning a meta-analysis. In: Koricheva, J., Gurevitch, J, Mengersen, K. (Eds.), Handbook of Meta-analysis in Ecology and Evolution. Princeton University Press, Princeton, New Jersey, pp. 27-36.

Testa, J., 2006. The Thomson Scientific journal selection process. Int. Microbiol. 9, 135-138.

Testa, J., 2012. The book selection process for the Book Citation Index in Web of Science. http://wokinfo.com/media/pdf/BKCI-SelectionEssay_web.pdf (Accessed 28th June, 2017).

Thomson Reuters, 2016. Web of Science™ All Databases help. http://images.webofknowledge.com/images/help/WOK/hp_database.html. (Accessed April 17th, 2017).

Tripathi, H.K., Garg, K.C., 2014. Scientometrics of Indian crop science research as reflected by the coverage in Scopus, CABI and ISA databases during 2008-2010. ALIS 61, 41-48.

Van Dijck, J., 2013. Google scholar as the co-producer of scholarly knowledge. In: Takševa, T. (ed.), Social Software and the Evolution of User Expertise: Future Trends in Knowledge Creation and Dissemination. Information Science Reference (an imprint of IGI Global), Hershey PA USA 17033, pp. 130-146.

Walters, W. H., 2011. Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases. Portal 11, 972-1006.

Table 1. Total number of unique references retrieved from each database for each of five separate searches (i.e. after removal of duplicates). The numbers in parentheses represent Efficiency[4].

Data Source:	Search terms: "dugong" + "Australia"	"koala" + "Australia"	"polychaetes" + "Australia"	"waterbirds" + "Australia"	"wildlife tourism"
Total number of references[1]	1151	2105	1416	1063	1051
Number of references common to the four primary databases	71	249	67	35	53
Web of Science (Sydney)	324 (28%)	1436 (68%)	543 (38%)	165 (16%)	186 (18%)
Web of Science (Perth)	234 (20%)	1214 (58%)	457 (32%)	111 (10%)	
Web of Science Core Collection (Sydney)	140 (12%)	798 (38%)	248 (18%)	50 (5%)	97 (9%)
Web of Science Core Collection (Perth)	126 (11%)	785 (37%)	238 (17%)	49 (5%)	
Google Scholar[2]	963 (84%)	933 (44%)	967 (68%)	981 (92%)	948 (90%)
Scopus	134 (12%)	683 (32%)	232 (16%)	65 (6%)	135 (13%)
Scopus Secondary Documents[3]	54	920	113	12	510
Mean Efficiency	28%	46%	31%	22%	32%

[1] References retrieved by searches from Perth Web of Science databases were excluded from this total so as to avoid double-counting those that also came from the Sydney collection (both were based on the same search terms). Scopus Secondary Documents were also excluded.
[2] Only the top 1,000 (by citations) of hits in Google Scholar were analysed. After removal of duplicates less than 1,000 unique references remained.
[3] No attempt was made to remove duplicates from Scopus secondary documents.
[4] The 'Efficiency' of each data source varies with the different search terms and is shown by the number of references retrieved for a data source as a percentage of the total number retrieved.

Table 2. Proportion of duplicates returned measured as the number of duplicate references over the cumulative total number of references retrieved across the four taxonomic search terms and the wildlife tourism term (as in the column headings of Table 1).

Source and Search term	Number of Records	Number of Duplicates	Percentage of Duplicates
Google Scholar - "Dugong" + "Australia"	1000	30	3.0%
Google Scholar - "Koalas" + "Australia"	1000	21	2.1%
Google Scholar - "Polychaetes" + "Australia"	1000	20	2.0%
Google Scholar - "Waterbirds" + "Australia"	1000	14	1.4%
Google Scholar - "Wildlife tourism"	980	32	3.3%
Scopus - "Dugong" + "Australia"	148	2	1.4%
Scopus - "Koalas" + "Australia"	750	13	1.7%
Scopus - "Polychaetes" + "Australia"	249	0	0.0%
Scopus - "Waterbirds" + "Australia"	70	0	0.0%
Scopus - "Wildlife tourism"	135	0	0.0%
Web of Science (Perth) - "Dugong" + "Australia"	253	6	2.4%
Web of Science (Perth) - "Koalas" + "Australia"	1330	61	4.6%
Web of Science (Perth) - "Polychaetes" + "Australia"	515	16	3.1%
Web of Science (Perth) - "Waterbirds" + "Australia"	118	3	2.5%
Web of Science (Sydney) - "Dugong" + "Australia"	347	13	3.7%
Web of Science (Sydney) - "Koalas" + "Australia"	1555	102	6.6%
Web of Science (Sydney) - "Polychaetes" + "Australia"	604	55	9.1%
Web of Science (Sydney) - "Waterbirds" + "Australia"	172	6	3.5%
Web of Science - "Wildlife tourism"	190	4	2.1%
Web of Science Core Collection (Perth) - "Dugong" + "Australia"	133	1	0.8%
Web of Science Core Collection (Perth) - "Koalas" + "Australia"	843	5	0.6%
Web of Science Core Collection (Perth) - "Polychaetes" + "Australia"	251	0	0.0%
Web of Science Core Collection (Perth) - "Waterbirds" + "Australia"	52	0	0.0%
Web of Science Core Collection (Sydney) - "Dugong" + "Australia"	147	1	0.7%

Web of Science Core Collection (Sydney) - "Koalas" + "Australia"	860	4	0.5%
Web of Science Core Collection (Sydney) - "Polychaetes" + "Australia"	261	0	0.0%
Web of Science Core Collection (Sydney) - "Waterbirds" + "Australia"	53	0	0.0%
Web of Science Core Collection - "Wildlife tourism"	97	0	0.0%

Table 3. Conservation biologists' knowledge of and use of 18 databases, based on 27 respondents (24%) from SCB (Oceania) and 87 (76%) from authors who had published in *Biological Conservation* in 2014.

Answer Options	I know of this database but do not use it	I know of this database but rarely use it	I know of this database and often use it	I do not know of this database	Response Count (maximum 114)
ASFA (Aquatic Sciences and Fisheries Abstracts)	19	6	3	80	108
Biological Abstracts	39	40	7	26	112
Biosis Citation Index	31	23	2	55	111
Biosis Previews	27	18	0	65	110
CAB Abstracts	32	14	2	61	109
Chinese Science Citation Index	9	1	0	100	110
Conference Proceedings Citation Index	10	9	0	91	110
Current Contents Connect	26	16	3	65	110
Derwent Innovations Index	6	0	0	102	108
FSTA (Food Science and Technology Abstracts)	6	0	0	103	109
Google Scholar	1	13	100	0	114
Inspec	5	2	0	102	109
Medline	41	22	5	40	108
SciELO Citation Index	16	5	11	79	111
Scopus	29	36	30	17	112
Web of Science (All Databases)	12	25	66	9	112
Web of Science Core Collection	14	27	64	6	111
Zoological Record	30	28	6	45	109

Table 4. Responses to 15 statements regarding the widely used databases Scopus, Google Scholar (GS), Web of Science (WoS) and Web of Science Core Collection (WoSCC).

Statement	True	False	Unsure	Response count (maximum 114)
GS is effective at finding both scientific 'grey literature' (e.g. government reports, conference presentations, theses) and peer reviewed literature	87	9	18	114
Even if a journal is indexed in Scopus, Scopus may not include all papers published in that journal prior to 1996	25	0	89	114
WoSCC offers a range of subscriptions that vary in how far back they extend in time	38	2	73	113
WoS subscriptions always include the same component databases	9	9	95	113
GS citation data can be manipulated fraudulently	12	14	88	114
GS offers a service called Google Scholar citations	75	1	38	114
Scopus can retrieve citations to documents not in Scopus by documents that are in Scopus using a Secondary Documents Search	4	0	108	112
Scopus offers researchers a unique Researcher ID	28	4	82	113
WoSCC can retrieve citations to documents not in WoSCC by documents that are in WoSCC using a Cited Reference Search	23	4	84	114
WoSCC offers researchers a unique Researcher ID	24	3	84	114
WoS offers researchers a unique Researcher ID	26	5	81	113
GS permits users to request a correction for an incorrect entry	24	5	84	113
Scopus permits users to request a correction for an incorrect entry	15	0	97	112
WoS permits users to request a correction for an incorrect entry	13	1	98	112
WoSCC permits users to request a correction for an incorrect entry	14	1	97	112

Online Table A1. Respondents were asked to indicate whether each of these statements was True, False of if they were Unsure.

Statement	Correct answer
GS is effective at finding both scientific 'grey literature' (e.g. government reports, conference presentations, theses) and peer reviewed literature	True. GS is acknowledged as excellent in locating grey literature (Harzing and van der Wal 2008).
Even if a journal is indexed in Scopus, Scopus may not include all papers published in that journal prior to 1996	True. However, Scopus is implementing a major project to extend coverage back to 1970 (http://blog.scopus.com/posts/breaking-the-1996-barrier-scopus-adds-nearly-4-million-pre-1996-articles-and-more-than-83)
WoSCC offers a range of subscriptions that vary in how far back they extend in time	True (Jacsó, 2008).
WoS subscriptions always include the same component databases	False. WoS subscriptions can be customised to include different component databases.
GS citation data can be manipulated fraudulently	True. Labbé (2010) and López-Cózar *et al*., (2012) explain how.
GS offers a service called Google Scholar citations	True. See https://scholar.google.com/intl/en/scholar/citations.html
Scopus can retrieve citations to documents not in Scopus by documents that are in Scopus using a Secondary Documents Search	True. There is a tutorial in using Secondary Documents searches at http://help.scopus.com/flare/Content/tutorials/sc_CitRefSearch.html?swfTarget=label03.
Scopus offers researchers a unique Researcher ID	True. Scopus author IDs are explained at https://www.elsevier.com/solutions/scopus/support/authorprofile.
WoSCC can retrieve citations to documents not in WoSCC by documents that are in WoSCC using a Cited Reference Search	True. Directions are available at https://images.webofknowledge.com/WOK50B6/help/WOS/hcr_search.html.
WoSCC offers researchers a unique Researcher ID	True, this is the same for WoS. Directions are available at http://www.researcherid.com/Home.action?SID=W1ZLifVHqKrEdzjeadq&returnCode=ROUTER.Success&SrcApp=CR&Init=Yes
WoS offers researchers a unique Researcher ID	True, this is the same for WoSCC. Directions are available at http://www.researcherid.com/Home.action?SID=W1ZLifVHqKrEdzjeadq&returnCode

Statement	Correct answer
	=ROUTER.Success&SrcApp=CR&Init=Yes
GS permits users to request a correction for an incorrect entry	False, although users can delete incorrect entries from their own profiles if they have one.
Scopus permits users to request a correction for an incorrect entry	True. Follow the help links from the Scopus home search page.
WoS permits users to request a correction for an incorrect entry	True. On the home page, choose ‘Request a data change’ from the drop down menu under ‘Customer feedback and support’
WoSCC permits users to request a correction for an incorrect entry	True. The procedure is the same as for WoS.
GS is effective at finding both scientific 'grey literature' (e.g. government reports, conference presentations, theses) and peer reviewed literature	True. GS is acknowledged as excellent in locating grey literature (Harzing and van der Wal, 2008).

Online Table A 2. Results of GLMs testing whether authors' publication history or whether or not English was their first language influenced their response to specific survey questions. Results significant after Bonferroni correction are shown in **bold**.
Abbreviations: GS = Google Scholar, WoS = Web of Science, WoSCC = Web of Science Core Collection.

Question	GLM Result			
		Wald Chi-Square	d.f.	Sig.
I know of GS and use it frequently	Intercept	32.33	1	<0.01
	English as first language	6.58	1	**0.01**
	Publication history	1.15	3	0.76
I know of WoS and use it frequently	Intercept	4.26	1	0.03
	English as first language	2.65	1	0.10
	Publication history	3.64	3	0.30
I know of WoSCC and use it frequently	Intercept	2.70	1	0.10
	English as first language	0.23	1	0.63
	Publication history	7.06	3	0.07
I know of Scopus and use it frequently	Intercept	12.90	1	<0.01
	English as first language	4.52	1	0.03
	Publication history	1.83	3	0.61
GS is effective at finding both scientific 'grey literature' (e.g. government reports, conference presentations, theses) and peer reviewed literature	Intercept	33.04	1	<0.01
	English as first language	<0.01	1	0.97
	Publication history	5.00	3	0.17
Even if a journal is indexed in Scopus, Scopus may not include all papers published in that journal prior to 1996	Intercept	23.53	1	<0.01
	English as first language	2.06	1	0.15
	Publication history	1.84	3	0.60

Question		GLM Result		
		Wald Chi-Square	d.f.	Sig.
WoSCC offers a range of subscriptions that vary in how far back they extend in time	Intercept	8.09	1	<0.01
	English as first language	1.14	1	0.28
	Publication history	1.51	3	0.68
WoS subscriptions always include the same component databases	Intercept	28.22	1	<0.01
	English as first language	3.64	1	0.06
	Publication history	3.68	3	0.30
GS citation data can be manipulated fraudulently	Intercept	23.38	1	<0.01
	English as first language	1.29	1	0.26
	Publication history	3.53	3	0.32
GS offers a service called Google Scholar citations	Intercept	11.81	1	<0.01
	English as first language	4.10	1	0.04
	Publication history	3.14	3	0.37
Scopus can retrieve citations to documents not in Scopus by documents that are in Scopus using a Secondary Documents Search*	Intercept	-	-	-
	English as first language	-	-	-
	Publication history	-	-	-
Scopus offers researchers a unique Researcher ID	Intercept	15.00	1	<0.01
	English as first language	2.06	1	0.15
	Publication history	1.58	3	0.66
WoSCC can retrieve citations to documents not in WoSCC by documents that are in WoSCC using a Cited Reference Search	Intercept	20.41	1	<0.01
	English as first language	0.91	1	0.34
	Publication history	1.29	3	0.73

Question		GLM Result		
		Wald Chi-Square	d.f.	Sig.
WoSCC offers researchers a unique Researcher ID	Intercept	18.68	1	<0.01
	English as first language	8.64	1	**<0.01**
	Publication history	3.42	3	0.33
WoS offers researchers a unique Researcher ID	Intercept	13.97	1	<0.01
	English as first language	6.72	1	0.01
	Publication history	3.60	3	0.31
GS permits users to request a correction for an incorrect entry	Intercept	18.23	1	<0.01
	English as first language	1.06	1	0.30
	Publication history	1.27	3	0.74
Scopus permits users to request a correction for an incorrect entry	Intercept	29.00	1	<0.01
	English as first language	1.41	1	0.23
	Publication history	8.80	3	0.03
WoS permits users to request a correction for an incorrect entry	Intercept	26.93	1	<0.01
	English as first language	8.41	1	**<0.01**
	Publication history	9.21	3	0.03
WoS permits users to request a correction for an incorrect entry	Intercept	27.74	1	<0.01
	English as first language	7.24	1	0.01
	Publication history	11.04	3	0.01

* With 96.4% of respondents choosing ‘unsure,’ this item was not valid for analysis.

Online Table A3. Assessments of strengths and limitations for searching the scientific literature for biological conservation, based on assessments of the main databases available and current idiosyncrasies.

Database	Limitations	Strengths	Idiosyncracies
Web of Science (WoS)	A subscription database, so users need to be affiliated with a subscribing institution Acknowledged to be biased to the sciences and to publications in English	Now includes coverage of books, book chapters, and conference proceedings Includes citation searching to find literature that has cited a specific work Sophisticated advanced search features Diverse download options, including to .csv files and bibliographic software	Different subscriptions are available that vary in coverage over time and in the subsidiary databases included
Web of Science Core Collection (WoSCC)	A subscription database, so users need to be affiliated with a subscribing institution Acknowledged to be biased to the sciences and to publications in English Restricted to the most influential journals as defined by Thomson Reuters' selection criteria. Regional journals may be under-represented.	Now includes coverage of books, book chapters, and conference proceedings Provides a 'Cited Reference Search' to locate literature not included in WoSCC but cited by literature that is included Includes citation searching to find literature that has cited a specific work Includes citation searching to find literature that has cited a specific work Sophisticated advanced search features Diverse download options, including to .csv files and bibliographic software	Different subscriptions are available that vary in coverage over time.
Google Scholar (GS)	Poor specification of the scope of searches Errors in data retrieval Susceptible to fraudulent	Includes citation searching to find literature that has cited a specific work Exceptionally broad coverage of 'grey literature' including theses and government reports Strong coverage of book, book chapters	

Database	Limitations	Strengths	Idiosyncracies
	manipulation	and conference proceedings	
	Cannot download results of searches in .csv format (needs the separate freeware ‘Publish or Perish’, or Zotero (www.zotero.org) as an add-on to a browser)	Options under settings allow users to download references to four different reference manager software options	
Scopus	A subscription database, so users need to be affiliated with a subscribing institution Acknowledged to be biased to the sciences and to publications in English	Now includes coverage of books, book chapters, and conference proceedings Provides a ‘Secondary Documents Search’ to locate literature not included in Scopus but cited by literature that is included Includes citation searching to find literature that has cited a specific work Includes the Scopus journal analyser to compare the citation performance of different journals Sophisticated advanced search features Diverse download options, including to .csv files and bibliographic software	Has incomplete coverage prior to 1996, although there is currently a project to extend coverage back further (http://blog.scopus.com/posts/breaking-the-1996-barrier-scopus-adds-nearly-4-million-pre-1996-articles-and-more-than-83)

Online Appendix 1. The survey of conservation biologists

1. We, Mike Calver from Murdoch University (m.calver@murdoch.edu.au) and Richard Kingsford from the University of New South Wales (richard.kingsford@unsw.edu.au), are interested in understanding how conservation scientists, managers, policymakers and educators and other users of scientific publications find specific information relevant to their conservation problem, solution or research interest.

We have devised this short survey which will take you about 15 minutes and hope you will assist us with it, given your interest in conservation biology. If, after reading the information above, you agree to continue with the survey, choose 'yes' from the options below and proceed. Should you change your mind at any time and decide to withdraw, simply close your browser and you will automatically exit the survey. Note, though, that once you click the 'done' button at the end of the survey your responses will be uploaded and it will not be possible to withdraw or amend them because we cannot tie responses to you as an individual.

Should you wish to receive a summary of the results of the study, there is an opportunity in the survey for you to give an email address to which the summary will be sent. Email addresses will be used solely for sending a summary of the study and not to identify respondents. All responses are anonymous and confidentiality will be kept.

If you do not agree to proceed with the survey, please choose 'no' from the options below and close this window to leave the survey.

This study has been approved by the Murdoch University Human Research Ethics Committee (Approval 2014/157). If you have any reservation or complaint about the ethical conduct of this research, and wish to talk with an independent person, you may contact Murdoch University Human Research Ethics Committee (Approval 2014/157). If you have any reservation or complaint about the ethical conduct of this research, and wish to talk with an independent person, you may contact Murdoch University's Research Ethics Office (Tel. 08 9360 6677 (for overseas studies, +61 8 9360 6677) or email ethics@murdoch.edu.au). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

○ Yes

○ No

2. Please indicate which of the following databases you know of or use.

	I know of this database but do not use it	I know of this database but use it rarely	I know of this database and use it often	I do not know of this database
ASFA (Aquatic Sciences and Fisheries Abstracts)	○	○	○	○
Biological Abstracts	○	○	○	○
Biosis Citation Index	○	○	○	○
Biosis Previews	○	○	○	○
CAB Abstracts	○	○	○	○
Chinese Science Citation Index	○	○	○	○
Conference Proceedings Citation Index	○	○	○	○
Current Contents Connect	○	○	○	○
Derwent Innovations Index	○	○	○	○
FSTA (Food Science and Technology Abstracts)	○	○	○	○
Google Scholar	○	○	○	○
Inspec	○	○	○	○
Medline	○	○	○	○
ScieLO Citation Index	○	○	○	○
Scopus	○	○	○	○
Web of Science All Databases (before 2014, known as Web of Knowledge)	○	○	○	○
Web of Science Core Collection (before 2014, known as Web of Science)	○	○	○	○
Zoological Record	○	○	○	○

3. Please list any other database you use.

4. Please indicate whether each of the following statements is true or false about the searching power and coverage of the widley used databases Web of Science (All Databases, before 2014, known as Web of Knowledge) (WoS), Web of Science (Core Collection, before 2014, known as Web of Science) (WoSCC), Google Scholar (GS) and Scopus. If you are uncertain, choose 'unsure'

	True	False	Unsure
GS is effective at finding both scientific 'grey literature' (e.g. government reports, conference presentations, theses) and peer reviewed literature	○	○	○
Even if a journal is indexed in Scopus, Scopus may not include all papers published in that journal prior to 1996	○	○	○
WoSCC offers a range of subscriptions that vary in how far back they extend in time	○	○	○
WoS subscriptions always include the same component databases	○	○	○

5. Please indicate whether each of the following statements is true or false about obtaining citation data fromthe widley used databases Web of Science (All Databases, before 2014, known as Web of Knowledge) (WoS), Web of Science (Core Collection, before 2014, known as Web of Science) (WoSCC), Google Scholar (GS) and Scopus. If you are uncertain, choose 'unsure'

	True	False	Unsure
GS citation data can be manipulated fraudulently	○	○	○
GS offers a service called Google Scholar citations	○	○	○
Scopus can retrieve citations to documents not in Scopus by documents that are in Scopus using a Secondary Documents Search	○	○	○
Scopus offers researchers a unique Researcher ID	○	○	○
WoSCC can retrieve citations to documents not in WoSCC by documents that are in WoSCC using a Cited Reference Search	○	○	○
WoSCC offers researchers a unique Researcher ID	○	○	○
WoS offers researchers a unique Researcher ID	○	○	○

6. Please indicate whether each of the following statements is true or false about correcting entries in the widley used databases Web of Science (All Databases, before 2014, known as Web of Knowledge) (WoS), Web of Science (Core Collection, before 2014, known as Web of Science) (WoSCC), Google Scholar (GS) and Scopus. If you are uncertain, choose 'unsure'

	True	False	Unsure
GS permits users to request a correction for an incorrect entry	○	○	○
Scopus permits users to request a correction for an incorrect entry	○	○	○
WoS permits users to request a correction for an incorrect entry	○	○	○
WoSCC permits users to request a correction for an incorrect entry	○	○	○

7. What is your gender?

- ○ male
- ○ female

8. In what country do you currently reside?

9. Is English your first languge?

- ○ Yes
- ○ No

10. Have you published:

- ○ less than 10 peer-reviewed publications
- ○ 11 - 30 peer-reviewed publications
- ○ 31 - 50 peer-reviewed publications
- ○ greater than 50 peer-reviewed publications

11. Select one of the following options that best describes your primary role in conservation?

- ○ Conservation policy-maker
- ○ Education practitioner
- ○ Research scientist (agency)
- ○ Research scientist (non-government)
- ○ Research scientist (university)
- ○ Conservation manager
- ○ Student

Other (please specify)

[ ]

12. Thank you. The survey is complete. If you have any further comments, please give them where indicated below. If you would like a summary of the results of the survey, please enter your email address where indicated below. To exit the survey and submit your results, click the 'Done' button.

Enter any additional comments here. The box will expand as you type. [ ]

If you would like a summary of the results, enter your email address here. [ ]

Online Appendix 2. Processing of data from the database searches.

The search results from the four primary database sources were initially saved as Excel spread sheets. Column headings and data types were standardised, then imported into an Access® database where we assigned a unique record identifier to each reference, and recorded its relevant database (data source), search term and subscription locality (either Sydney or Murdoch).

There are substantial differences in the quality of the raw data from all data sources. Formatting was inconsistent between (and even within) data sources for references including: absence of trailing full stops (periods) after the title; full given names or just initials; and Roman and Arabic numerals used inconsistently. GS listed the given names/initial before the surname, thus requiring reformatting; Volume and Issue number formatting was not consistent.

Digital object identifiers (DOI) were not available in more than half the publications retrieved. This required us to generate our own surrogate DOI (which we called the NewID) to uniquely identify a piece of work. We used an algorithm that filtered and sorted the complete set of 14,113 references by Title, Author, data source, year of publication and, if present, the Volume and Issue number. References were categorized as the same (i.e. a repeat), and given the same 'NewID', where the authors (ignoring capitalization and spelling of initials), the title (ignoring punctuation), the data source, the year of publication, and the volume and issue number (when present) were the same. The second and subsequent occurrences of a reference within the same data source (there were often 3 or more) were flagged as 'duplicates' so as to exclude them from comparisons.

Formatting for GS caused problems because of many spelling and typographical errors. Ninety-three GS references (2.3%) had incomplete titles, often starting with 3 ellipses (…). These references have not been excluded from the analyses. They would not affect the absolute number of shared references but will devalue the proportional shared references when GS is involved. We were confident that after several visual passes, we correctly classified virtually all of the references and duplicates.

Standard SQL was used to generate queries to variously count or compare references unique to, or shared between the four data sources, for each of the five search terms.

Figure 1. Comparison of search results based on the search terms as per column headings in Table 1, from the four primary bibliographic data sources (WoS (Sydney), WoSCC (Sydney), Scopus and Google Scholar). The two-dimensional simplification of a multidimensional relationship shows the total number of retrieved references and indicates the degree of overlap between them.