Where you search determines what you find: the effects of bibliographic databases on systematic reviews

a. Dr Seperia B. Wanyama, College of Business and Management Sciences, Makerere University, Uganda. E: swanyama@bams.mak.ac.ug *
b. Professor Ronald W. McQuaid, Management, Work and Organization, University of Stirling, UK E: ronald.mcquaid@stir.ac.uk
c. FH- Professor Markus Kittler, Management Department, MCI - Die Unternehmerische Hochschule, Innsbruck, Austria E: markus.kittler@mci.edu

*Corresponding author

Abstract

Systematic literature reviews are common in social research for integrating and synthesising existing research. This paper argues that the outcomes of such reviews are affected by the choice of bibliographic databases. It presents evidence of substantial variation across three large electronic databases (Scopus, Web of Science and EBSCO) in a study on employee retention and staff turnover. It considers the specific articles, numbers returned, numbers shared across databases and perceived quality of journals hosting the retrieved articles. Results show that only 130 articles (5.7% of 2267 retrieved) were found common to all three databases, suggesting that decisions on how and where literature is retrieved can substantially affect the results of systematic reviews and meta-analyses. The findings caution against the use of single databases and claiming comprehensiveness. The paper reflects on how additional literature search methods (e.g., contacting experts, citation indices) and their sequence of use can affect systematic review quality.

Key words:

Systematic Literature Review; Database Search; Literature Review; Meta-analysis

This is an Accepted Manuscript version of the following article, accepted for publication in *International Journal of Social Research Methodology*. Seperia B. Wanyama, Ronald W. McQuaid & Markus Kittler (2021) Where you search determines what you find: the effects of bibliographic databases on systematic reviews, *International Journal of Social Research Methodology*, 25:3, pp. 409-422 10.1080/13645579.2021.1892378. It is deposited under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

There has been a considerable increase in the use of systematic literature reviews in social research, including Management and Human Resource Management research (e.g., Alagaraja 2012). It has been argued that these studies are an adequate means for gathering knowledge and providing a thorough overview of available evidence (Tranfield et al. 2003) often claiming comprehensiveness. While expertise, insights and the connections made by authors remain a crucial element to literature reviews, the quality of the data basis and consequently the review process itself is affected by decisions made in relation to the identification and retrieval of existing evidence. Several search methods including searching databases, checking reference lists and contacting experts in the area (Snyder 2019) have been advanced with an emphasis on following systematically documented and reported procedures that are replicable and minimise bias in the selection and inclusion of articles in a study (Mallett et al. 2012; Milne 2007; Moher et al. 2015). A prior step, often inadequately justified in the methods section of systematic reviews, is the decision of where to look (e.g. which databases to interrogate), and a clear justification of this sampling decision.

Whereas there are many literature search sources including physical libraries, contacting subject experts and reference checks, databases are often used in preference to other search methods (Green et al. 2006; Siddaway et al. 2019) due to their ability to access literature from a wide range of sources in a short time. Despite demanding rigour in the review method, the selection of specific bibliographic database(s) can be influenced by convenience and availability (such as those accessible in the researcher's organization or library) rather than being based on an informed assessment and understanding of the databases' appropriateness or quality (Reeves and Bednar 1994). Such a convenience-led choice is even more likely if researchers lack a clear understanding of what different databases have to offer and what their limits are (e.g., Mongeon

and Paul-Hus 2016). While Ciccone and Vickery (2015) found no statistical difference between the relevance of results for two specific databases, ProQuest's Summon and EBSCO Discovery Service, this conclusion does not appear to apply to other searches or databases more commonly used in management research. We argue that the approach to database choice can be expected to yield differing findings despite a systematic application of widely accepted search protocols/procedures and inclusion-exclusion criteria (Martín-Martín et al. 2018).

There are increasing demands to focus on a defined field of knowledge (Bradbury-Jones et al. 2019), and the greater orientation by databases towards particular fields (Rosenstreich and Wooliscroft 2012). However, little is known about the appropriate number of databases for an almost exhaustive search, variation or overlaps across sources, and the perceived quality of journals for articles retrieved by different major databases, especially in Human Resource Management (HRM) (Bosman et al. 2006; Falagas et al. 2008). While there is ample guidance on reporting results and emphasis is usually placed on a critical and analytical synthesis of ideas from reviewed literature, limited guidance exists in respect to the systematic and logical extraction of the readings that generate these ideas.

This paper considers results retrieved from three electronic repositories which are popular among HRM and management researchers (Scopus, Web of Science and EBSCO) for a systematic literature search concerning relatively common search terms related to employee retention strategies and turnover. The paper contributes to further developing literature review methods examining the role of database choice on search results in management research. Second, it contributes by examining the commonality and the quality of search results in a prominent research area (covering employee retention or turnover, as associated with human resource retention strategies, employee engagement, job satisfaction, organizational commitment, and turnover intention). Third, the study provides a literature review framework as a guide in actual

systematic literature review. The findings inform systematic literature reviewers in their database choice decision by advocating the use of multiple sources/databases, providing comparative insights into major databases, including their research procedures. The study emphasizes the need for additional methods such as reference checks, contacting experts, and physical searches in addition to databases to ensure rigour in the review method.

The next section discusses the process of systematic literature reviews. This is followed by an explanation of the databases and methods used. Findings of the systematic review process are then presented and discussed. We conclude our paper with a reflection on key limitations and recommendations for conducting future systematic reviews.

Systematic literature reviews and search databases

Literature reviews are a common means in academia to establish foundations for studies by providing an account and evaluation of different theories and arguments, theoretical development and relevant empirical evidence; combining findings of different studies and reviewing different methods; and refining the research problem, research questions and hypotheses (Baumeister 2013). For these reasons, emphasis is not only on the process of reviewing and writing, but also on searching, obtaining and determining the kinds of literature selected, and the extraction of key ideas for synthesis and comprehensive analysis. However, there is less understanding of search processes and their likely influence on results returned for inclusion in the review.

Both the search process, and approach to the review, determine its broad type (narrative or systematic). Narrative literature reviews are commonly used in empirical articles and dissertations as a basis for defining research questions and hypotheses. They provide insights

into theory, previous findings and methodologies (Baumeister 2013). However, many do not provide an account of the literature search and selection process, and hence, any likely bias (Green et al. 2006). Second, focus has shifted towards greater use of systematic literature reviews, with more rigorous methods that seek to ensure objectivity, transparency and replicability in article search and selection. Systematic literature reviews emphasize documenting and reporting every step that is taken (e.g., Moher et al. 2015), ensuring that the researchers, within their means, try to access all significant articles on the topic. Each step of the research has clearly defined criteria: in the inclusion or exclusion of articles for review and in the synthesis of existing knowledge.

As with traditional narrative reviews, narrative systematic reviews provide qualitative analysis of different previous studies including the theories, assumptions, arguments, methodologies, results, and conclusions of several objectively (seeking to minimise bias) selected articles that are appraised and synthesized into a comprehensive analytical paper. It has an advantage of dealing with several studies that may have used different methodologies. Meta-analyses on the other hand, use available empirical studies as primary data to consider issues such as ‘how large an effect something has’ or ‘to what extent does something change’, and review and mathematically combine or synthesize results of different studies that used comparable methods to address the research question or hypothesis (Baumeister 2013; Nielsen et al. 2017).

Systematic literature reviews generally emphasize accumulating a relatively complete census of relevant literature that is later examined for its appropriateness for inclusion into the analysis (based on clear eligibility and selection criteria) to produce a more comprehensive study. However, they may still miss a number of relevant articles, despite the rationale being to accumulate as much literature on a topic as possible to the point of data saturation, i.e. when

further search ceases to yield anything substantially different (Saunders et al. 2018; Webster and Watson 2002).

While there are many methods for literature search and retrieval, including physical searches and reference checks among others, electronic databases are now the main literature search method as they are considered more convenient and efficient. For instance, Morris et al. (2009) demonstrate how databases provide an interface that allows easier location and access to articles as compared to systematically working through laborious paper indexes and different journal series. In spite of technological malfunctions at times, the speed with which the search can be conducted fits with frequently observed resource constraints and researcher efficiency needs, which often accompany increased pressures to publish. While it is relatively easy for the researcher to systematically organize the results that databases return, only a fraction of the output may be relevant to the specific research topic or question. While the level of redundant papers can depend on the choice of keywords and search strings, it increases with the number of databases used, which might make a researcher reluctant to use multiple sources. This can be an issue in management research where topics under investigation tend to be broad and not as precisely defined as in other areas, such as medicine (Bradbury-Jones et al. 2019).

Our concerns over the rigorous selection and use of databases during a systematic review is shared within the research literature, which points to a set of caveats regarding: the number of databases selected for a systematic search (Green et al. 2006); the appropriateness of the selected database(s) to a subject area of research (Rosenstreich and Wooliscroft 2012); the ability to produce relevant results (Gough et al. 2017); the ability to produce common results with other databases; and the quality of returned results in terms of their journal ratings (Bosman et al. 2006). The degree of uniformity of the search parameters used by each database when applying search protocols is also unclear.

Different studies have used a varying number of databases without clear justification for such decisions, although reviewers often recommend searching more than one (Daigneault et al. 2014). However, the number used seems to be merely convention rather than the result of an examination of the variation across those repositories. Webster and Watson (2002) recognize that the search processes will not yield a comprehensive and complete census of literature and consequently suggest that experts should also be used, in order to identify further critical papers or other research sources. The degree to which critical evidence is omitted from an apparently rigorous systematic review procedure can undermine follow-up steps in reviewing the evidence, such as contacting experts, reference checks (Milne 2007), and 'hard copy' searches. Crucially for the current paper, the potential implications of choosing two or three databases in management research have not been fully analyzed empirically.

The choice of databases should be based on their relevance and appropriateness to the topic area, although these vary by subject area, research objectives (Daigneault et al. 2014; Phelps and Campbell 2012) and the specificity of the topic being searched. Consequently, the level of variation and overlap of results across different databases remain opaque as little research has assessed variation across bibliographic databases. For the field of marketing, Rosenstreich and Wooliscroft (2012) show how EBSCO, SSCI, JCR and Scopus are not representative of the full field of management, and that these databases also disproportionately cover US-based journals in terms of international comparisons, with only weak coverage of publications from other regions. Web of Science, on the other hand, retrieved more journal articles than Scopus when using search key terms in the social sciences (Bosman et al. 2006), although Scopus had broader overall coverage than it in terms of journals and number of documents.

The journal quality of the papers retrieved from databases has received even less attention in systematic review literatures and practice. Some reviews use perceived journal quality as a filter

for meeting minimum “quality” requirements (Palmatier et al. 2018), usually reflecting different measures such as impact factor and journal quality ranking (Morris et al. 2009; Rosenstreich and Wooliscroft 2012). While Scopus was rated higher than Web of Science in terms of covering more journals and publications, Web of Science was considered as holding more journals in the top-quality segment (Bosman et al. 2006). This is supported by Falagas et al.’s (2008) finding that Scopus had more non-academic articles, who further report that Google Scholar sometimes offers results of inconsistent accuracy or quality. Hardly any evidence of such analysis in the field of HRM has been established despite the increasing use of systematic literature reviews.

The search procedures and algorithms of the different search databases vary, affecting their ability to retrieve and generate relevant results, and resulting in variations based on the precise topic or terms being searched. For example, searches can be performed based on abstract, topic/title, author, all fields and results can be limited to and refined by double-blind reviewed journal articles, books and book chapters, and the year of publication (the year chosen on the basis of theoretical, or sometimes pragmatic, reasons) (Keupp et al. 2012). Understanding the search structure and processes of the different databases can add to the understanding of variations in results and their implications.

In addition to a thorough understanding of the databases as a gateway to the evidence available, obtaining adequate results requires judgment beyond a clear definition of search terms and rules when sifting through them in order to select the most relevant work (Gough et al. 2017). Care is needed when refining search keywords so as not to narrow them too much and hence exclude relevant studies (Falagas et al. 2008). In addition, using the number of returns following basic keyword searches is of limited help in choosing a database, particularly if it retrieves irrelevant work. Identifying applicable search keywords or strings is in itself a challenging task that is achieved after several search trials and adjustments following resultant returns. Acknowledging

that database searches are a central element in systematic review procedures, it appears reasonable that adequate knowledge of search databases is a prerequisite for their selection (beyond a mere convenience choice) and should be an essential stage in conducting systematic reviews.

Extracting the required literature from a large set of selected readings can be a daunting exercise that requires a logical and systematic approach for a comprehensive and comparative analysis of the readings for key ideas along specific parameters such as theory, method, findings, key arguments, limitations and recommendations. A framework setting out a guide on the required aspects for extraction and summary for eventual synthesis and analysis should support researchers in identifying and examining key readings to acquire knowledge.

Databases and method

We examined variations in results across databases stemming from research on employee retention and staff turnover. Building on preliminary research, we generated a list of search keywords and strings guided by commonly observed themes on our study topic (de Menezes and Kelliher 2011; Tranfield et al. 2003). The choice of databases was informed by relevancy to the subject area of business management and HRM (Rosenstreich and Wooliscroft 2012), perceived quality, and consistence of results (Bosman et al. 2006; Falagas et al. 2008), all fitting with the objective of this paper. We systematically searched for relevant publications using three databases: Scopus, Web of Science and EBSCO (used through a University research gateway that accesses other databases through EBSCO, e.g. Business Source Premier, Science Direct, JSTOR, so it could be considered a group of databases). Scopus, operated by a major academic publisher, argues that it is the largest abstract and citation database of peer-reviewed literature (including articles, books and conference proceedings); Web of Science claims to be the world's

largest publisher-neutral citation index (searching article titles, abstracts etc.); and EBSCO searches the full-text of peer-reviewed journals. These databases are not exhaustive of all those available, but are commonly used through academic libraries, and focus on peer reviewed (and hence good quality) outputs.

In line with Randolph's (2009) suggestion for pilot testing, we refined our search strings over time with different searches and discussions (see Appendix 1 for the search words/strings used). Depending on the structure of a particular database, the searches were based on document, title, abstract and keywords (Petticrew and Robertst 2006). The entire process was documented for each search and any refinements made were recorded. For example, we kept the chosen years open, refined for peer-reviewed articles, and included articles in press at the time of the search. The results were retrieved and categorized in a three-step process (Table 1).

Table 1. Search results at different stages of screening

Step/Action	**Searched databases**					**Refinement criteria/ Limiters**
Displayed Results per search per database ↓	Search string	WoS	EBSCO	Scopus	**Total**	Title, abstract, key words
	1	1488	70046	1090	**72624**	
	2	773	1488	516	**2777**	
	3	2428	6605	1332	**10365**	
	4	1329	3560	1021	**5910**	
	5	83	138	46	**267**	
	6	294	407	94	**795**	
	Total	**6395**	**82244**	**4099**	**92738**	
Refine results per search string for each database ↓	Search string	WoS	EBSCO	Scopus	**Total**	Peer review journals, academic journals, articles, relevance of the journal field/subject
	1	626	993	978	**2597**	
	2	465	362	483	**1310**	
	3	1388	1110	1268	**3766**	
	4	655	634	1013	**2302**	
	5	26	23	39	**88**	
	6	75	57	86	**218**	
	Total	**3235**	**3179**	**3867**	**10281**	
Selected and saved Records ↓	Search string	WoS	EBSCO	Scopus	**Total**	Full article, relevance to the study subject areas/ themes
	1	275	180	347	**802**	
	2	330	227	327	**884**	

	3	756	605	638	**1999**
	4	395	334	360	**1089**
	5	19	2	9	**30**
	6	32	13	23	**68**
	Total	**1807**	**1361**	**1704**	**4872**
Screening Results from each database		WoS	EBSCO	Scopus	**Total**
		1049	905	1025	**2979**
Clean results from the 3 databases		2979	Less	712	**2267**

Using RefWorks & manual scrutiny, 730 duplicates & anonymous records removed

Unique records coded as 1-7 categories* (1 record per publication)

**See figure 1 and table 2 for the distribution of the final records and Appendix 1 for search words and strings.*
Note: WoS = Web of Science; EBSCO accessed through the university access.

In the first step all papers matching the search protocol were retrieved and an 'Overall Total' folder created containing the combined references from all searches across all databases (using subfolders for different databases). As shown in Table 1, the initial number of selected studies (4,872) was reduced to 2,979 articles through the deletion of duplicates *within* each database and further condensed to 2,267 unique articles. Each article was allocated uniquely to one of the categories 1-7 (comprising results found only in a single database, or in two of the databases or in all the three, see below).

In the second step, the perceived quality of the retrieved articles was assessed by journal impact factors and rankings using the SC Imago journal ranking list 2010 (based on quantitative measures), the ABS journal list 2010 and the ERA list 2010 (both combining quantitative measures and expert opinion to achieve the rankings) (Morris et al. 2009). The choice of the lists was based on availability and uniformity in the period of assessment, and for consistent comparison across the three journal ranking lists at the time. The journal titles were matched against the different 'quality' lists (using journal serial numbers) to retrieve the journal rankings.

The third step was to use other frequently used search methods including the use of reference checks, hand and internet search, and expert networks (Webster and Watson 2002), to identify

significant contributions to the topic and compare them to the results of the databases. Checking references of identified articles was conducted to establish common and seminal papers, while some identified authorities were contacted for recommendation of some key articles on the topic, beyond the database searches. These processes identified some seminal papers, e.g. Mathieu and Zajac (1990), that were not retrieved in any of the database searches.

Additionally, we examined the search and profile pages of each of the databases to compare their search features. Key features such as indexing and archiving period, searching parameters, subject areas, document types, and indexed search databases that affect the output of each database were examined and their potential impact on results are discussed in the remainder of this paper. Lastly, we considered key areas of research topic focus in order to develop a literature review framework.

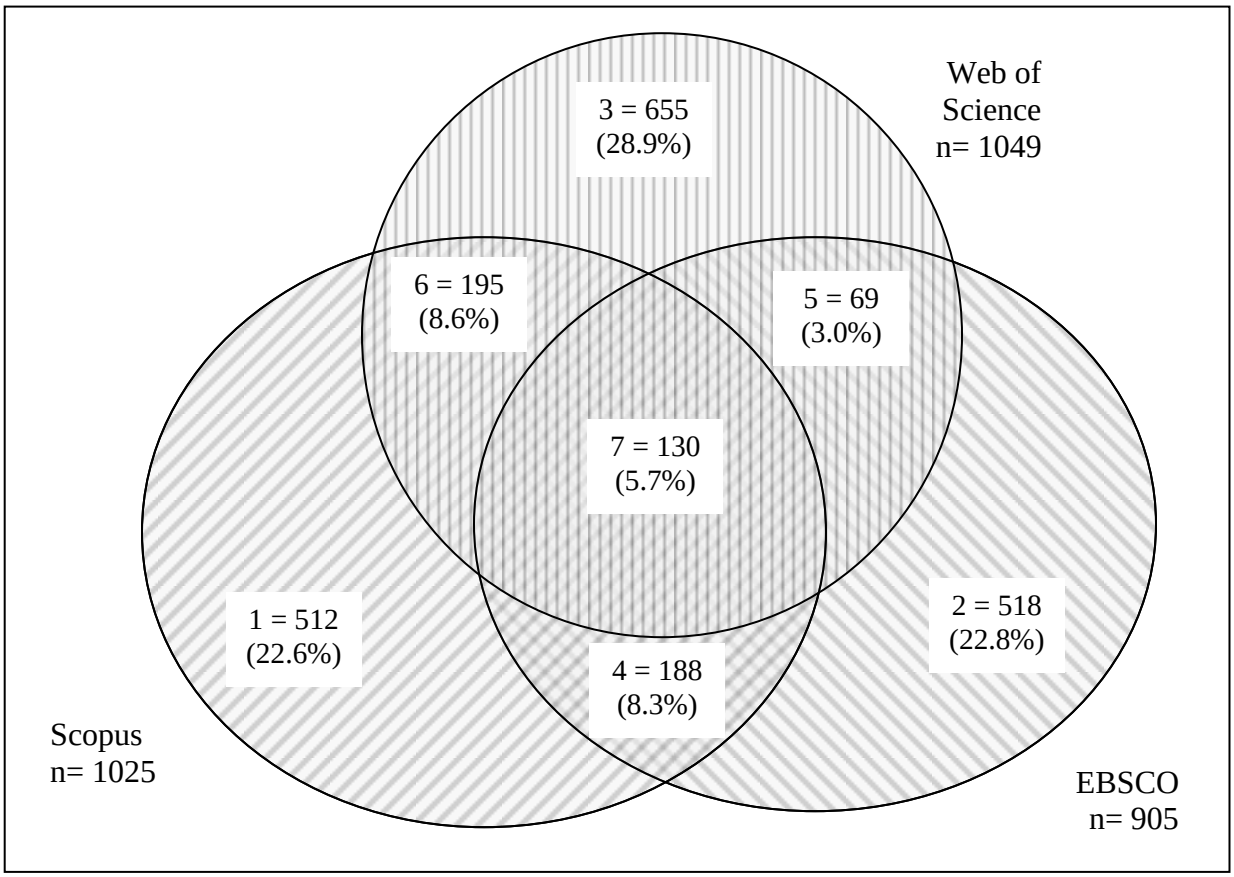
Findings and discussion

Our study provides insights into the number and quality of studies retrieved within each database and the overlaps in results across databases. It also provides an analysis of database search structures and processes that seem to account for much of the variation in the search and the results retrieved. Finally, the framework for summarizing the findings from the selected and reviewed readings can guide researchers in extracting and summarizing key points for analysis and synthesis. First, in terms of identifying differences in the quantity of results returned, the study reveals substantial differences between the different databases. We initially retrieved 4,872 articles across all three databases, which were identified as being related to the themes and area of study based on predetermined search procedures.

In the first step all papers matching the search protocol were retrieved and after deleting duplicates within each database (subfolders), we used codes to identify the source of each record

in terms of its host database(s) (1= Scopus only, 2= EBSCO only, 3= Web of Science only, 4= Scopus and EBSCO, 5= EBSCO and Web of Science, 6= Scopus and Web of Science and, 7= found in all the three databases). We were therefore able to organize records as appearing in only *one*, *two* or all *three* of the databases (see Figure 1).

Figure 1. Distribution of all results across the 3 databases



For instance, the results illustrate that only 130 unique articles (5.7%) appear in all the three databases (category 7), 452 articles (20.0%) are found in *two* of the three databases (categories 4-6,) and 1,685 articles (74.3%) appear in *only one* of the three (categories 1-3), signaling limited sharing of journal articles. So only a quarter (25.7%, n=582) of the unique articles appeared in more than one database. Hence, the findings indicate that the decision on which and how many bibliographic databases form the basis of a literature review is likely to have significant impacts

on the results. With very few articles (5.7%) being common to all the three databases, and almost three-quarters of them being unique to only one database, it means that databases largely host and as such, retrieve predominantly unique results. Searching more than two databases is therefore, we submit, an essential component of a rigorously conducted systematic review, particularly if aiming to ideally identify most of the relevant literature.

Variations in perceived quality of the retrieved articles across the databases were found. A total of 1,926 records out of 2,267 (85.0%) retrieved had an impact factor on the SC Imago list. The number of ranked records, the total and average impact factors following our grouping of articles by databases (1-7), are shown in Table 2.

Table 2. Average (mean) Impact Factor (Science Imago) of database results

Database(s)	Ranked Records	Total Impact factor	Average impact factor (per article)	Unranked Records	Total Records	Mean impact including unranked as 0	Difference
1 - Scopus Only	453	401.2	0.89	59	512	0.78	0.11
2 - EBSCO Only	360	745.8	2.07	158	518	1.44	0.63
3 - WoS Only	554	1086	1.96	101	655	1.66	0.30
4 - Scopus & EBSCO	176	192.3	1.09	12	188	1.02	0.07
5 - EBSCO & WoS	68	179.9	2.65	1	69	2.61	0.04
6 - Scopus + WoS	189	292.9	1.55	6	195	1.50	0.05
7 - Scopus + EBSCO + WoS	126	229.3	1.82	4	130	1.76	0.06
Total	**1926**	**3127**	**12.02**	**341**	**2267**	**1.38**	**10.64**

Note: WoS = Web of Science; EBSCO accessed through the university access.

Considering results for each database (both unique and shared results), it is evident that Web of Science hosted the most highly ranked results portfolio (mean impact factor = 1.91, based on a total impact of 1787.8 over 937 articles) closely followed by EBSCO (1.85, based on total impact of 1347.3 over 730 articles), while Scopus had a significantly lower mean impact score (1.18, based on total impact of 1115.8 over 944 articles). Whilst the degree to which the ranking

of an outlet, such as a Journal, can inform about the quality of individual articles is limited, and subject to an ongoing debate, we do not speculate about individual paper quality, but note that at least based on the journal rankings, there is an aggregate variation across the databases regarding perceived outlet quality[1].

Considering the records common across the different databases, the mean impact factor of the articles included in all the three databases (category 7) was 1.82 (Table 2). However, while this was expected to contain the highest-ranking articles, it was actually slightly lower than the average impact factor for categories 5 (EBSCO and Web of Science, 2.65), 2 (EBSCO only, 2.07), and 3 (Web of Science only, 1.96), reflecting the downward weight of Scopus. The influence of the lower impact factor of Scopus is evident for the different database combinations. These findings are also consistent when other perceived journal quality measures are used - the ABS and ERA journal rankings.

Table 3. Average (mean) Impact Factor (ABS and ERA) of database results

	No. per Rank					**% per Rank**				
	1	**2**	**3**	**4**	***Total***	**1**	**2**	**3**	**4**	***Total***
Scopus (ABS)	82	143	130	157	***512***	16%	28%	25%	31%	***100%***
EBSCO (ABS)	65	117	144	213	***539***	12%	22%	27%	40%	***100%***
WoS (ABS)	44	161	251	288	***744***	6%	22%	34%	39%	***100%***
Scopus (ERA)	252	224	198	130	***804***	31%	28%	25%	16%	***100%***
EBSCO (ERA)	180	168	186	163	***697***	26%	24%	27%	23%	***100%***
WoS (ERA)	145	218	347	175	***885***	16%	25%	39%	20%	***100%***

Note: Totals may not sum to 100% due to rounding, WoS= Web of Science.

Table 3 shows number and percentage of records per journal rank (1-4) based on ABS and ERA journal ranking, for each of the three databases. Consistent with the results in table 2, Web of

[1] While taking the mean assumes a consistency between the points on the impact factor scales, median measures produce similar results.

Science reported more of journal ranked articles than EBSCO and Scopus. Moreover, articles returned by Web of Science were mainly in highly ranked journals (rank 3&4= 73% and 59% for ABS and ERA respectively). These findings may be of particular interest to scholars who base their systematic reviews on perceived journal quality as a basis for the quality of the article reviewed under the notion of quality evidence-based research (Keupp et al. 2012). It also means that particular databases tend to host more articles from highly rated journals in comparison to others.

An examination of the parameters and conditions underlying the search procedures of the three databases also illustrates considerable differences that affect the outcomes of the searches across the databases. The indexing time span varies for each of the databases. While the Web of Science core collection search spanned from 1900 to date, Scopus included publications from around the 1960s, and with citation analysis information being available only for articles published after 1996 (Falagas et al. 2008; Bosman et al. 2006). EBSCO did not specifically provide information on the indexing time coverage on its search page, but the results display page shows that the results can be refined to stretch from as far back as 1401 to date. The varying periods of indexing might add to the explanation of different numbers of papers retrieved from the three databases, although the overlap of results shows that this does not explain the full differences in results.

All three databases cover various subject areas including Business Management, Social Science, Social Work, and Arts and Humanities; although Web of Science and EBSCO provide options for limiting the searches to specific subject areas. The results display page suggests that Scopus covered the Natural Sciences more comprehensively than the Social Sciences. This implies that the magnitude of the retrieved results varies by database depending on the subject area of the research, and the options for refining the search. Additionally, Scopus allowed the researcher to refine results by year, author name, subject area, document type e.g., article, review, or

conference paper. EBSCO allowed limiting the search to full article, and peer reviewed documents, while source type can be restricted to one or more of academic journals, reports, news, magazines and trade publications. Web of Science allowed the refining of results by highly cited papers in the field, year, web of science categories (disciplines), and document type.

Finally, the 'search by' parameters showed additional variation across the three databases. While it was optional with EBSCO to select the focus of the search in the document for the defined search words, searching by topic, and by article title, abstract, and key words, were the default search options for Web of Science and Scopus. However, whereas Scopus had an option for searching in 'all fields' in addition to authors, article title, abstract, and key words among others, EBSCO allowed searching all text, but with other search options such as author, title, and abstract operating independently. It appears ironic that, while Scopus and EBSCO provided options for a broader search of the specified words through all the fields, and all texts respectively, Web of Science's broadest search focus was by topic or title. Such variations (time span, subject areas, document type, and search target for the key words) mean differing findings, and hence affect the numbers and the quality of the returned journal articles as discussed above.

To shed light on the usefulness of relying on the database search, we also carried out other frequently used search methods: our findings suggest that the credibility of (systematic) literature review results is enhanced if the database search is complemented by mechanisms such as paper-based/physical searches, reference checks, internet searching and expert networks aimed at identifying significant or seminal contributions which may not be identified despite a rigorous application within the review procedure. This multi-method approach to the literature search revealed that some seminal works (e.g., Mathieu and Zajac 1990; Meyer and Allen 1991; Schaufeli et al. 2002) relate to the thematic scope of our study, exemplified by organizational commitment and employee retention being omitted from all three databases despite the rigorous

application of systematic review proceedings. The omission of seminal work confirms earlier concerns about the shortcomings of single method approaches when identifying key literature (e.g., McManus et al. 1998), and thus, our study not only supports the use of more than one or two databases, but also supports a multi-method approach for providing a relatively comprehensive overview of relevant literature focused on a particular research question.

Systematic reviews suggest a useful objective and transparent process, following explicit, pre-determined and consistent search protocols (Randolph 2009), which may aid greater replicability in studies. The benefits of systematic reviews are advocated in the literature (Baumeister 2013), and assume the potential to deliver a (near) comprehensive picture of existing evidence on a specific review question (see Briner and Denyer, 2012, for examples on narrowing down review questions). This is subject to some method-related challenges ranging from the identification of adequate search terms matching a review aim and scope, to the development and meticulous implementation of tactics for the inclusion and exclusion of studies for review (Phelps and Campbell 2012). The current study suggests that databases, while presenting a basis for well-organized and objective/non-biased searches, can generate results that are likely to vary in terms of number, uniqueness and quality of returned search articles per database. Our observations are in line with the findings in a recent analysis of differences in database coverage comparing Scopus and Web of Science (Mongeon and Paul-Hus 2016), but ours goes further, considering three databases and including perceived quality criteria. We also observed that, even when using three prominent databases, some seminal works identified using additional means of identifying literature were missed.

While systematic review literature advocates exhaustive literature searches (Tranfield et al. 2003) in order to collate all relevant evidence, research contexts are often characterized by finite personal resources and limited access to databases or research budgets. This suggests a

compromise between comprehensiveness, relevance and efficiency when initiating research. To communicate such a compromise, we recommend that scholars should clarify their retrieval strategy in the protocol of literature reviews and keep a record (memo) of their protocol and decisions relating to choices they make. A clear account of review decisions and procedures will help ensure transparency by informing the reader of the context within which the review was conducted. In the pre-review stage, decisions that compromise comprehensiveness need to be considered explicitly. Deciding which and how many databases to use, and their characteristics, are central questions with, as our review illustrates, considerable consequences for the results to be expected.

Lastly, the objectivity, fairness, rigour, and robustness emphasized in systematic literature reviews should not only be illustrated and reflected in the search process, but also in the extraction of ideas from the selected sources. A logical and generally comprehensive approach, with consistency and uniformity in extracting and summarizing key points of interest, is of great value. A literature review recording framework may be useful for considering and systematically recording, for instance, each article's reference information, the key aims of literature, their focus in terms of aims or research questions, the context, sector, theory used, methods, findings, key arguments, limitations and recommendations. A clear framework can help improve rigour, simplicity, consistency, comprehensiveness, and fairness in considering different readings.

Conclusion

This paper focuses on database choice as an important decision when conducting systematic literature reviews and meta-analyses and scoping papers in HRM and management research. Our findings on the overlap in results from three major electronic databases (Scopus, Web of Science and EBSCO) show that search results vary considerably across databases in terms of the number

of articles identified, overlap across databases and their perceived quality. Only 5.7% of the results were common to all the three databases chosen and only a quarter (24.7%) of the records were common to two or more databases. Three-quarters of the results were found in only a single database which suggests the use of supplementary tools for identifying and retrieving literature and relevant evaluation of research. Each database has different purposes, procedures and rationales and so the limited overlap of search results between them may reflect these. The results do not evaluate the effectiveness of the search engines in achieving their objectives, but rather indicate the importance of researchers understanding the focus, coverage and methods of each database so as to ensure that it is appropriate for their particular topic and field of study.

While our study is based on the application of systematic search procedures, some limitations have to be noted. The variations in results across databases and (missing) overlaps observed in this study cannot simply be generalized as they are subject to decisions on the scope and methods chosen for a specific review topic. In this paper, the context of our search focused on a broad area, employee retention strategies and turnover within a low-income economy context, including the intermediate variables of employee engagement, job satisfaction, organizational commitment and turnover intention. Nevertheless, we assume that our observation of limited overlap might also occur with other research foci. While we were explicitly looking for journal articles, different results may occur if researchers decide to widen the scope of the search including books, chapters or ‘grey’ literature or to narrow the topic to a more precise area. An additional limitation is the use of journal quality as a proxy for the quality of the results retrieved. However, over such a large sample we assume some overlap between journal and article quality.

Our findings have significant implications for researchers, information specialists, libraries and practitioners using research, as well as the bibliographic database providers themselves. Drawn

from observations regarding the literature search, our results suggest that searches focusing on only one database seem to be limited in providing robust results. As the results between databases tend to vary in quantity and uniqueness, as well as in perceived quality (based on journal rankings) this research cautions against the use of single databases for systematic reviews and meta-analyses. The findings of this study suggest the use of two or more databases with their selection based on a clear understanding of their contents and fit with the rationale for the review itself. This would suggest that systematic reviews (Moher et al., 2015) should not only describe the information sources used, but also explicitly present the rationale for their inclusion (or exclusion) within the context that the review takes place. As comprehensiveness (Denyer and Tranfield, 2006) and data saturation are limited even with the use of two or more databases, our findings suggest the use of additional methods for identifying literature (such as contacting experts, reference checks and checking for citation indices). Such additional methods can help to identify omitted seminal articles and new unpublished thinking on topics, and enrich the review beyond the results stemming from database searches only.

Whereas systematic reviews are expected to provide clear and replicable results, linking research across domains and reaching outside the specified codes, their discussion and result sections could still be open to creativity and insight beyond the systematic approach itself (with findings outside the replicable search protocol clearly labelled as such). We acknowledge that investigative research often goes together with reaching outside the specific search codes to reduce the danger of missing key or seminal work as shown above. Yet, within the review procedure, the search mechanisms used, and the underlying sampling procedure, must be treated with care to ensure rigour in method and the validity of insights generated from such reviews. The expression ‘caveat emptor’ (or ‘let the buyer beware’) may well apply to both the selection of databases that institutions choose to offer and those individual databases that scholars decide

to use. Our findings illustrate the path dependency of insights stemming from the choice of a database or a combination of databases, suggesting the use of additional search (or sampling) techniques. Careful consideration of the differences in the search profiles and search procedures such as indexing time, search fields, and subject areas of the databases, is required to avoid potentially excess biased results. For the consistent application of search protocols, it would be useful to have information on the relative uniformity of common database profiles across such fields. Reporting such data could help database selection in future review projects.

While the paper mainly contributes to the literature on search processes, it also contributes to methods for selecting articles for inclusion in a review. For example, systematic reviews should explicitly consider the number of databases a work is included in and the perceived quality of work and/or outlet it is published in, together with citation indices and recommendations by established experts. We suggest that inclusion criteria for the final selection of articles are integrated into the methods of the search strategy when forming eligibility criteria and making decisions about what data will be sought. First, considering the commonality of papers across databases may help in assessing the data retrieved for meta-biases. Availability through different search platforms may increase the likelihood of article readership and hence could affect citations. Second, journal rankings may be considered but with caution; although imperfect they are often perceived to infer the quality of the journal and its articles. Third, reference checks and non-database searches, and fourth, recommendations by experts, can inform the researcher about seminal articles on a particular subject, that could have been missed through database searches alone. While our research results indicate that care must be taken with claims about comprehensiveness in systematic literature reviews, the paper suggests additional ways to enhance data saturation and identifies the need for systematic reviews to include explicit information on, and consideration of, databases and search engines used in literature reviews.

References

Alagaraja, M. (2012) HRD and HRM Perspectives on Organizational Performance: A Review of Literature. *Human Resource Development Review,* 12(2), pp. 117–143.

Baumeister, R.F. (2013) Writing a literature review. *The Portable Mentor.* Heidelberg: Springer, pp. 119-132.

Bosman, J., van Mourik, I., Rasch, M., Sieverts, E. and Verhoeff, H. (2006) Scopus reviewed and compared. *The Coverage and Functionality of the Citation Database Scopus, Including Comparisons with Web of Science and Google Scholar. Utrecht University Library*, 73.

Bradbury-Jones, C., Breckenridge, J.P., Clark, M.T., Herber, O.R., Jones, C. and Taylor, J. (2019) Advancing the science of literature reviewing in social research: the focused mapping review and synthesis, *International Journal of Social Research Methodology*, 22(5), pp. 451-462.

Briner, R.B. and Denyer, D. (2012) Systematic review and evidence synthesis as a practice and scholarship tool. *Handbook of Evidence-Based Management: Companies, Classrooms and Research,* pp. 112-129.

Ciccone, K., and Vickery, J. (2015) A comparison of search performance using user queries. *Evidence Based Library and Information Practice,* 10(1), pp. 34-49.

Daigneault, P., Jacob, S. and Ouimet, M. (2014) Using systematic review methods within a Ph.D. dissertation in political science: challenges and lessons learned from practice. *International Journal of Social Research Methodology*, 17(3), pp. 267-283.

de Menezes, L.M. and Kelliher, C. (2011) Flexible Working and Performance: A Systematic Review of the Evidence for a Business Case. *International Journal of Management Reviews,* 13(4), pp. 452-474.

Denyer, D., & Tranfield, D. (2006). Using qualitative research synthesis to build an actionable knowledge base. *Management Decision,* 44(2), pp.213-227.

Falagas, M.E., Pitsouni, E.I., Malietzis, G.A. and Pappas, G. (2008) Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 22(2), pp. 338-342.

Gough, D., Oliver, S. and Thomas, J. (2017) *An introduction to systematic reviews.* 2 ed. London: Sage.

Green, B.N., Johnson, C.D. and Adams, A. (2006) Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of Chiropractic Medicine,* 5(3), pp. 101-117.

Keupp, M.M., Palmié, M. and Gassmann, O. (2012) The Strategic Management of Innovation: A Systematic Review and Paths for Future Research. *International Journal of Management Reviews,* 14(4), pp. 367-390.

Mallett, R., Hagen-Zanker, J., Slater, R. and Duvendack, M. (2012) The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness,* 4(3), pp. 445-455.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., and López-Cózar, E.D. (2018) A systematic comparison of citations in 252 subject categories, J*ournal of Informetrics*, 12(4), pp. 1160-1177.

Mathieu, J.E. and Zajac, D.M. (1990) A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin,* 108(2), pp. 171.

McManus, R.J., Wilson, S., Delaney, B.C., Fitzmaurice, D.A., Hyde, C.J., Tobias, R.S., Jowett, S. and Hobbs, F.D. (1998) Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *BMJ (Clinical Research Ed.),* 317(7172), pp. 1562-1563.

Meyer, J.P. and Allen, N.J. (1991) A three-component conceptualization of organizational commitment. *Human Resource Management Review,* 1(1), pp. 61-89.

Milne, D. (2007) Review of Systematic reviews in the social sciences: A practical guide. *Psychology Learning & Teaching,* 6(1), pp. 58-59.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Stewart, L.A. et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews,* 4(1), pp. 1-9.

Mongeon, P. and Paul-Hus, A. (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics,* 106(1), pp. 213-228.

Morris, H., Harvey, C. and Kelly, A. (2009) Journal rankings and the ABS journal quality guide. *Management Decision,* 47(9), pp. 1441-1451.

Nielsen, K., Nielsen, M.B., Ogbonnaya, C., Känsälä, M., Saari, E. and Isaksson, K. (2017) Workplace resources to improve both employee well-being and performance: A systematic review and meta-analysis, *Work & Stress,* 31(2), pp. 101-120.

Palmatier, R. W., Houston, M. B., and Hulland, J. (2018). Review articles: Purpose, process, and structure, *Journal of the Academy of Marketing Science*, 46, pp. 1–5.

Petticrew, M. and Roberts, H. (2006) *Systematic reviews in the social sciences. A practical guide.* London: Blackwell.

Phelps, S.F. and Campbell, N. (2012) Systematic reviews in theory and practice for library and information studies. *Library and Information Research,* 36(112), pp. 6-15.

Randolph, J.J. (2009) A guide to writing the dissertation literature review. *Practical Assessment, Research & Evaluation,* 14(13), pp. 1-13.

Reeves, C.A. and Bednar, D.A. (1994) Defining quality: alternatives and implications. *Academy of Management Review,* 19(3), pp. 419-445.

Rosenstreich, D. and Wooliscroft, B. (2012) Assessing international journal impact: the case of marketing. *European Business Review,* 24(1), pp. 58-87.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H. and Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & Quantity*, 52(4), pp. 1893-1907.

Schaufeli, W.B., Salanova, M., González-Romá, V. and Bakker, A.B. (2002) The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3(1), pp. 71-92.

Siddaway, A.P. Wood, A.M. and Hedges, L.V. (2019) How to Do a SystematicReview: A Best Practice Guidefor Conducting and ReportingNarrative Reviews,Meta-Analyses, andMeta-Syntheses. *Annual Review of Psychology*, 70(1), pp. 747-770.

Snyder, H, (2019) Literature review as a research methodology: An overview and guidelines, *Journal of Business Research*, 104, pp. 333–339

Tranfield, D., Denyer, D. and Smart, P. (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), pp. 207-222.

Webster, J. and Watson, R.T. (2002) Analyzing the Past to Prepare for the Future: Writing a Literature Review. *Management Information Systems Quarterly*, 26(2), pp. xiii-xxiii.

Appendix 1: Search words and strings used (with inclusion and exclusion terms)

1. HR*or*"HumanResourc*"*or*Staff*or*worker**or*Employ*ANDStrateg**or*framework*or*"frame work"*or*Theor**or*Approach*ANDCommitment*or*Identif**or*retention*or*Turnover*or*Engag ement*or*stay**or*Leav*

2. HR*or*“Human Resourc*”*or*“Human Resourc* Management”*or*HRM AND Turnover*or*Retention*or*Leav**or*Stay*ANDCommitment*or*Engagement*or*Identif*

3. Staff*or*Employ**or*Worker*ANDTurnover*or*Retention*or*Leav**or*Stay*ANDCommitment

4. Strateg**or*Practic**or*Approach*ANDRetention*or*Turnover*or*Leav**or*Stay*ANDCommit ment*or*Engagement **NOT** patient*or*customer**or*consum**or*student**or*patient**or*bed**or*plant*or*blood*or*veg**or*chemi **or*drug**or*chain*or*biolog**or*nutri**or*climate*or*purchase**or*stigma**or*footboll*or*ecology*or* DNA*or*Therapy*or*Volunteer**or*Union**or*Food*or*Soil

5. Public*or*"Civil Service"orGovernmentorPrivateANDUganda*or*Africaor"Developing World"*or*DevelopingCountr*ANDRetention*or*Turnover*or*Leav**or*Stay*ANDCommitme ntorEngagementorIdentif* **NOT** patient*or*customer**or*consum**or*student**or*patient**or*bed**or*plant*or*bloodorveg*orchemi **or*drug**or*chain*or*biolog**or*nutri**or*climate*or*purchase**or*stigma**or*footboll*or*ecology*or* DNA*or*Therapy*or*Volunteer**or*Union**or*Food*or*Soil

6. "Developing World"*or*Developing Countr**or*Africa*or*UgandaANDRetention*or*Turnover*or*Leav**or*Stay*ANDCommitmento *r*Engagement*or*Identif* **NOT** patient*or*customer**or*consum**or*student**or*patient**or*bed**or*plant*or*blood*or*veg**or*chemi **or*drug**or*chain*or*biolog**or*nutri**or*climate*or*purchase**or*stigma**or*footboll*or*ecology*or* DNA*or*Therapy*or*Volunteer**or*Union**or*Food*or*Soil