Can the impact of grey literature be assessed? An investigation of UK government publications cited by articles and books

Matthew S. Bickley[1], Kayvan Kousha[2] and Michael Thelwall[3]

[1] *M.Bickley@wlv.ac.uk*

[2] *K.Kousha@wlv.ac.uk*

[3] *M.Thelwall@wlv.ac.uk*

[1,2,3] Statistical Cybermetrics Research Group (SCRG), University of Wolverhampton, Wulfruna Street, Wolverhampton, WV1 1LY (United Kingdom)

Abstract

Grey literature encompasses a range of relatively informal textual outputs that are not indexed in citation databases. Although they are usually ignored in research evaluations, it is important to develop methods to assess their impact so that their contributions can be recognised, and successful types of grey literature can be encouraged. This article investigates the extent to which 97,150 UK government publications were cited by Scopus articles and Google Books during 2013-2017 in eleven broad subject areas. A method was used to semi-automatically extract citations to the UK government publications from articles and books with high recall and precision. The results showed that Scopus citations are more common than Google Books citations to UK government publications, especially for older documents, and for those in Healthcare, Education and Science. Since the difference is not huge, both may provide useful grey literature impact data.

Introduction

'Grey Literature' or 'Gray Literature' is a term which describes textual documents that are not published in a standard academic format, such as a book or journal article. The term includes reports, regulations, and policy documents, which are important outputs from many governments and organisations. Fuzzy for many years and still not concrete due to the boundaries between grey literature and non-grey literature varying depending on the situation (IGLWG 1995), the Prague definition of 2010 seems to be now accepted: "*Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body*" (Schöpfel, 2010, p.11). The US Interagency Gray Literature Working Group has given the following alternative definition: "*Foreign or domestic open source material that usually is available through specialized channels and may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents*" (IGLWG, 1995). Hence, grey literature publications can include, but are not limited to, unpublished research, governmental reports, policy statements conference proceedings, and theses or dissertations (GreyNet, 2019, UNE, 2019).

There are many high-profile grey literature repositories, confirming that this is an important document type. The UK government publication repository includes almost 120,000 annual reports, regulations, statistics, or policy documents in different topics (https://www.gov.uk/government/publications). This is a specialised source of grey literature in government policy making. The repository hosts many policy-making papers, such as healthcare reports, which are of high value to society and can be used to improve information on risk factors and how healthcare research is used (Institute of Medicine, 2009).

Other grey literature repositories include those of the World Health Organization (WHO, https://www.who.int/publications/en/), the United Nations (https://digitallibrary.un.org) and the World Bank (http://www.worldbank.org/en/research/brief/publications). Given that large amounts of grey literature have been created by governments and other important organisations, it would be useful to know if they have an impact so that their creators can decide which types of document are worth producing. This article focuses the academic impact as a first step towards this goal.

Citation analysis is commonly used to assess scientific impact of published research. However, there seems to be no practical or standard method to identify grey literature citations. Grey literature publications do not have well-established, centralised and standardised sources, and hence impact indicators are more difficult to calculate.

Google Scholar has been suggested as a good source for monitoring the impact of grey literature (Orduna-Malea, Martín-Martín & López-Cózar, 2017) and dissertations (Kousha & Thelwall, submitted). However, Google Scholar queries cannot be automated on a large scale, except for the facilities of Publish or Perish (Harzing, 2010) and it is therefore not suitable for large scale grey literature evaluations. Web queries have also been proposed for small sets of documents (Wilkinson, Sud, & Thelwall, 2014), but these do not necessarily reflect academic impact.

Given the lack of an accepted solution for determining the academic impact of grey literature, this article proposes and demonstrates two new approaches. First, Scopus (API) cited reference searches can be used to find citations to non-standard academic outputs (Kousha, Thelwall, & Rezaie, 2011) and complex queries can be designed to identify citations to large numbers of documents. Second, the Google Books API can also be used to automatically identify citations to monographs with high accuracy (Kousha & Thelwall, 2015). These strategies are proposed and are important to determine if feasible for grey literature. This paper describes the two new methods in detail and compares their results for 97,150 UK government publications from 2013-2017 across eleven broad subject areas.

Research questions

The underlying goal is to assess if Scopus and Google Books citation searches can be automated for capturing citations to grey literature publications. UK government publications are the focus of the study because the UK government publishes a large number free online, its repository can be crawled, and the authors are familiar with the UK context.

1. Can academic citations to grey literature publications be automatically extracted from Scopus and Google Books on a large scale?
2. Which citation search strategy or indicator is most useful for the impact assessment of UK government publications?
3. Are there disciplinary differences in the answer to the above question?

Methodology

This section describes how the new method was developed through small scale pilot studies.

Data sets

The online repository of documents released by the UK government (held at https://www.gov.uk/government/publications, hereafter: 'the repository') is classified by government-defined policy area and year of release (see Table 3 in the online Appendix (https://figshare.com/s/51a8308bdf43772820b3). This data was collected in July 2018 by a bespoke crawl routine added to the free Webometric Analyst (lexiurl.wlv.ac.uk) software. Each policy area was combined into more general topic areas (Table 1). The most recent five years were chosen to be most relevant for use in this method due to the increase in uploads to

the repository at that time. Out of 137,559 documents available, 97,150 (70.6%) are from the years 2013-2017. Each document has a unique URL as well as a title. The URLs were used in subsequent searches to identify citations.

Table 1. All 11 grey literature areas used by combining policy areas as defined in the repository, split by years used, along with total over 2013-2017 (grey and policy areas sorted by largest size).

Grey literature area	*Policy areas merged*	*2013-2017*	*2013*	*2014*	*2015*	*2016*	*2017*
Economics	Business and enterprise; UK economy; Tax and revenue; Employment; Trade and investment; Financial services	21112	2346	5373	4287	4155	4951
Government	Government efficiency, transparency and accountability; Local government; Government spending; Regulation reform; Media and communications	11399	1618	2987	2343	2005	2446
Environment	Environment; Food and farming; Climate change; Wildlife and animal welfare; Rural and countryside	10997	1591	2557	2378	2175	2296
Security	Crime and policing; Law and the justice system; Defence and armed forces; Public safety and emergencies; National security	9729	1096	2308	2030	2028	2267
Housing and travel	Transport; Housing; Planning and building	8995	1281	2028	1766	1574	2346
Healthcare	National Health Service; Public health; Social care	8836	892	1535	1910	2156	2343
International affairs	Borders and immigration; Foreign affairs; International aid and development; Wales; Northern Ireland; Scotland; Europe	8376	1129	1494	2115	1759	1879
Society	Community and society; Children and young people; Welfare; Equality, rights and citizenship; Pensions and ageing society; Consumer rights and issues	6823	994	1500	1604	1381	1344
Education	Schools; Further education and skills; Higher education	6045	597	1171	1295	1497	1485
Science	Energy; Science and innovation	4134	653	979	901	818	783
Leisure	Arts and culture; Sports and leisure	704	141	117	158	116	172
Total		97150	12338	22049	20787	19664	22312

Scopus API citation searches

To find citations to one or more URLs from documents indexed by Scopus, a query of the following form can be used in either the Advanced Search interface or submitted to the Scopus API:

REF("[*search term*]") OR REF("[*search term*]") OR REF("[*search term*]")…

The result is a set of journal articles, magazines, conference papers or books indexed by Scopus that contain a citation in their reference section that matches any [*search term*]. Grey literature titles were not effective as search terms because they were often too short. For example, the UK government report, "Ahead of the curve" has a subtitle of "How UK motorsport technology and innovation can benefit your company". Due to the subtitle not being part of the title, almost exclusively false matches were found in Scopus (1288) when using the article title as the [*search term*]. In comparison, only one match was found when using the URL and omitting https://www., here, REF("gov.uk/government/publications/ahead-of-the-curve"), and this was a correct match. This strategy was not perfect because some URLs can be contained within longer URLs and documents could be cited by title without an URL. Nevertheless, the method can identify citations with high precision. These queries were submitted via the Scopus API to automatically gather the results.

For Scopus API to search the database, a text file for each grey literature area in each year was created. The file contained each query term, listed one per line. Each query included the "REF" part as above, as it is still required to search only the reference sections within Scopus. To match Google Books searches (discussed below), queries without the leading part (www.gov.uk/government/) were used. An example of such a query, using the example above, is:

REF("publications/ahead-of-the-curve")

The list of queries was then input into Scopus API search which automates the search process. In total, 235 query files were produced (47 policy areas per year across 5 years). Results returned are files of all query matches found. After some cleaning and matching, results files were combined into grey literature areas per year (Table 1). The number of matches per policy and grey literature area, and per year can then be calculated, and hence, impact assessed.

Google Books API citation searches

Google Books indexes a substantial fraction of the world's books. The academic books in its collection may contain references to grey literature. The free Google Books API can be queried via Webometric Analyst (WA) for URLs, as in the case of Scopus above. Whilst Scopus only returns a result if an exact match is found within the reference section, Google Books also returns close matches but highlights the matched section in the results returned. WA contains routines to filter out false positives by excluding results that do not contain the original query URL. However, due to the length of the original query on some URLs and imperfections in the Google Books description field (such as additional spaces or text wrapping issues), matches can be missed. Due to this, a second matching method was also used, as described below. All URLs have the form:

gov.uk/government/[*article-title-separated-by-hyphens*]

Here, "gov.uk" is the hostname and "government/[*article-title-separated-by-hyphens*]" is the path. The hostname and first part of the path (gov.uk/government/) are common to all grey literature references within this repository and are therefore useful to match true citations.

Nevertheless, text wrapping could cause a problem due to the length of some URLs. If the URL part of the reference were to wrap to more than one line, URLs referenced might change due to the addition of an extra hyphen or a line break, causing a match to be missed. To avoid this issue the hostname and first part of the path (gov.uk/government/) were removed and two Google Books search strategies were formed.

For Google Books API to search the database, a text file for each grey literature area in each year was created. The file contained each query term, listed one per line. Here, each query did not include the "REF" part (as in Scopus), as Google Books does not have the ability to search a reference section specifically. Examples of each search strategy for the example above, to find matches for the document at URL gov.uk/government/publications/ahead-of-the-curve, are:

publications/ahead-of-the-curve
www.gov.uk

The list of queries for each search strategy was then separately input into Google Books API search contained within Webometric Analyst which automates the search process. As before, 235 query files were produced per search strategy (47 policy areas per year across 5 years). Results returned are files of all query matches found, including false positives where a similar match is found. Webometric Analyst also includes further routines to match the original query to the description output for each result, to ascertain true matches.

In pilot studies, comparisons between the two Google Books search strategies were performed to determine if one is inherently more suitable than the other. It was decided that the second search strategy using only the hostname (queries matching only www.gov.uk) was too general, causing matches to general webpages on the UK government website. The first search strategy, although possibly missing some matches due to the length of each query, was more specific and has better precision than the other search strategy.

From this decision, the Scopus search strategy defined above was finalised to be the same as Google Books – so both Scopus and Google Books were searched with the same part of the URL per query. This should help equate precision levels across the separate digital library searches.

After some further cleaning and matching, results files were combined into grey literature areas per year (Table 1). The number of matches per policy and grey literature area, and per year can then be calculated, and hence, impact assessed and compared to Scopus.

Following some pilot studies, some of the highest-ranked documents have very generic URLs. These may be overrepresented in this study as the citation count for the URL may include other URLs within the repository that start with exactly this URL, followed by further phrases.

Manual checking of results is needed, so precision was also calculated due to help remove the inclusion of false positives, estimated from a sample. A random sample of 50 documents in the original data that had at least one citation in Scopus API was extracted and manually searched in Scopus Advanced Search. This was then repeated for a further random sample of 50 with at least one citation in Google Books API and checked manually in Google Books. Precision for each document was calculated by comparing the automated citation count and manual citation count, and the smaller of the two was divided by the larger. The overall precision of each online library was then estimated by taking the geometric mean of the 50 document's precision levels.

Results

Proportions of UK government publications with Scopus or Google Books citations

Since most documents received no citations, the results focus on the proportion cited rather than the average number of citations per document. Other measures of impact exist that can deal with mostly uncited datasets, such as (Equalised) Mean-based Normalised Proportion Cited (MNPC and EMNPC) or Mean Normalised Log-transformed Citation Score (MNLCS) but require a comparison to a world average (Thelwall, 2017). Here, comparisons are between different online libraries across different disciplines, not compared to similar non-grey literature articles.

The results are split by year because comparing the proportion cited between years may be misleading due to the different lengths of time for a document to be cited; older documents with lower impact may report higher than newer document with a higher potential impact. Comparisons between the original 47 policy areas as defined in the repository between the two search strategies are in the online Appendix (Table 3: https://figshare.com/s/51a8308bdf43772820b3).

The proportion cited from Scopus article path matching are always significantly above the proportion cited from Google Books with article path matching (all lower 95% confidence intervals for Scopus are larger than upper 95% confidence intervals for Google Books article path), across all years and all grey literature areas (55 occasions, 11 areas per year across 5 years) (Figures 1-5).

The more impactful grey literature areas have a proportion cited on Scopus >10% for most years, and some lesser impactful areas still have a proportion cited on Scopus >5% for older years, so a substantial minority have been cited.

As can be seen in figures 1-5, the proportion cited is generally higher in Scopus, and it seems that journals may cite grey literature more often than books. Nevertheless, the difference may be due to different levels of recall for the two search strategies.

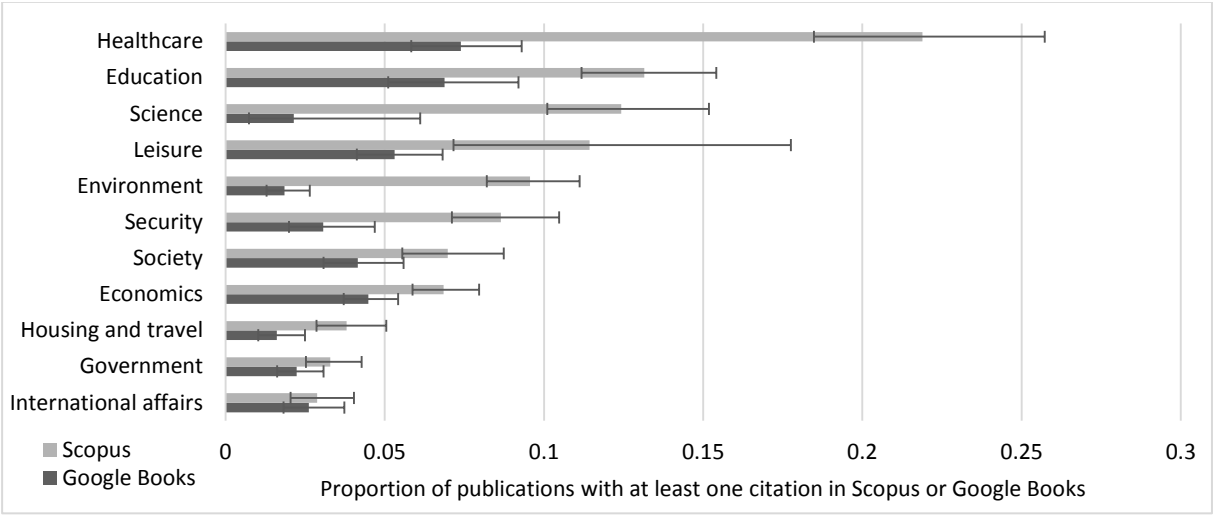

Figure 1. Proportion of UK government publications in 2013 with at least one citation in Scopus or Google Books with 95% confidence interval across 11 areas (Sorted by largest Scopus cited).

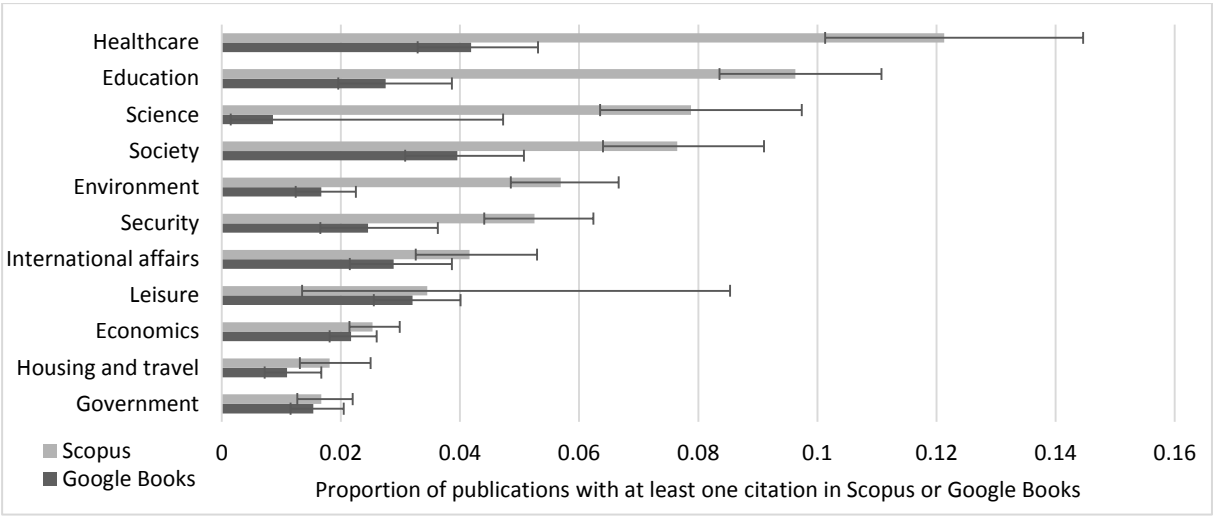

Figure 2. Proportion of UK government publications in 2014 with at least one citation in Scopus or Google Books with 95% confidence interval across 11 areas (Sorted by largest Scopus cited).

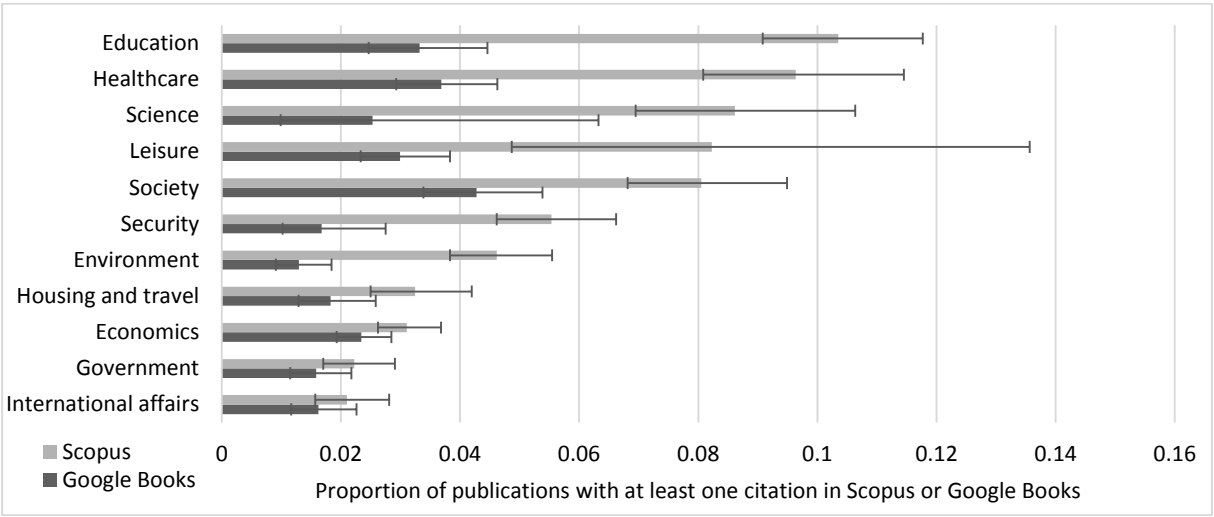

Figure 3. Proportion of UK government publications in 2015 with at least one citation in Scopus or Google Books with 95% confidence interval across 11 areas (Sorted by largest Scopus cited).

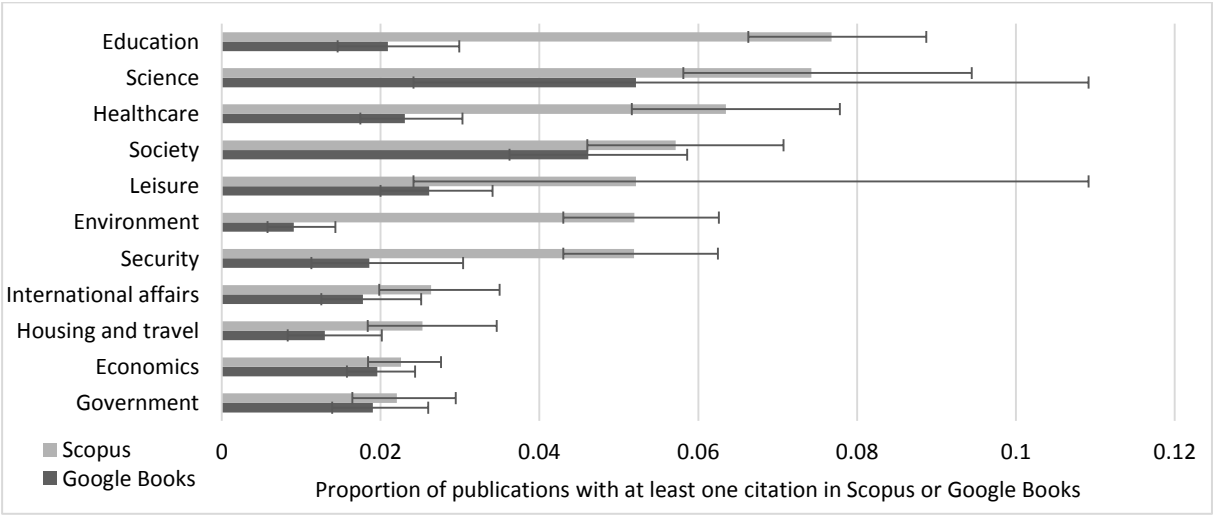

Figure 4. Proportion of UK government publications in 2016 with at least one citation in Scopus or Google Books with 95% confidence interval across 11 areas (Sorted by largest Scopus cited).

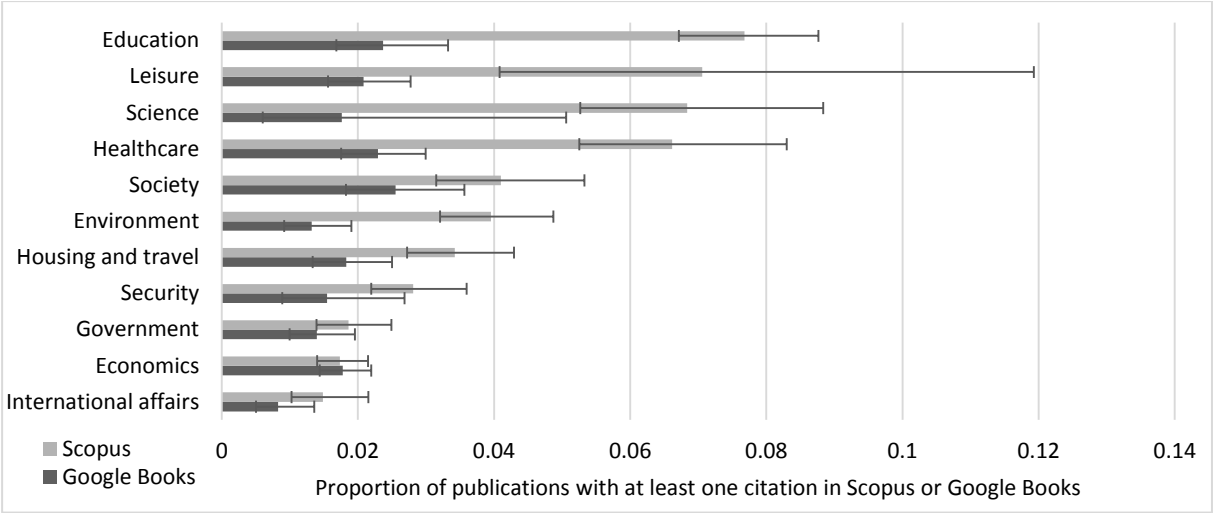

Figure 5. Proportion of UK government publications in 2017 with at least one citation in Scopus or Google Books with 95% confidence interval across 11 areas (Sorted by largest Scopus cited).

Characteristics of the most cited UK government grey literature in Scopus

The top three grey literature areas by proportion cited are Healthcare, Education and Science for each of the years 2013-2016 within Scopus references, and the same three are in the top four in 2017, with Leisure as second most cited, although with a large confidence interval.

In the first two years, Healthcare had the most impact, and has the second, third and fourth highest for 2015-2017 respectively. Education is in the top 2 most impactful grey literature areas; highest in 2013 and 2014, and second in all other years. Science is always third most impactful except for 2016, when it was second. In contrast, the grey literature areas International affairs, Economics and Government regularly finished bottom or near-bottom of the most impactful topics.

The grey literature area Healthcare in 2013 appears to be an anomaly due to its relatively high proportion cited (Scopus 0.22), with no other Scopus measurement above 0.13 for any year. A specific event, such as a national news story or major change in guidelines, may have caused a relative increase in 2013 research citing grey literature.

Table 2 shows the 25 most Scopus-cited grey literature documents across all subject areas. The five most cited grey literature documents in each subject area are shown in the online Appendix (Table 4: https://figshare.com/s/51a8308bdf43772820b3).

Table 2. Top 25 most cited UK government publications as found by Scopus.

Title (in bold) *URL (preceded by gov.uk/government/)*	*Year*	*Policy area*	*Grey literature area*	*Scopus citations*
Prisoners' criminal backgrounds and proven re-offending after release publications/2012	2013	Crime and policing	Security	3933
Housing publications/housing	2016	Tax and revenue	Economics	472
Climate change publications/climate-change	2017	Environment; Food and farming; Wildlife and animal welfare	Environment	333
Costs in disputed applications (PG38)	2017	Housing; Business	Housing and	260

publications/costs		and enterprise	travel; Economics	
Mental health and travelling abroad publications/mental-health	2014	Foreign affairs	International affairs	224
Bridges publications/bridges	2015	Government efficiency, transparency and accountability	Government	129
English indices of deprivation 2015 statistics/english-indices-of-deprivation-2015	2015	Community and society	Citizenship	121
Sustainability publications/sustainability	2013	Tax and revenue	Economics	112
NHS reference costs 2012 to 2013 publications/nhs-reference-costs-2012-to-2013	2015	National Health Service	Healthcare	97
NHS reference costs 2014 to 2015 publications/nhs-reference-costs-2014-to-2015	2015	National Health Service	Healthcare	84
NHS reference costs 2013 to 2014 publications/nhs-reference-costs-2013-to-2014	2015	National Health Service	Healthcare	83
Staffing publications/staffing	2017	Government efficiency, transparency and accountability	Government	72
NHS Constitution for England publications/the-nhs-constitution-for-england	2015	National Health Service	Healthcare	65
E-cigarettes: an evidence update publications/e-cigarettes-an-evidence-update	2015	Public health	Healthcare	56
Start active, stay active: report on physical activity in the UK publications/start-active-stay-active-a-report-on-physical-activity-from-the-four-home-countries-chief-medical-officers	2016	National Health Service; Public health	Healthcare	54
Energy consumption in the UK statistics/energy-consumption-in-the-uk	2017	Energy; Climate change	Science; Environment	50
NDNS: results from Years 1 to 4 (combined) statistics/national-diet-and-nutrition-survey-results-from-years-1-to-4-combined-of-the-rolling-programme-for-2008-and-2009-to-2011-and-2012	2017	National Health Service; Public health; Children and young people	Healthcare; Citizenship	46
Facts and figures statistics/facts-and-figures	2014	Business and enterprise	Economics	45
Social media publications/social-media	2015	Wales	International affairs	44

Websites publications/websites	2014	Transport; UK economy	Housing and travel; Economics	44
Open Data Charter publications/open-data-charter	2013	Government efficiency, transparency and accountability	Government	42

The top-ranked documents have generic URLs, such as gov.uk/government/publications/2012 (first in Table 2) and are overrepresented here as this URL does not represent the entire article title, and there are other URLs within the repository that start with this URL (gov.uk/government/publications/2012-user-event-taking-part-survey for example). Following this, URLs such as gov.uk/government/publications/open-data-charter (25th in Table 2) appear to be a generic URL due to words used and length, but will not be as generic as the first one.

For example, documents with citation counts that matched between Scopus API and Scopus Advanced Search had an accuracy of 1 (100%). Those with citation counts of one in either method and two in the other had an accuracy of 0.5 (50%), and vice versa. This way, each non-agreement results in a fall in accuracy, whether the non-agreement is due to a false positive or a missed match. A combined precision of 0.82 (82%) was estimated for Scopus and 0.71 (71%) for Google Books, each calculated using the geometric mean of 50 text's precision levels.

Excluding these general URLs (Table 2), the themes of the most cited articles (articles with >60 citations) are statistics of an annual report, multiple annual healthcare reports, general healthcare updates/studies and the NHS Constitution. This agrees with the results at the start of this section, showing that healthcare is generally the most cited topic within grey literature. This is possibly due to the importance that current healthcare policy has on relevant practice from medical professionals, teaching within the sector and future policy changes in a publicly transparent field. Furthermore, an example such as "E-cigarettes: an evidence update" is one of the most cited, non-generic URL documents. It is of note due to the rising amount of healthcare research now surrounding the use of electronic cigarettes and derivatives, due to the unknown long-term problems with their use (Callahan-Lyon, 2014).

Another example of time-appropriate research is that of the document "Start active, stay active: report on physical activity in the UK". It has a very specific URL but is relatively highly cited. Physical activity is a useful tool for combatting many issues such as obesity (Bray et al, 2016) and cardiovascular disease (Wilson, Ellison & Cable, 2016), and with an increase of these problems in recent years, it is important to make sure research incorporates all aspects of research, including that of grey literature.

As shown in the online Appendix (Table 4: https://figshare.com/s/51a8308bdf43772820b3), and ignoring the generic URLs cited (as in Table 2), the types of publication within each grey literature area appear to vary. For example, like Healthcare as mentioned in the analysis of Table 2, the grey literature areas Housing and Travel, Science and Security all have highly cited annual reports that would naturally be updated yearly. These may be highly cited as they are updated each year, so the most recent version is always relevant. As new versions are released, old forms may be cited for comparative reasons.

Education, for example, features highly cited articles that centre around unique-to-the-field reasons, namely the National Curriculum. Four of the top five most cited articles are focussed on different subjects or levels within the curriculum, across all ages from school entry to leaving at age 16 or 18. Education is arguably one of the most important areas of research due to the importance of learning from a young age, in addition to the increasing adoption of

technology in the classroom at all levels in recent years (Davison & Lazaros, 2015, Domingo & Garganté, 2016). Alongside this, a highly cited article on SEND (Special Educational Needs and Disabilities) is about codes of practice within this area (also classified as a Society grey literature document in this study). This may be due to an increasing focus on inclusion of children with special education needs in the classroom within the regular school lesson (Hornby, 2015, Bryant, Bryant & Smith, 2017).

Discussions and limitations

Using Google Books and matching just the term www.gov.uk in the description field gives more results due to the inclusion of extra spaces and line breaks in the description. However, this is not a good strategy because it introduces extra false matches. Any mention of any governmental page within the description field will cause a match due to all pages starting with the hostname, even if the match is a non-article such as a general webpage. Following, it can be suggested that Google Books article path has a higher precision but likely will miss some matches. Scopus appears to have a balance in terms of higher recall and improved precision compared to Google Books search strategies – a more specific matching term with no major issues found when matching article path to generate results.

From this, Google Books article path has been shown to display a lower proportion cited overall. Although no 'gold standard' to measure online impact within grey literature exists, the results suggest that Scopus API references when matched with the article path part of the URL is likely the best search strategy from those studied here.

To ease the collection and impact assessment of grey literature in future, it may be useful for publishers of these documents to provide their publications with persistent identifiers like DOI.

Limitations

The results are limited using a single case study (UK government publications). The Scopus API requires a paid subscription to use and is limited to 10,000 queries per week. Research of this size may take 10 weeks (n=97,150 for this study), and larger studies may take longer.

Merging of UK government policy areas are somewhat arbitrary for certain areas. Although the policy areas 'schools', 'further education and skills' and 'higher education' form a logical group, others are less intuitive, such as 'food and farming' within 'environment' and the 'housing and travel' grouping. It appears that the more ambiguous groupings were the less impactful, so should not affect the results much, but care should be taken if grouping into grey literature areas. In addition to this, the policy areas defined in the repository used have changed since data was gathered for this study, reducing the number of policy areas. As the total is reduced, it is likely that this may counter some of the problems when defining grouping into grey literature areas.

Several generic URLs were found within the repository that produced many incorrect search matches. This problem needs to be mitigated by data cleaning. The removal of generic URLs may be necessary if studying characteristics of specific documents. Determining which URLs are generic and specific requires manual checking of results, which increases time needed. For the results with extreme citation counts (publications/2012 with 3922 citations, for example), a sample of these matches must be checked to assess the proportion (accuracy and coverage) of false matches to generate an estimate of the total number of correct matches.

Conclusions

In answer to the research questions, a semi-automatic method can be used to identify grey literature publications for both Scopus and Google Books. Although some data collected may need to be cleaned and some text editing required for matching in Webometric Analyst, most

steps of the method can be run automatically. From this, the impact of a grey literature article can be gauged using a specific repository. If the repository can be crawled or data can be manually gathered, Scopus can be used to determine how often it has been cited. In addition, the impact of grey literature documents can also be assessed through Google Books.

Scopus appears to be a better measure of impact for grey literature compared to Google Books, at least in terms of generating more matches in addition to a higher level of precision (generated from a random sample of 50 cited documents). Pilot studies showed a larger impact measurement if matching Google Books to a more generic but still suitable matching term. Although recall will be higher, precision would be lost due to the matching term not including any part of the article title (or URL equivalent). Precision and recall are acceptable when using this method for grey literature, as judged for Scopus API, showing clear differences in impact for each grey literature area across all years, when it exists. Google Books suffers with precision if the matching term is too generic, and recall is lower with equivalent matching terms.

Finally, Healthcare, Education and Science seem to be the most cited type of grey literature, at least in terms of UK government documents. Researchers assessing document-based knowledge flows in these areas should include grey literature within their analysis in order to get a more complete picture, who can be assisted by publishers of grey literature by including persistent document identifiers such as DOI.

Appendices

Appendix 1 (Table 3), Appendix 2 (Table 4), and Appendices 3-7 (Figures 6-10), as referred to above, can be found in the online Appendices (https://figshare.com/s/51a8308bdf43772820b3).

References

Bray, G.A., Frühbeck, G., Ryan, D.H. & Wilding, J.P.H. (2016). Management of obesity. *The Lancet*, 387(10031), 1947–1956.

Bryant, D., Bryant, B. & Smith, D. (2017). Teaching Students with Special Needs in Inclusive Classrooms. *ELT Journal*, 71(4), 659.

Callahan-Lyon, P. (2014). Electronic cigarettes: human health effects. *Tobacco Control*, 23(2), 36–40.

Davison, C.B. & Lazaros, E.J. (2015). Adopting Mobile Technology in the Higher Education Classroom. *The Journal of Technology Studies*, 41(1), 30–39.

Domingo, M.G. & Garganté, A.B. (2016). Exploring the use of educational technology in primary education: Teachers' perception of mobile technology learning impacts and applications' use in the classroom. *Computers in Human Behavior*, 56(3), 21–28.

GreyNet International (2019). *Document Types in Grey Literature*. Retrieved January 4, 2019 from: http://www.greynet.org/greysourceindex/documenttypes.html.

Harzing, A. W. K. (2010). *The publish or perish book*. Melbourne: Tarma software research.

Hornby, G. (2015). Inclusive special education: development of a new theory for the education of children with special educational needs and disabilities. *British Journal of Special Education*, 42(3), 234–256.

Institute of Medicine (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research* (p. 112). Washington, DC: The National Academies Press.

Interagency Gray Literature Working Group (IGLWG, 1995). *Gray Information Functional Plan (GIFP)*, Retrieved January 7, 2019 from: https://apps.dtic.mil/dtic/tr/fulltext/u2/b300928.pdf.

Kousha, K. & Thelwall, M. (2015). An automatic method for extracting citations from Google Books. *Journal of the American Society for Information Science and Technology*. 66(2), 309–320.

Kousha, K. & Thelwall, M. (Submitted). Can Google Scholar and Mendeley help to assess the scholarly impacts of dissertations? *Journal of Informetrics*.

Kousha, K., Thelwall, M. & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, 62(11), 2147–2164.

Orduna-Malea, E., Martín-Martín, A. & López-Cózar, E. D. (2017). Google Scholar and the gray literature: A reply to Bonato's review. arXiv preprint arXiv:1702.03991.

Schöpfel, J. (2010). Towards a Prague Definition of Grey Literature. *Twelfth International Conference on Grey Literature: Transparency in Grey Literature*. Prague, Czech Republic, 6-7 December 2010. Grey Tech Approaches to High Tech Issues, 11-26.

Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, 11(1), 128–151.

University of New England (UNE, 2019). *Grey literature*. Retrieved January 4, 2019 from: https://www.une.edu.au/library/support/eskills-plus/research-skills/grey-literature.

Wilkinson, D., Sud, P., & Thelwall, M. (2014). Substance without citation: Evaluating the online impact of grey literature. Scientometrics, 98(2), 797-806.

Wilson, M.G., Ellison, G.M. & Cable, N.T. (2016). Basic science behind the cardiovascular benefits of exercise. *British Journal of Sports Medicine*, 50(2), 93–99.