Scientometrics (2022) 127:3489–3504
https://doi.org/10.1007/s11192-022-04398-3

The high scholarly value of grey literature before and during Covid-19

Kayvan Kousha[1] · **Mike Thelwall**[1] · **Matthew Bickley**[1]

Received: 28 October 2021 / Accepted: 27 April 2022 / Published online: 21 May 2022

Abstract
New academic knowledge in journal articles is partly built on peer reviewed research already published in journals or books. Academics can also draw from non-academic sources, such as the websites of organisations that publish credible information. This article investigates trends in the academic citing of this type of grey literature for 17 health, media, statistics, and large international organisations, with a focus on Covid-19. The results show substantial and steadily increasing numbers of citations to all 17 sites, with larger increases from 2019 to 2020. In 2020, Covid-19 citations to these websites were particularly common for news organisations, the WHO, and the UK Office for National Statistics, apparently for up-to-date information in the rapidly changing circumstances of the pandemic. Except for the UN, the most cited URLs of each organisation were not traditional report-like grey literature but were other types, such as news stories, data, statistics, and general guidance. The Covid-19 citations to most of these websites originated primarily from medical research, commonly for coronavirus data and statistics. Other fields extensively cited some of the non-health websites, as illustrated by social science (including psychology) studies often citing UNESCO. The results confirm that grey literature from major websites has become even more important within academia during the pandemic, providing up-to-date information from credible sources despite a lack of academic peer review. Researchers, reviewers, and editors should accept that it is reasonable to cite this information, when relevant, and evaluators should value academic work that supports these non-academic outputs.

Keywords Grey literature impact · Non-academic impact · Covid-19 · Coronavirus

✉ Kayvan Kousha
k.kousha@wlv.ac.uk

Mike Thelwall
m.thelwall@wlv.ac.uk

Matthew Bickley
m.bickley@wlv.ac.uk

[1] Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

Introduction

References in academic research publications fulfil different needs, not all of which can be adequately served by other academic publications. Their primary purposes are to acknowledge previous contributions and support the argument within a paper that the new research is valid and important enough to be published (Small, 2004). The sections of an article may be supported by references in different ways, however (Thelwall, 2019). Peer reviewed articles are often cited to justify the validity of a method, identify a research gap to be filled, and introduce theory underpinning the study (e.g., Spiegel-Rosing, 1977). In contrast, the following list illustrates some purposes that might be better served by citing documents (including web pages) that have not been formally published and peer reviewed.

- Statistics to justify the importance of a topic (e.g., Swales, 2011), such as the death toll of Covid-19 in a country or the world.
- Practical issues addressed by the research, such as national guidelines for Covid-19 social distancing.
- Influential non-academic documents analysed by the research, such as United Nations Food and Agricultural Organisation policy recommendations for protecting the food supply chain during pandemic social distancing restrictions.
- Facts relevant to the context of a study, such as a news story about a politician's speech in an article about political responses to the pandemic.
- Non-academic data analysed in a study, such as datasets from the UK's Inter-Departmental Business Register collected by the government through its official functions, used to assess the impact of the pandemic on different business types.

An important advantage of non-academic citations is that they can provide a bridge to the external world and ground academic research in practical concerns. Thus, it seems possible that as academia in some countries increasingly conducts research with societal impact (e.g., as assessed in the UK Research Excellence Framework impact case studies: Watermeyer & Hedgecoe, 2016), the prevalence of non-academic citations should increase. The Covid-19 pandemic seems to have accelerated this by partly refocusing the attention of academia on immediate practical issues, such as identifying characteristics of the virus and illness (Di Girolamo & Reynders, 2020; Kousha & Thelwall, 2020). The pandemic may therefore have led to an increase in the citing of non-academic sources, reflecting an enhanced value for grey literature within academia. Identifying such changes may help authors, reviewers, and research evaluators to judge the appropriateness of non-academic sources of reference for academic publications.

Previous analyses of the prevalence of citations to grey literature in academic research have tended to have a relatively narrow scope because of the lack of a universal database of citations to grey literature. For example, studies have investigated citations to UK government publications (Bickley et al., 2020), biotechnology articles citing patents (Glänzel & Zhou, 2011), grey literature cited by six nursing journals (Woods et al., 2020), and grey literature cited in impact case studies (Kousha et al., 2021). The limited nature of previous studies reflects the difficulty in identifying citations to grey literature. Small scale investigations can manually check reference lists in sets of articles for grey literature publications (Woods et al., 2020), presumably capturing all relevant sources. Moreover, analyses of specific subsets of the grey literature may be able to define Scopus queries to identify

relevant citations (e.g., those including patent terminology). It is also possible to identify documents that cite online grey literature by running Scopus queries for URLs in reference lists and automatically extracting the URL in references that seem likely to be for grey literature (e.g., for YouTube: Kousha, et al., 2012). This facility makes it possible to run large-scale grey literature studies based on finding references that include distinctive URLs (e.g., www.india.gov.in/*). Although such results are restricted to online grey literature, there seems to be a shift to putting all information free online and so an increasing (but unknown) proportion of the grey literature can now be found in this way. For example, government publications are increasingly online, as are major patent databases, such as that of the USA (www.uspto.gov). This improves the validity of heuristics that use URLs in reference lists as a device to find grey literature citations.

This article assesses the extent to which grey literature from a set of large organisations has been cited in academic publications to assess whether citations to grey literature are common and whether they have become more prevalent during Covid-19. The focus is on citations to a set of large organisations' URLs rather than citations to all grey literature because it is not yet practical to automatically identify all citations to grey literature with URLs because many published academic articles may also be cited with a variety of URLs (from publishers, institutions, personal home pages, repositories). This article also discusses the practical implications for policymakers.

Grey literature background

The widest definition of grey literature is that it is text-based output that is not formally published by commercial or academic publishers (i.e., not a book or academic journal: Carpenter, 2005). This includes, but is not limited to, webpages, patents, preprints, white papers, and social media (Aloia & Naughton, 2016; Schöpfel, 2006). The more formal and credible side of grey literature includes relatively substantial documents, such as unpublished reports, theses, conference proceedings, and government documents (Pejšová & Vaska, 2011). The difference between formal report-like grey literature and other types has been blurred by online publishing. For example, whereas previously a government might produce annual documents summarising an aspect of national statistics, this information might now be published instead, or in addition, as a series of webpages. As an example of this, one webpage contains current statistics on staff disability in UK higher education (www.hesa.ac.uk/data-and-analysis/staff/table-5), which represents the end result of a large scale data collection and collation operation but is reported as a web page rather than as a PDF report with associated methods details and commentary.

The importance of formal grey literature can be illustrated by the main differences between it and standard academic outputs. Research published in books or journal articles typically has a longer delay between analysis and publication (Pappas & Williams, 2011), and in some fields, such as biomedical sciences, only a third of research reported in meeting abstracts ever gets formally published (von Elm et al., 2003). This shows that a significant proportion of research undertaken is never made available in an academic journal or book. Even if the research is published, there may still be substantial publication delays. This suggests that much research is unavailable to be cited, at least as a standard peer reviewed article or book. Of course, not all formal grey literature is of good quality or easily and freely accessible (Batt et al., 2004). Moreover, since peer review is a cornerstone

of academic knowledge, it would be inappropriate to cite unrefereed documents in many contexts.

Even though formal grey literature has been produced for a long period (Alberani et al., 1990), its digitisation and the mass production of online information (Savić, 2018) reflects an ever-expanding scope for usage. For example, many healthcare policy guidelines from institutions such as the World Health Organization (WHO; Nove et al., 2017) and Centres for Disease Control and Prevention (CDC; Shrivastava & Mahajan, 2021), are primarily published as online webpages or PDF documents. Evaluating the citations to grey literature outputs can also help to assess the academic influence of the cited organisations, or the success of their in-house publication strategy (Bickley et al., 2020).

Research questions

This study investigates how different sources of grey literature (in the wider sense) were cited in academic articles before (2005–2019) and during (2020–2021) the coronavirus pandemic and assesses how non-academic sources helped Covid-19 research. For this, 17 grey literature sources were chosen, covering health care, news, data and statistics, and large international organisations. This ad hoc selection includes an illustrative range of different non-academic sources that could support research articles. It is not practical to cover all grey literature sources and there seems to be no systematic way to select a range of different types of source, so the 17 sites were instead manually selected to cover prominent websites of different relevant types. The research questions are as follows.

1. How frequently are the 17 chosen sources of grey literature cited in Scopus indexed journal articles during 2005–2021, is there a trend, and has Covid-19 influenced any trend?
2. Which types of grey literature documents are cited in Covid-19 research?
3. What are the broad topics of Covid-19 articles citing grey literature?

Methods

Grey literature sources

Seventeen websites hosting academic-relevant grey literature were selected for analysis, to include health care, news, statistics, and international organisations (Table 1). The selections were based on prior experience analysing grey literature citations.

Scopus URL citation search for grey literature

Elsevier's Scopus citation database was used to count URL citations (i.e., documents with references containing relevant URLs) to the 17 grey literature websites. Scopus was selected instead of Clarivate's Web of Science because it has wider coverage of academic journals and supports searches to identify cited URLs in the references of academic publications. All data collection from Scopus was carried out during the 13th of October 2021.

For the Scopus URL citation searches, the main web addresses of the seventeen grey literature sources were searched using the "Reference Website" Scopus field, such

Table 1 Seventeen sources of grey literature selected in the study and their scope of activities

Category	Source (URL)	Main activity
Health Care	World Health Organization—WHO (https://www.who.int/)	International public health policy making and cooperation
	The National Institute for Health and Care Excellence—NICE (https://www.nice.org.uk/)	The UK evidence-based healthcare guidance
	ClinicalTrials.gov (https://clinicaltrials.gov/)	A database of global clinical studies and trials
	Food and Drug Administration—FDA (https://www.fda.gov/)	Authorising the safety, efficacy, and security of drugs and medical devices
International Organisations	United Nations—UN (https://www.un.org/)	Maintaining international peace, security, and cooperation among nations
	UNESCO (https://www.unesco.org/en)	International cooperation in Education, the Sciences and Culture
	UNICEF (https://www.unicef.org/)	International aid to children worldwide
	World Bank (https://www.worldbank.org/en/home)	International financial institution providing loans and grants to the governments
	The European Union—EU (https://european-union.europa.eu/)	Political and economic cooperation between EU member states
	Food and Agriculture Organization—FAO (https://www.fao.org/home/en)	International organisation in agriculture, forestry, fisheries, and food security
Statistics	The Census Bureau (https://www.census.gov/)	Producing and analysing data and statistics about the American people and economy
	National Center for Health Statistics—NCHS (https://www.cdc.gov/nchs/)	Providing data and statistics about the public health of the American people
	Office for National Statistics (https://www.ons.gov.uk/)	Providing data and statistics about the economy, population, and society in the UK
News	BBC (http://bbc.co.uk/; http://bbc.com/)	UK based news broadcaster
	CNN (http://cnn.com/)	US based news broadcaster
	The Guardian (http://theguardian.com/)	British daily newspaper
	The New York Times (https://www.nytimes.com/)	American daily newspaper

as through the queries *WEBSITE("who.int/*")* or *WEBSITE("un.org/*")*. The query results were restricted to journal articles published between January 2005 and October 2021 to estimate the role of these non-academic web sources in formal research communication before (2005–2019) and during the Covid-19 pandemic (2020–2021). Two domains were searched in Scopus for the BBC (*bbc.co.uk* and *bbc.com*) and the Guardian (*theguardian.com* and *guardian.co.uk*) to locate citations to both of their main websites (see an example below).

> WEBSITE ("bbc.co.uk/*") OR WEBSITE ("bbc.com/*") AND PUBYEAR > 2004 AND (LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (SRCTYPE, "j"))

Because Scopus increases in size annually and to allow the analysis of the partial year 2021 data, the numbers of URL citations to grey literature were normalised by dividing by the number of Scopus journal articles for each year to examine their growth as a percentage. For instance, although there were more citations to The European Union website (europa.eu) from Scopus articles published during January to December 2020 (35,599) than from articles published during January to October 2021 (34,475), the normalised results were higher in 2021 (0.016) than in 2020 (0.014).

Additional searches were conducted to locate citations to the 17 grey literature websites from coronavirus-related research published between January 2020 and October 2021, as shown in the example below for WHO. For this, the following set of distinctive terms was searched for in titles, abstracts or keywords: *covid, covid-19, SARS-CoV-2, 2019-nCoV, coronavirus, "corona virus"*.

> WEBSITE ("who.int/*") AND TITLE-ABS-KEY (covid) OR TITLE-ABS-KEY ("SARS-CoV-2") OR TITLE-ABS-KEY ("2019-nCoV") OR TITLE-ABS-KEY (coronavirus) OR TITLE-ABS-KEY ("corona virus") OR TITLE-ABS-KEY ("2019-nCoV") AND (LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020)) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (SRCTYPE, "j")).

To assess the broad fields of the Covid-19 articles citing each website, related subcategories were merged into four subjects using the predefined Scopus SUBJAREA search field for Health Care (*MEDI OR NURS OR VETE OR DENT OR HEAL OR MULT*), Life Sciences (*AGRI OR BIOC OR IMMU OR NEUR OR PHAR*), Physical Sciences (*CENG OR CHEM OR COMP OR EART OR ENER OR ENGI OR ENVI OR MATE OR MATH OR PHYS*) and Social Sciences (*ARTS OR BUSI OR DECI OR ECON OR PSYC OR SOCI*). The query below was used to identify URL citations to the United Nations (un.org) from Covid-19 articles in Social Science subjects between 2020 and October 2021. For example, the query retrieved the article "Pandemics, tourism and global change: a rapid assessment of COVID-19" published in the *Journal of Sustainable Tourism* in 2020, citing statistics about air transport passengers.

> WEBSITE("who.int/*") AND TITLE-ABS-KEY(covid) OR TITLE-ABS-KEY("COVID-19") OR TITLE-ABS-KEY("SARS-CoV-2") OR TITLE-ABS-KEY("2019-nCoV") OR TITLE-ABS-KEY(coronavirus) OR TITLE-ABS-KEY("corona virus") OR TITLE-ABS-KEY("2019-nCoV") AND SUBJAREA(ARTS OR BUSI OR DECI OR ECON OR PSYC OR SOCI) AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020)).

Content analysis of cited grey literature by Covid-19 research

A content analysis was conducted on the 340 grey literature URLs with the most citations from Covid-19 articles to identify the characteristics of cited grey literature sources. For this, the free Webometric Analyst software (lexiurl.wlv.ac.uk) was used to automatically convert Scopus CSV outputs to text files (see "Tab-Sep", "Convert CSV to Tab-separated") and to extract the URL citations to each grey literature source from the references of the citing Scopus articles found (see "Citation" menu, "Extract URLs from Scopus Web").

The frequencies of the cited URLs were calculated after manual checking of the 20 URLs with the most citations for each grey literature website, excluding general citations to their homepages (e.g., who.int/ or fao.org/). The first author checked the contents of the 20 URLs with most Scopus citations from Covid-19 articles ($20 \times 17 = 340$) to assess what type of information was commonly cited in academic articles and whether the cited webpages were relevant to Covid-19. After examining the data for common themes, the cited HTML and PDF webpages were classified into the following five broad categories.

- *Data and statistics* This category includes webpages providing any kind of data, statistics, survey results, census, situation report or numeric indicators, graphs, charts, maps, or other numerical formats. This could be about Covid-19 (e.g., WHO global Covid-19 situation) or not (e.g., FAO Food Price Index).
- *Policy document or report* This class includes webpages reflecting policy, strategies or agendas of international organisations or governments. These were mainly about towards different development plans such as for health care, science, socio-economic actions and education. For instance, this class includes UN reports on the impact of COVID-19 on children, EU discharge criteria for confirmed COVID-19 cases, and the universal declaration of human rights.
- *News* This category includes news stories, features and analysis from news agencies, organisations or other institutions to inform the general public about an event or situation. Examples include a WHO director briefing about Covid-19, an FDA announcement about first treatments for COVID-19 and BBC news stories about reducing air pollution and CO2 during the pandemic.
- *Clinical guideline or trial* This category contains clinical documents providing recommendations, observations or trials for treatments, diagnosis, or management of diseases or other medical conditions. Examples include WHO Covid-19 clinical management guidance, the NICE (National Institute for Health and Care Excellence) guideline for post-traumatic stress disorder and a trial of remdesivir in adults with severe Covid-19.
- *General guidance* This category includes webpages providing general guidance, advice, questions and answers. It includes guidance for certifying deaths due to Covid-19, tips and guidance for washing hands to protect against Covid-19 and breastfeeding during the pandemic

Results

Normalised comparison of citations to grey literature

Figures 1, 2, 3 and 4 show the annual proportions of Scopus articles with citations to the seventeen health care, international organisation, news, and statistics websites during

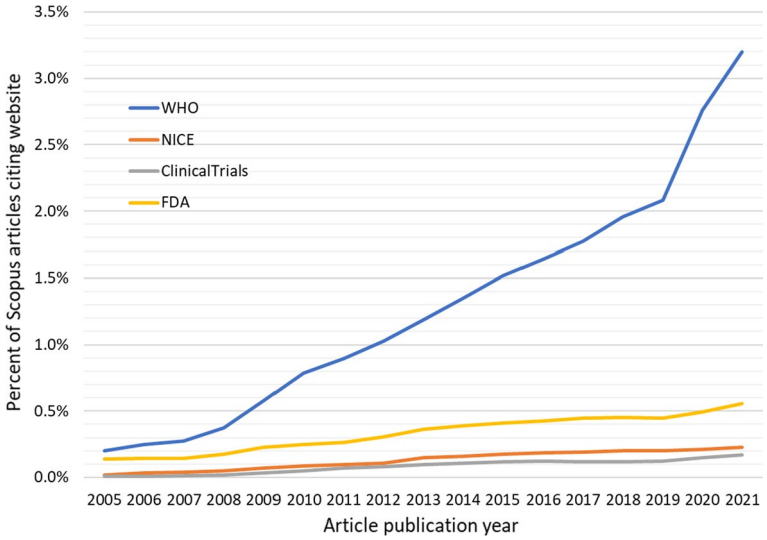

Fig. 1 The percentage of Scopus articles citing four health care websites between 2005 and October 2021

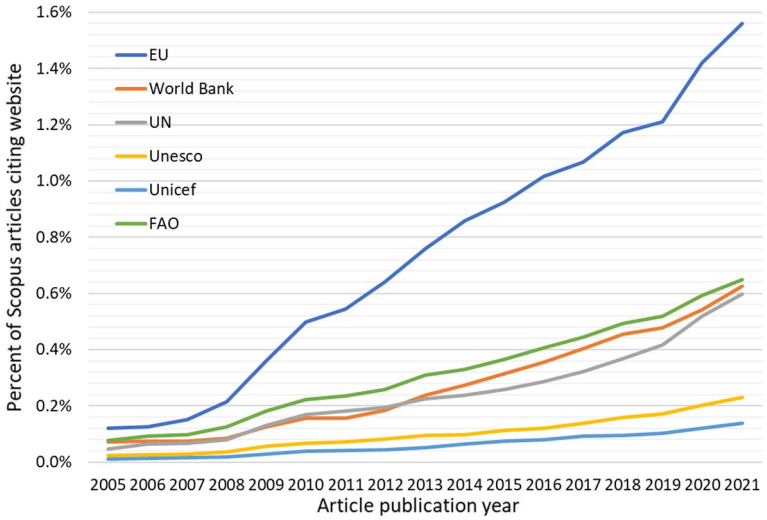

Fig. 2 The proportion of Scopus articles citing five large organisation websites between 2005 and October 2021

2005–2021. The results show that the proportions of articles citing these grey literature sources have grown over time, although there are also differences between the websites. In particular, the WHO and European Union websites had attracted the most citations from academic articles. There is also in many cases a greater increase during the pandemic (2020–2021) and the increase is greatest for WHO, the European Union, the US Center for Health Statistics, and most news websites. For instance, the proportion of academic citations to WHO during the pandemic in 2020 and 2021 was 1.3 and 1.5 times higher than

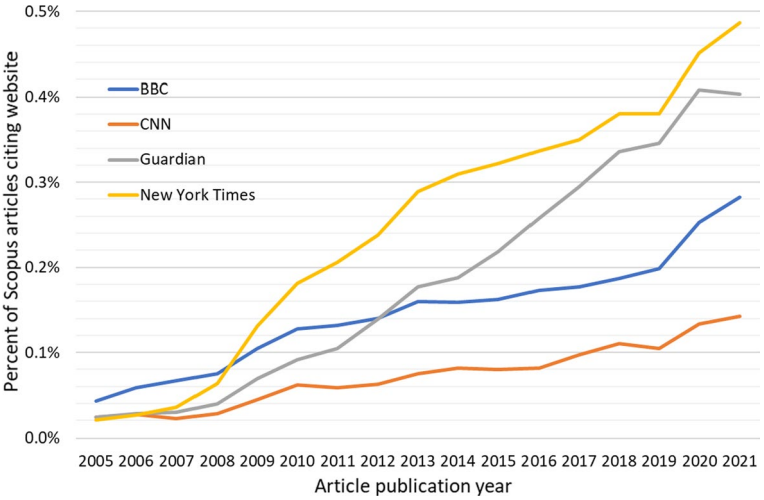

Fig. 3 The proportion of Scopus articles citing four news websites between 2005 and October 2021

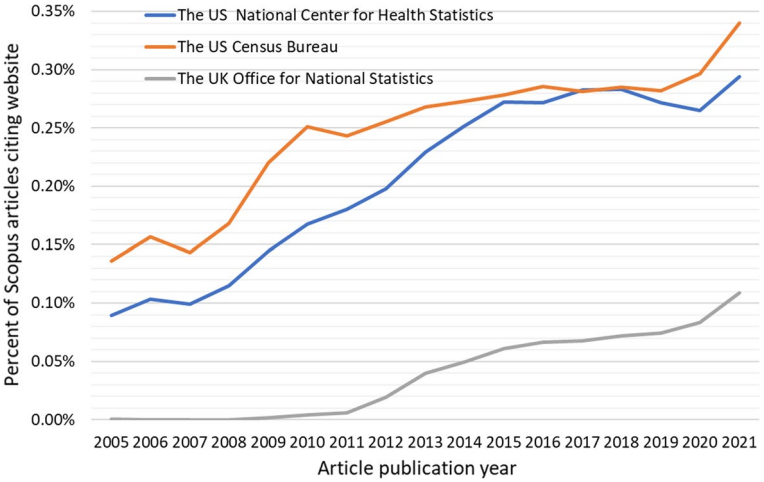

Fig. 4 The proportion of Scopus articles citing three statistics websites between 2005 and October 2021

before the pandemic in 2019. Tables A1–A4 in the online appendix report the number of Scopus articles with at least one citation to the 17 websites during 2005–2021 (https://doi.org/10.6084/m9.figshare.16810471.v1).

Academic use of grey literature for coronavirus research

Figure 5 shows the percentage of citations to the 17 grey literature sources from Covid-19 and non-Covid-19 articles published between January 2020 and October 2021. To identify citations from non-Covid-19 articles, the AND NOT operator was used to exclude

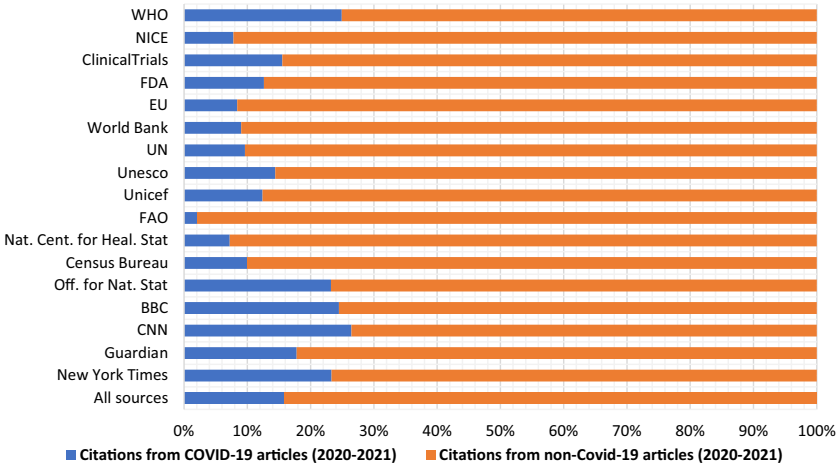

Fig. 5 The percentage of citations to grey literature sources from Covid-19 and non-Covid-19 journal articles between January 2020 and October 2021

coronavirus relevant articles.[1] Overall, 16% of the citations found were from Covid-19 research. Nevertheless, there are some differences between the 17 non-academic sources. A quarter of the citations to WHO (24.9%) and Office for National Statistics (23.2%) documents were from Covid-19 research, reflecting the greater importance of this organisation for the pandemic. Three of the news sources also attracted an above average share citations from Covid-19 articles (CNN: 26.4%, BBC: 24.5%, New York Times: 23.3%), suggesting that coronavirus information disseminated by news agencies was often useful for research. In the following section, the types of grey literature sources with the most citations from articles is discussed.

Figure 6 shows the broad subjects of the Covid-19 articles citing the 17 websites. Just under half (48%) of the Medicine and Health Science citations were from Covid-19 research, followed by Social Sciences (24.6%), Physical Sciences (14.4%) and Life Sciences (13%). This suggests that grey literature has been useful for coronavirus research outside the medical sciences, but at a lower rate. For instance, Covid-19 Social Science research was particularly likely to cite UNESCO (55%), the UN (44%), and the World Bank (43%) in addition to the Guardian (52%), BBC (45%), CNN (44%) and the New York Times (42%).

[1] An example of query used to identify citations to FAO from non-Covid-19 articles: WEBSITE ("fda.gov/*") AND NOT TITLE-ABS-KEY (covid) AND NOT TITLE-ABS-KEY ("SARS-CoV-2") AND NOT TITLE-ABS-KEY ("2019-nCoV") AND NOT TITLE-ABS-KEY (coronavirus) AND NOT TITLE-ABS-KEY (" corona virus ") AND NOT TITLE-ABS-KEY ("2019-nCoV") AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020)).

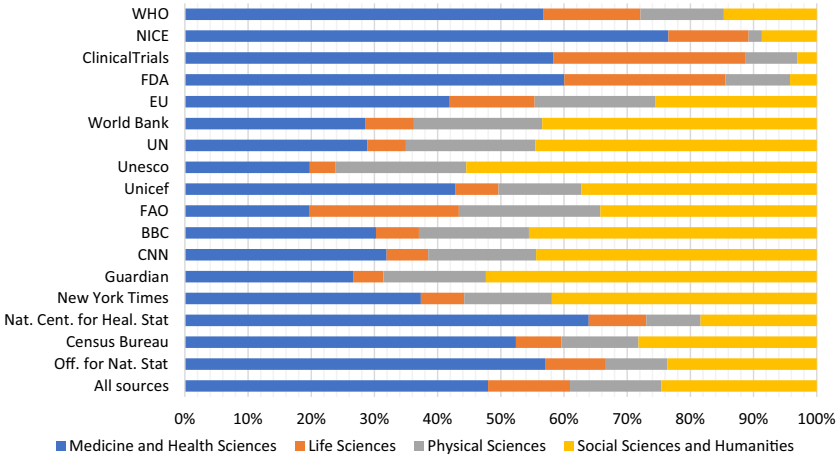

Fig. 6 The percentage of citations to grey literature sources from Covid-19 journal articles between January 2020 and October 2021 by broad subject

Types of grey literature cited in Covid-19 research

A content analysis of 340 grey literature URLs with the most Scopus citations from Covid-19 articles (20 per website) found that 81.8% (278) were Covid-relevant. The types of information usually cited varied between the sources. For instance, news (37%), data and statistics (29%) and general health guidance (25%) from the WHO website were most cited by Covid-19 research, whereas clinical guidelines or trials were predominantly cited from NICE (98%), ClinicalTrials.gov (100%), and FDA (89%). With one exception (the UN), the majority of the most cited webpages were not traditional grey literature (policy documents, reports, clinical guidelines or trials) but were newer types of grey literature.

Surprisingly, the New York Times (69%) and the BBC (42%) were often cited for data and statistics rather than for news stories. Similarly, data and statistics from four of the international organisations had also attracted many citations from coronavirus research: World Bank (79%), UNESCO (77%), EU (51%) and FAO (38%). In contrast, policy documents or reports from the UN had received the most citations (53%), followed by FAO (25%) and EU (14%). Unsurprisingly, nearly all citations to statistics websites were to data, statistics, survey results, census information, or indicators: The UK Office for National Statistics (97%), The US Census Bureau (95%), and The National Center for Health Statistics (90%). Overall, it seems that data and statistics published by many non-academic sources had an important role in supporting Covid-19 research. The citation counts, types and description of 340 grey literature URLs with the most Scopus citations from Covid-19 articles are available online (https://doi.org/10.6084/m9.figshare.16810471.v1).

Discussion

This study is primarily limited by the restriction to 17 websites rather than covering all types of grey literature. Since the results differed substantially between websites, any other selection of websites might have given substantially different results overall. Thus, the results for individual sources provide only a weak indication of the likely pattern for others. Future studies may assess the role of national grey literature sources within academia (e.g., India, Brazil, China and Russia). For instance, out of all Scopus citations to Chinese Government websites (".gov.cn/*") during the pandemic in 2020 (11,290) and 2021 (13,138) about 22 and 17% were from COVID-19 relevant articles (2227 and 2507 respectively) Similarly, about 11 and 16% of Scopus citations from journal articles to Indian Government websites (".gov.in/*") in 2020 and 2021 were from COVID-19 related research, suggesting that non-academic information published by governments can be valuable in research communication.

The almost universally increasing citing of all 17 websites (Figs. 1, 2, 3, 4) is perhaps surprising, despite the arguments in the Introduction about the increasing prevalence of online information and the increasing need to make research relevant to society. The online availability arguments do not seem relevant to news websites, for example, which were established many years ago (e.g., 1997 for the BBC: BBC, 2007) and do not seem to be expanding. In contrast, an increased rate of citation to government websites might be expected as they increase the amount of material that they put in the public domain (Bickley et al., 2020). The almost universally increasing citation rates therefore seem to reflect a greater willingness to cite non-academic sources, rather than an increase in the citable content. This may be due to the increased academic focus on societal impact (e.g., Watermeyer & Hedgecoe, 2016) and consequent need to reference external events and data, as also suggested in the Introduction. The apparent greater willingness to cite web sources may also partly stem from an easing of early fears that such citations were unhelpful because webpages could be deleted or changed (Germain, 2000; Wagner et al., 2009). Even though the issue of disappearing URLs is still relevant, authors and referees may have seen enough web citations in published research that they are rarely noticed, except perhaps in exceptional cases. They may also accept disappearing URLs for citations that contextualise a study rather than underpinning its data, methods or theory. The Internet Archive is increasingly used to fix URLs, when necessary, but still only by a minority of articles citing URLs (e.g., 3195 Scopus articles referenced it in 2020, according to: WEBSITE("www.archive.org/*") AND PUBYEAR = 2020 AND (LIMIT-TO(DOCTYPE, "ar"))). Another possible explanation for the increased citing of URLs is that online publication is standard now, with many articles being online-only, where URLs may be seen as a natural part of the format.

Although causal connections have not been directly investigated, the influence of the pandemic is clear from some of the trends. The sharp increase in citing WHO during the pandemic is expected but it is still perhaps surprising that over 3% of articles in Scopus cited it in 2021: one article out of every 31. Despite WHO not publishing formally peer reviewed documents on its website, its staff are clearly international academic medical experts, and its work is presumably internally peer reviewed or checked thoroughly by academic medical experts and so the core information on its website might reasonably be thought to have been through a form of quality control that is at least comparable to that of a rigorous academic journal. It is not clear whether this applies to the WHO Global Covid-19 situation dashboard (https://covid19.who.int), however, which had attracted the most

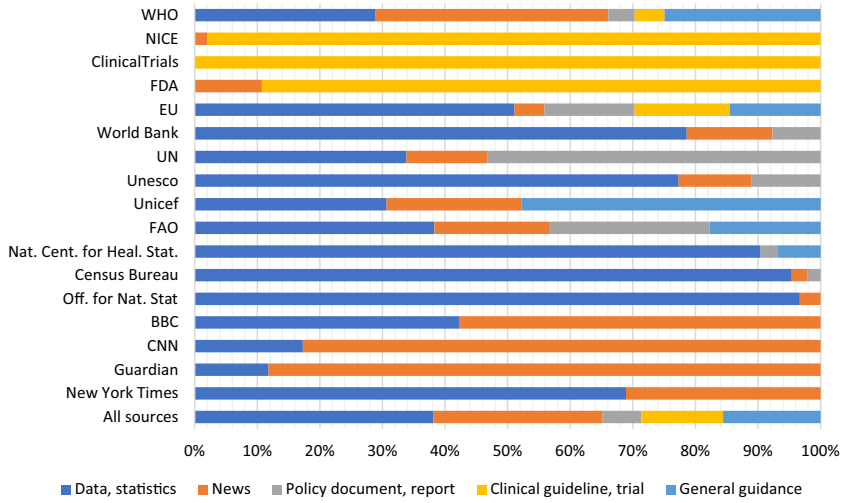

Fig. 7 The percentage of type of top 20 websites with most citations from Covid-19 articles during January 2020 to October 2021 across different grey literature sources

Scopus citations (3427) of any page from the selected grey literature sources, providing authoritative global data and statistics and other information about Covid-19.

The trends for the other three medical websites examined are also plausible. The FDA published important Covid-19 information and Covid-19 presumably prompted a dramatic increase in rapid clinical trials to test vaccines and treatments for Covid-19 as a novel virus needing urgent medical interventions. For example, the FDA's "Instructions for using real-time RT-PCR Covid-19 test" (165 Citations: https://www.fda.gov/media/134922/download) had attracted many Scopus citations. In contrast, there did not seem to be a pandemic citation increase for NICE guidelines. Whilst NICE publishes many guidelines on the treatment of conditions for UK health professionals based on expert group reviews of the literature, it also produced a series of "Covid-19 rapid guidelines" (e.g., 19 by October 2021: NICE, 2021). These guidelines are signposted on the NICE website but are hosted on a different international collaborative site, app.MAGICapp.org. Articles do not seem to be citing MAGICapp.org (e.g., it has 18 citations in 2020) instead of NICE, so it seems that the rapid guidelines are either rarely cited or that other NICE guidelines are less cited to balance their citations. For example, a NICE Covid-19 rapid guideline for managing children, young people and adults (NG159 and updated version NG191: https://www.nice.org.uk/guidance/ng191) was cited 213 times by Scopus articles.

The remaining site types (news, international organisation, statistics) all seemed to have extra citations due to the pandemic. These citations partly originated from non-medical research engaging with aspects of the pandemic and citing information about its social context, such as UNESCO statistics about Covid-19's impact on education and school closures (683 citations), death statistics involving Covid-19 in England and Wales from the Office for National Statistics (455 citations), or The New York Times Coronavirus Vaccine Tracker (145 citations). Thus, these increases presumably reflect a more immediate connection between academic research and societal needs during the pandemic. They also reflect the wide variety of non-academic sources that have helped academic contributions to addressing the needs of the pandemic such as the World Bank Indicator (643 citations),

the UN 2019 Revision of World Population Prospects (167 citations) or FAO statistics about food production, security, trade, and prices (67 citations).

In terms of the type of webpage cited, whilst traditional grey literature in the form of reports published online were sometimes cited from international organisations (mainly FAO, UN), news or media posts, or statistics were far more commonly cited (Fig. 7). Thus, academic research does not seem to be extensively citing reports that have not been through academic peer review but is instead citing other types of content, presumably for facts about the pandemic, medical information, or government actions. This partly aligns with the open science movement aims to encourage researchers to share more non-refereed content, such as preprints, software, and data. For example, the FORCE11 declaration on data sharing stresses the need to formally cite data used (to credit the authors) (Martone, 2014). This practice may have helped to normalise citing data so that non-academic data creators (e.g., Census Bureau, FAO) are also frequently cited formally when their data is used.

Conclusions

The results show that citing grey literature is common in academic research, with a single organisation, WHO, now being cited in over 3% of all papers. Citations to grey literature are common in all broad areas of science, and the usual targets are not reports but data, statistics, news stories and other webpage types. Moreover, there seems to be an increasing use of grey literature citations in academic research, including to all 17 of the organisations examined here. This increase seems to have been accelerated by Covid-19 for nearly all organisations. Two partial causes may be increased online publishing and increased acceptance of citing non-refereed and online sources, perhaps as a by-product of the open access movement. Another partial cause may be the increased importance of societal relevance for research, leading to a need to cite documentation produced by important organisations, facts about the world (such as in news stories), or to analyse (and cite) real-world data.

From the perspective of research evaluation, the results suggest that citing grey literature has become an accepted and standard part of research. Whilst this is concerning from a knowledge integrity perspective, because the cited URLs could disappear, it is positive in terms of the suggested closer connections between academic research and society, and a disappearing URL may not be critical for context-setting citations (e.g., Covid-19 citing death rates to emphasise the importance of pandemic-relevant studies). The increase in citing grey literature may reflect a greater tendency of scholarship to directly engage with societal problems, such as the pandemic, and should not therefore be discouraged unless the integrity of an article is compromised. Thus, research evaluators should be accepting of the potential value of grey literature citations from academic publications. In parallel with this, academic contributions to cited grey literature produced by non-academic organisations, such as WHO, should also be valued for their contribution to research as well as for the direct societal impact of the organisation. This should include not just reports but also general guidance and carefully collected data, statistics and facts.

Supplementary Information The online version of this article (https://doi.org/10.1007/s11192-022-04398-3) contains supplementary material, which is available to authorized users.

Data availability The shared data provides citation counts, types and descriptions of 340 grey literature URLs with the most Scopus citations from Covid-19 articles. It also includes four graphs showing the

number of Scopus articles with at least one citation to 17 grey literature sources between 2005 and October 2021 (https://doi.org/10.6084/m9.figshare.16810471.v1).

References

Alberani, V., De Castro Pietrangeli, P., & Mazza, A. M. (1990). The use of grey literature in health sciences: A preliminary survey. *Bulletin of the Medical Library Association, 78*(4), 358–363.

Aloia, D., & Naughton, R. (2016). Share #GreyLit: Using social media to communicate grey literature. *Grey Journal (TGJ), 12*(2), 97–103.

Batt, K., Fox-Rushby, J. A., & Castillo-Riquelme, M. (2004). The costs, effects and cost-effectiveness of strategies to increase coverage of routine immunizations in low- and middle-income countries: Systematic review of the grey literature. *Bulletin of the World Health Organization, 82*(9), 689–696.

BBC (2007). 10 years of the BBC News website. http://news.bbc.co.uk/1/hi/in_depth/629/629/7057140.stm

Bickley, M. S., Kousha, K., & Thelwall, M. (2020). Can the impact of grey literature be assessed? An investigation of UK government publications cited by articles and books. *Scientometrics, 125*(3), 1425–1444.

Carpenter, B. B. (2005). GrayLit network: A science portal of technical papers. *Reference Reviews, 19*(5), 40–41. https://doi.org/10.1108/09504120510604292

Di Girolamo, N., & Reynders, R. M. (2020). Characteristics of scientific articles on COVID-19 published during the initial 3 months of the pandemic. *Scientometrics, 125*(1), 795–812.

Germain, C. A. (2000). URLs: Uniform resource locators or unreliable resource locators. *College & Research Libraries, 61*(4), 359–365.

Glänzel, W., & Zhou, P. (2011). Publication activity, citation impact and bi-directional links between publications and patents in biotechnology. *Scientometrics, 86*(2), 505–525.

Kousha, K., & Thelwall, M. (2020). COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit Posts. *Quantitative Science Studies, 1*(3), 1068–1091.

Kousha, K., Thelwall, M., & Abdoli, M. (2012). The role of online videos in research communication: A content analysis of youtube videos cited in academic publications. *Journal of the American Society for Information Science and Technology, 63*(9), 1710–1727.

Kousha, K., Thelwall, M., & Abdoli, M. (2021). Which types of online evidence show the non-academic benefits of research? Websites cited in UK impact case studies. *Quantitative Science Studies*. https://doi.org/10.1162/qss_a_00145

Martone, M. (2014). *Data citation synthesis group: Joint declaration of data citation principles*. FORCE11.

NICE (2021). COVID-19 products: A list of all our products on COVID-19. https://www.nice.org.uk/guidance/conditions-and-diseases/respiratory-conditions/covid19/products?ProductType=Guidance&Status=Published

Nove, A., Cometto, G., & Campbell, J. (2017). Assessing the health workforce implications of health policy and programming: How a review of grey literature informed the development of a new impact assessment tool. *Human Resources for Health*. https://doi.org/10.1186/s12960-017-0252-x

Pappas, C., & Williams, I. (2011). Grey literature: Its emerging importance. *Journal of Hospital Librarianship, 11*(3), 228–234.

Pejšová, P., & Vaska, M. (2011). An analysis of current grey literature document typology. *Grey Journal (TGJ), 7*(2), 72–80.

Savić, D. (2018). Rethinking the role of grey literature in the fourth industrial revolution. *Grey Journal (online), 14*(2018), 7–14.

Schöpfel, J. (2006). Observations on the future of grey literature. *Grey Journal (TGJ), 2*(2), 67–76.

Shrivastava, R., & Mahajan, P. (2021). Analysis of the usage and diversity of grey literature in addiction research: A study. *Collection and Curation, 40*(3), 93–99.

Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. *Scientometrics, 60*(1), 71–79.

Spiegel-Rosing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science, 7*(1), 97–113.

Swales, J. M. (2011). *Aspects of article introductions (No. 1)*. University of Michigan Press.

Thelwall, M. (2019). Should citations be counted separately from each originating section? *Journal of Informetrics, 13*(2), 658–678.

von Elm, E., Costanza, M. C., Walder, B., & Tramèr, M. R. (2003). More insight into the fate of biomedical meeting abstracts: A systematic review. *BMC Medical Research Methodology, 3*(12), 1.

Wagner, C., Gebremichael, M. D., Taylor, M. K., & Soltys, M. J. (2009). Disappearing act: Decay of uniform resource locators in health care management journals. *Journal of the Medical Library Association: JMLA, 97*(2), 122.

Watermeyer, R., & Hedgecoe, A. (2016). Selling 'impact': Peer reviewer projections of what is needed and what counts in REF impact case studies. A retrospective analysis. *Journal of Education Policy., 31*(5), 651–665.

Woods, S., Phillips, K., & Dudash, A. (2020). Grey literature citations in top nursing journals: A bibliometric study. *Journal of the Medical Library Association: JMLA, 108*(2), 262–269.