

News stories as evidence for research? BBC citations from articles, books and Wikipedia¹

Kayvan Kousha and Mike Thelwall

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

E-mail: {k.kousha, m.thelwall}@wlv.ac.uk

Although news stories target the general public and are sometimes inaccurate, they can serve as sources of real-world information for researchers. This article investigates the extent to which academics exploit journalism using content and citation analyses of online BBC News stories cited by Scopus articles. A total of 27,234 Scopus-indexed publications have cited at least one BBC News story, with a steady annual increase. Citations from arts and humanities (2.8% of publications in 2015) and social sciences (1.5%) were more likely than citations from medicine (0.1%) and science (<0.1%). Surprisingly, half of the sampled Scopus-cited science and technology (53%) and medicine and health (47%) stories were based on academic research, rather than otherwise unpublished information, suggesting that researchers have chosen a lower quality secondary source for their citations. Nevertheless, the BBC News stories that were most frequently cited by Scopus, Google Books and Wikipedia introduced new information from many different topics, including politics, business, economics, statistics, and reports about events. Thus, news stories are mediating real world knowledge into the academic domain, a potential cause for concern.

Introduction

Citing reputable information sources to support an argument is a hallmark of academic research and of other publications that claim to be high quality, such as encyclopedias. Nevertheless, scholars seeking information to support their research claims may find some non-peer-reviewed sources useful, such as scientific videos, images, datasets, software (Kousha & Thelwall, 2015), and web sources (Yang, Qiu, & Xiong, 2010). News stories could also support scholars' claims if they contain information that is not otherwise released into the public domain, or that has not been otherwise published by a credible source.

If information of non-academic origin needs to be cited by an article, then a story published by a reputable news organisation may be a reasonable choice, especially if they have accuracy and impartiality policies (e.g., www.bbc.co.uk/editorialguidelines/guidelines/accuracy). Nevertheless, even news stories about important topics, such as health, may contain errors, omissions or misinterpretations (MacDonald & Hoffman-Goetz, 2002; Chang, 2015), and so citations to news stories risk tainting research with false information or partial information that omits critical aspects, such as the study limitations (Brechman, Lee, & Cappella, 2009).

Despite the potential use and abuse of news stories as evidence for academic research, there is little information about how widespread it is, which types of news (e.g., scientific or non-scientific) are cited

¹ This is a preprint of an article to be published in the Journal of the Association for Information Science and Technology © copyright 2017 John Wiley & Sons, Inc.

and why. This is an important omission because if news citing is common then this phenomenon needs to be investigated more thoroughly to check whether it is always used in a safe way.

The role of citations in academic research

The primary function of citations is often to provide evidence to support research. For example, a paper may cite evidence for the effectiveness of its methods or the reasonableness of any assumptions made in its research design (Case & Higgins, 2000). Partly because of these reasons, citations often acknowledge prior influences on a study (Merton, 1973). Nevertheless, citations are a biased (Borgman & Furner, 2002) and incomplete (MacRoberts & MacRoberts, 2010) reflection of those influences. Moreover, a substantial number of citations may not be essential to the new work, but may have a secondary role (Chubin & Moitra, 1975; Moravcsik & Murugesan, 1975) and perhaps less than 10% provide a central contribution to the citing paper (Hanney, Grant, Jones, & Buxton, 2005). For example, citations may serve to introduce a topic, review prior work in the same area, or acknowledge a relevant research direction or influential author (Case & Higgins, 2000).

Individual citations seem to be less critical to scholarship in the humanities and social sciences than in science because they more frequently serve to introduce context, engage additional audiences and introduce different perspectives (Hyland, 2004). Thus, inaccurate information cited by a social sciences or humanities article is less likely to fundamentally undermine the new work.

Citations to news stories

Previous analyses of types of citations in academic publications have often ignored citations to news stories, allocating them to an “Other” category (e.g., Cullars, 1992; Knievel & Kellsey, 2005; Ucak & Al, 2009; Creaser, Oppenheim, & Summers, 2011). Those that explicitly identified news have found small percentages of references in different contexts. Out of 20,000 citations in 555 English and American literature journal articles published in 2003, 1.5% targeted newspapers (Heinzkill, 2007). Swedish literary journals (2000-2009) and research applications submitted to the *Swedish Research Council* in literary studies (2006-2009) included 1%-3% citations to the daily press (Hammarfelt, 2012). Out of 18,901 references from clinical and nursing journal articles during 2004-2005, only 19 were to newspapers (Oermann et al., 2008). From 665 references in computer science papers published in 2006, 1% were to newspapers or news services (Wainer, Przibiszki De Oliveira, & Anido, 2010). In contrast, 6% of the references from a sample of 130 *Universidad de Buenos Aires* humanities and social sciences articles were to newspapers (De Tiratel, 2000).

A similar low frequency is evident within dissertations and student research reports. Less than 1% of the citations from over 100 master’s degree and doctoral dissertations in the humanities from the *University of Malaya* during 1984 and 1994 were to newspapers (Zainab & Goi, 1997), as were 2% of the citations from 24 Applied Zoology PhD. dissertations in an Indian university (Kannappanavar & Banateppanavar, 2013). Substantially more news citations were found in *Cornell University* undergraduate term papers in microeconomics, increasing from 7% in 1996 (n=57 citations) to 19% in 1999 (n=130 citations). The increase might have been due to a university subscription to *Lexis/Nexis* newspaper databases in 1997 (Davis & Cohen, 2001) or an increasing reliance on non-scholarly sources for research (Davis, 2002). Microeconomics may be a special case because few citations from student research papers in an American university in 2002 in social sciences (3%) and the humanities (1%) were to newspaper articles (Carlson, 2006) and an examination of references in over 2,000 undergraduate social science research project

reports (2009 and 2013) in a Nigerian University reported 1.5% citations to newspapers (Iroaganachi, Itsekor, & Osinulu, 2014).

Although there are relatively few scholarly citations to newspapers, encyclopaedia citations to news stories seem to be more frequent. An investigation of external references from entity pages within English Wikipedia pages (2001 to 2007) found that 20% were to news articles (Fetahu, Anand, & Anand, 2015).

Academic-related news coverage

News organisations report many topics that are relevant to academia, as well as explicitly covering new research findings. UK press coverage of health research focuses on serious diseases, topical health problems, and new treatments (Entwistle, 1995), women's health, reproduction, and cancer (Bartlett, Sterne, & Egger, 2002). Topical issues, such as *Severe Acute Respiratory Syndrome (SARS)* health risks can expect substantial international coverage and subjective content in the early stages of a scare (Lewison, 2008).

In Europe, press coverage of cancer research publications is extensive and tends to focus mainly on ways to avoid cancer, with advances in cancer surgery attracting relatively little interest (Begum, Pallari, & Lewison, 2016). Two studies have analysed BBC science news reporting. An investigation of the BBC's coverage of cancer (1998-2006) found 260 stories per year. A third were about breast cancer followed by lung (10%), prostate (8%) and skin (6%) cancer. Two-thirds of the BBC News stories cited cancer research advances. Clinical articles were more commonly cited than other article types (Lewison, Tootell, Roe, & Sullivan, 2008). An analysis of 640 online BBC News stories with science themes in 2010 found that two thirds were about new research and two fifths had comments from independent scientists. A third were about medical science and a third of those that mentioned a publication were based upon a peer reviewed journal article. A fifth of the stories mentioning journal articles hyperlinked to them (Mellor, Webster, & Bell, 2011).

Some science magazines have the translation of scientific research for a non-specialist public as their main goal. Journalists working for the *New Scientist* read and use a wide range of academic journals, especially general and medical titles including *Proceedings of the National Academy of Sciences of the USA* (7%), *Nature* (6.5%), *Science* (4.6%), *Lancet* (3.9%), and *New England Journal of Medicine* (2.6%) (Lewison & Turnbull, 2010).

News stories may also cite research indirectly through Wikipedia. An early analysis of the site from 2004 found that journalists cited Wikipedia for a variety of topics, but that it was particularly useful as a source of evidence about the Middle East in the context of ongoing conflicts there (Lih, 2004).

Research questions

Although there is some evidence that online news is cited by academic publications, the extent and nature of this are unknown. In particular, it is not clear which types of news tend to be cited, such as news about research outcomes (e.g., academic publications) or non-research news (e.g., events, interviews and politics). Although there are many different news agencies and news websites, the BBC is a logical choice because of its coverage of different topics (e.g., science, health and politics), international scope and apparent reputation for reporting accuracy (e.g., Jones, 2011; Ofcom, 2015). The BBC News website attracts more citations from Scopus-indexed publications than CNN and Reuters (Figure 1), and so it is a reasonable source for the current study. Although BBC News stories can also be cited from radio or

television broadcasts (still the most common source in the UK: Ofcom, 2015), the main stories should have online versions and it seems logical to cite these, when possible, as a more easily accessible and relatively permanent record. Google Books was used to assess book-based impact, which is particularly relevant for the arts and humanities, and Wikipedia citations were used to seek evidence of wider interest in the cited BBC News stories.

1. How often are news stories cited by journal articles and are there disciplinary differences in the answer to this question?
2. Which BBC News sections are cited by journal articles, books and Wikipedia?
3. Which specific fields, as operationalised by individual academic journals, cite news the most?
4. Are there relationships between Scopus, Google Books and Wikipedia citations to BBC News stories?
5. Which BBC News story topics are cited in academic publications?
6. Which BBC News stories are most frequently cited by Scopus, Google Books and Wikipedia?

Methods

The Scopus citation database was used to identify BBC News stories cited in academic publications. Scopus supports cited reference searches for URLs and indexes more academic publications than other citation databases (Moed & Visser, 2008). Additional searches were conducted to locate citations to these URLs in Google Books and in Wikipedia. A content analysis was then conducted on a sample of the BBC News stories to identify their main characteristics.

Scopus publications citing BBC News stories

Scopus cited references searches were used to identify academic publications citing online BBC News stories. Common BBC News story URL formats were searched for in the Scopus references field with the following query.

```
REF ("*news.bbc.co.uk/*") OR REF ("*news.bbc.com/*") OR  
REF ("*bbc.co.uk/news/*") OR REF ("*bbc.com/news/*")
```

The above query matched 31,631 Scopus publications from 2000-2015. Because Scopus allows a maximum of 2,000 full records to be exported at one time, each year was downloaded separately, excluding unrelated broad subject areas from the search results (e.g., excluding Medicine and Engineering from Arts and Humanities).

Scopus articles from related subcategories were merged in Scopus, when possible, using the refine results option for Science and Technology, Medicine and Health Sciences, and Social Sciences, but not for the Arts and Humanities. For Medicine and Health Sciences, the Scopus sub-categories merged were: Medicine; Nursing; Health Profession; Pharmacology, Toxicology & Pharmaceutics; Immunology & Microbiology. For Science and Technology, the following sub-categories were merged: Chemistry; Mathematics; Physics & Astronomy; Environmental Science; Computer Science; Engineering; Materials Science; Earth & Planetary Sciences; Agricultural & Biological Sciences; Multidisciplinary; Chemical Engineering; Energy; Biochemistry, Genetics & Molecular Biology. Finally, for Social Sciences, the subjects combined were: Social Sciences; Psychology; Business, Management & Accounting; Economics, Econometrics & Finance; Decision Sciences. This stage therefore produced four sets of Scopus publications, each of which cited one or more BBC News URLs.

BBC News URL reference extraction and correction

The URLs of the BBC News stories cited by the Scopus publications identified in the previous stage were extracted by *Webometric Analyst* (lexiurl.wlv.ac.uk, “To extract URLs from Scopus” in the Citation tab). Because specific BBC News stories were required, home pages, sub-section pages (e.g., <http://www.bbc.com/news>, <http://www.bbc.co.uk/news/scotland>) and invalid URLs were excluded. URLs were then systematically corrected when obvious typos were present (e.g., “.stm” at the end of an URL replacing scanning errors like “/stm”, “.smt”, or “-stm”). News stories with less than three terms in their titles (164, 0.6%) were excluded to avoid retrieving false matches during the subsequent searches (e.g., “The pathologist”). This stage produced a large set of URLs of BBC News stories cited by Scopus publications.

BBC News story downloading, processing, and duplicate detection

The SocSciBot web crawler (socscibot.wlv.ac.uk) was used to download the URLs of the identified BBC News stories cited in Scopus. For all 27,234 pages downloaded, the news headline, section, description and publication date were extracted with a program designed for this in *Webometric Analyst* (see the Services menu listing for BBC News).

Related BBC News sections were combined to form broad topics for analysis. For example, results from regional BBC News sections, such as *England, Scotland, Northern Ireland, Wales, Manchester, Oxfordshire* and *Lancashire*, were merged into a *UK* category and *Arts, Entertainment, Music, Film, TV and Radio*, and *Education* were combined to form *Arts, Entertainment and Education*.

The next step was to identify duplicate BBC News stories within the downloaded set. Some Scopus references omitted or shortened news story titles (e.g., “South Korean”, “S Korean” or S. Korean” in “*S Korean dies after games session*”) and so story titles were not used to identify duplicate stories. BBC News stories can also have different URLs (e.g., news.bbc.co.uk/1/hi/technology/4137782.stm or news.bbc.co.uk/1/hi/world/asia-pacific/4137782.stm). Many URLs containing “/2/hi/” redirect to a newer address with “/1/hi/” instead (e.g., from news.bbc.co.uk/2/hi/technology/4137782.stm to news.bbc.co.uk/1/hi/technology/4137782.stm). Hence, in order to identify duplicate BBC News stories, the unique part at the end of each URL was used rather than the full URL. In the second example below, an alternative identifier could be *world-europe-11843820* but we chose to include only the last segment before the number for parsimony.

- “**4137782.stm**” instead of “<http://news.bbc.co.uk/2/hi/technology/4137782.stm>”
- “**europe-11843820**” instead of <http://www.bbc.co.uk/news/world-europe-11843820>

This stage resulted in a list of *unique* BBC News stories cited by Scopus publications, together with their official titles (as downloaded in the previous stage), URLs, categories, and unique URL segments.

Counting citations to each BBC News story

Scopus citations

To count Scopus citations to the main sample of 27,234 downloaded BBC News stories, a (second) general Scopus cited references search was used to retrieve publications with at least one citation to *any* BBC News story. This matches references that include the story URL or mention the BBC in the title but not those that omit the URLs and do not mention the BBC within citation title. This is a more inclusive variant of the previous Scopus query (at the start of the methods). Whilst the former query targeted Scopus

publications citing BBC News URLs, the new query targets publications that cite a BBC News story, whether or not the citation includes a URL.

```
REFSRCTITLE("BBC News") OR REFSRCTITLE("BBC World News") OR  
REFSRCTITLE("BBC Online News") OR REF("news.bbc.co.uk") OR  
REF("bbc.co.uk/news") OR REF("bbc.com/news") OR REF("news.bbc.com")
```

There were over 37,000 Scopus matches of the above query and these were downloaded. A program was written and added to Webometric Analyst to extract the BBC references from these publications and tally matches with either the title or the unique URL part of the main sample of 27,234 BBC News stories. For instance, the program captured 40 Scopus citations matching either the news title "*Iraq war illegal, says Annan*" or the URL part "*3661134.stm*" in the Scopus cited references. Citations to BBC News stories not present within the existing set were discarded. This stage produced Scopus citation counts for the main set of unique BBC News stories.

Google Books citations

Google Books API citation searches were used to count citations from books to the main sample of 27,234 BBC News stories. Webometric Analyst ("Books" tab) was used to automatically extract and report citations from books indexed by Google Books (see: Kousha & Thelwall, 2014). Citations in Google Books were searched for with a combination of phrase searches for "BBC News" and either the story title or URL part as phrase searches, as shown in the following examples. Because the Google Books API does not support the OR operator, the two queries were submitted separately and duplicate results were excluded.

Query 1: "BBC News" "Britain is 'surveillance society'"

Query 2: "BBC News" "6108496.stm"

This stage produced Google Books citation counts for the main set of unique BBC News stories.

Wikipedia citations

For Wikipedia citation searches, the Bing API in Webometric Analyst was used to identify the number of citations to all BBC News stories (for methods see: Kousha & Thelwall, M, 2016). The search strategy above for Google Books was applied to create Wikipedia queries but the command `site:wikipedia.org/wiki/` was added at the end of each query to restrict the results to Wikipedia articles and the vertical bar "|" was used to perform an OR operation (within Webometric Analyst), as in the example below.

```
"BBC News" "Web slows after Jackson's death" site:wikipedia.org/wiki|  
"BBC News" "8120324.stm" site:wikipedia.org/wiki
```

This stage produced Wikipedia citation counts for the main set of unique BBC News stories.

Content analysis of BBC News cited by Scopus

A content analysis was conducted on a random sample of 100 BBC News stories with at least one Scopus citation (i.e., the main sample) from each of three BBC News sections: Science, Environment & Technology, Health & Medicine, and Arts, Entertainment & Education. The predefined BBC classification scheme was used for this. Three experienced independent content analysts with PhDs in the social sciences conducted this task. Pilot testing was used to devise the scheme and description, and then the

samples were coded independently. For simplicity, and based upon the results of the pilot testing, each story was given a single categorisation even in cases where it could fit more than one. Cohen's kappa values were calculated pairwise between the three coders both for the 14 narrow categories (including 'Unclear' as a category) and the 4 broad categories (including 'Unclear' as a broad category). The Cohen's kappa values were compared to standard heuristic guidelines (at least 0.21+ indicates fair agreement, 0.41+ indicates moderate agreement, 0.61+ indicates substantial agreement and 0.81+ indicates almost perfect agreement) (Landis & Koch, 1977). The following categories were used.

1. News story about research results:

- 1.1. A research study published in a journal article (e.g., bbc.co.uk/news/health-25742467)
- 1.2. A research study published or presented at a scientific conference or meeting (e.g., bbc.co.uk/news/science-environment-16040655)
- 1.3. An academic research report or research project (e.g., news.bbc.co.uk/1/hi/health/845144.stm)
- 1.4. A research study by a non-academic institution or organization (e.g., news.bbc.co.uk/1/hi/health/1562147.stm).
- 1.5. Other research study not mentioned above (e.g., news.bbc.co.uk/1/hi/sci/tech/1050495.stm)

2. News story based on an interview (or an interview transcript):

- 2.1. An interview or talk with an academic (e.g., bbc.co.uk/news/science-environment-26014584)
- 2.2. An interview or talk with an expert in the science, environment or technology sectors such as a computer professional or an astronomer (e.g., news.bbc.co.uk/1/hi/sci/tech/1898803.stm)
- 2.3. An interview with a health worker or medical professional (e.g., news.bbc.co.uk/1/hi/health/3429269.stm)
- 2.4. An interview with a politician (e.g., bbc.co.uk/news/health-28098838)
- 2.5. Other interviews (e.g., news.bbc.co.uk/1/hi/health/1861002.stm)

3. News story about legislation, a published statistic or a scientific or non-scientific event:

- 3.1. Government policy, legislation or regulation (e.g., bbc.co.uk/news/science-environment-22167675)
- 3.2. Non-research statistics including annual statistics or periodic reports (e.g., bbc.co.uk/news/uk-15426720)
- 3.3. Any type of news event or incident not covered in the other categories including a scientific (e.g., news.bbc.co.uk/1/hi/health/287880.stm) or non-scientific (e.g., news.bbc.co.uk/1/hi/health/1435409.stm)

4. Not clear, such as a page with many sections or stories (e.g., news.bbc.co.uk/1/hi/uk/2481911.stm)

Results

RQ1: Scopus citations to BBC News stories by broad discipline

An increasing number of online news articles are cited in academic research (Figure 1). The BBC News website is more often cited by academic publications than CNN and Reuters, but there seems to have been a levelling off in 2013. The absolute numbers may be misleading, however, since they are dependent on the coverage of Scopus.

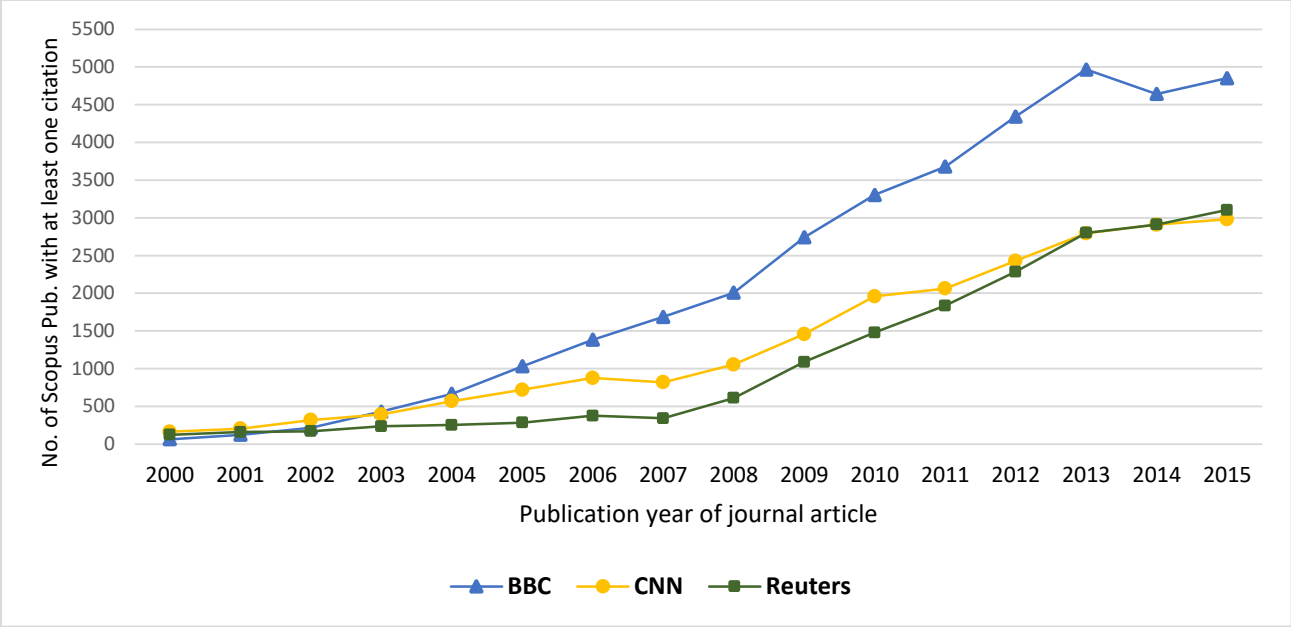


Figure 1. Scopus publications citing major news websites (May 2016). Citations to the BBC, CNN and Reuters news stories were estimated using Scopus cited references searches: For BBC News: REFSRCTITLE("BBC News") OR REFSRCTITLE("BBC World News") OR REFSRCTITLE("BBC Online News") OR REF("*news.bbc.co.uk*") OR REF("*news.bbc.com/*") OR REF("*bbc.co.uk/news*") OR REF("*bbc.com/news*"); for CNN: REFSRCTITLE("CNN") OR REFSRCTITLE("CNN News") OR REFSRCTITLE("CNN.com/") OR REF("*.CNN.com/*"); For Reuters: REFSRCTITLE("Reuters") OR REFSRCTITLE("Reuters.com/") OR REF("*.reuters.co.uk*") OR REF("*reuters.com/*") AND NOT REFSRCTITLE("*thomson*") AND NOT REF(*thomsonreuters*).

There are substantial disciplinary differences in the tendency to cite BBC News stories. The proportion of Scopus publications with at least one citation to a BBC News story is much higher in Arts and Humanities (from 0.24% in 2002 to 2.8% in 2015) and Social Sciences (from 0.15% in 2000 to 1.7% in 2013) than in Science and Technology (max: 0.05% in 2015) and Medicine and Health Sciences (max: 0.11% in 2012) (Figure 2). Although the numbers for science are perhaps negligible, news is clearly an important source of evidence in the humanities and social sciences. Taking into account the existence of other news organisations, presumably 5-10% of humanities articles now cite various news sources.

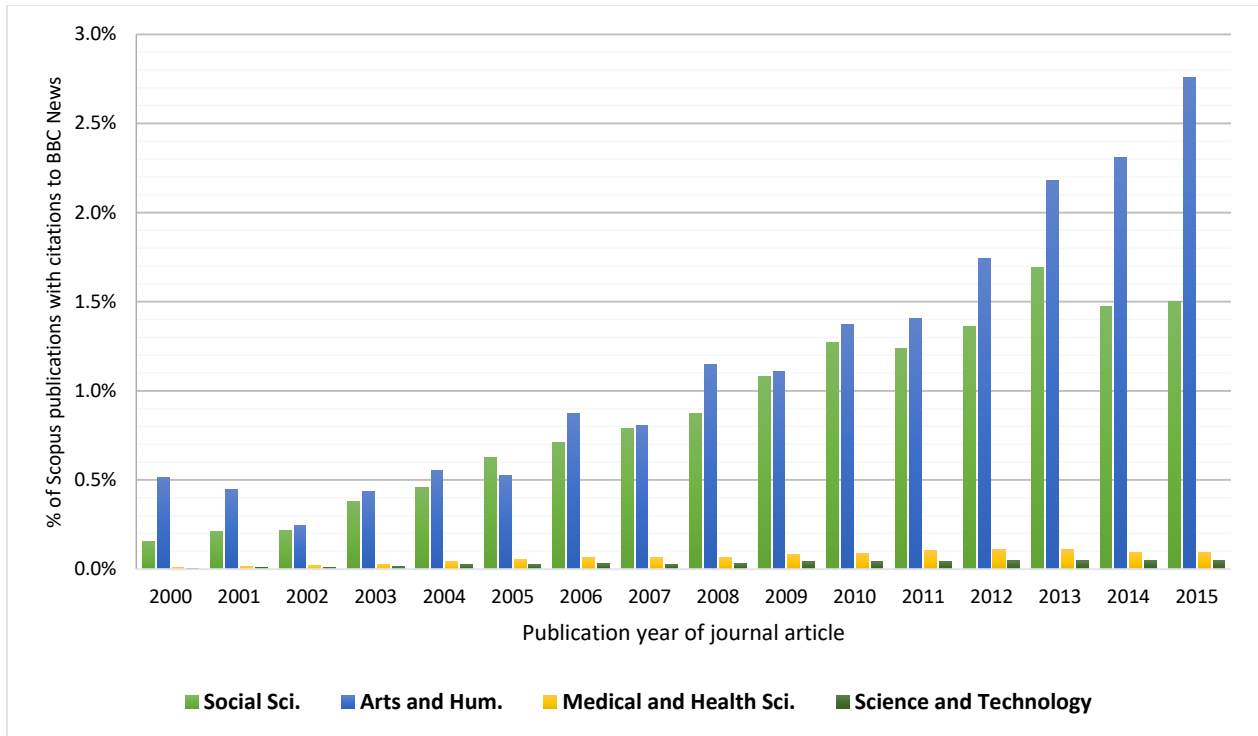


Figure 2. Scopus publications with at least one citation to BBC News stories in four broad fields 2000-2015.

RQ2: BBC News sections cited in Scopus, Google Books and Wikipedia

In terms of the types of news story cited (Figure 3), the UK section is the most common in all three databases, at least partly because of the BBC’s more comprehensive coverage of its home country. Asia is the second most cited news section. This might be due to the extent of BBC News coverage in these regions, their political and economic importance, or the historical UK connection with parts of Asia (e.g., India and Hong Kong). News about China (20%) and India (12%) were particularly common and both are large countries with growing economies. About 4% of the citations were to news stories about North Korea, mostly concerning political and nuclear issues.

The news section with the third most citations from Scopus and Google Books is Science, Environment and Technology. Many Scopus-cited BBC News stories (21%) were about computing, internet and web issues (e.g., computers, web, Google, Microsoft, Apple, Facebook) and 20% (122 of 614) of these were about computer hacking and viruses, which are presumably useful for research discussing computer security. A substantial minority (10%) of the Scopus-cited BBC News stories in the Science, Environment and Technology category were about climate change and global warming (e.g., tsunami, CO2, carbon emissions) and 5% were about astronomy and space (e.g., solar system, spacecraft, mars, telescope). In the Health section, 9% of the Scopus citations were to BBC News about the NHS (The UK National Health Service) and other health care system issues (e.g., hospitals, general practitioners), 4.1% were about babies or children (e.g., growth and obesity), 3.5% were about drugs (e.g., new drugs or safeguards), and 3.3% were about cancer.

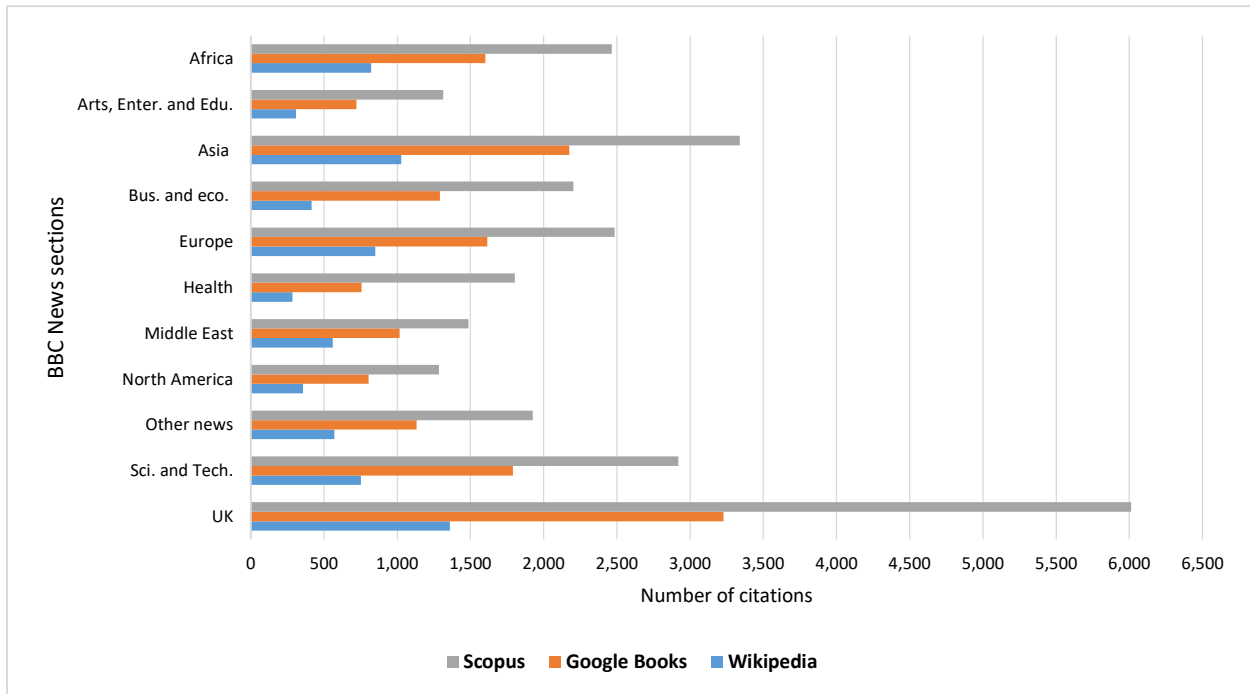


Figure 3. Percentage of BBC News stories with at least one Scopus citation across the main BBC News sections.

RQ3: Journals citing the most BBC News stories

Table A in the online appendix (<https://dx.doi.org/10.6084/m9.figshare.4003359.v1>) reports the 20 Scopus-indexed journals with the most citations to BBC News stories during 2000-2015. In the Social Sciences, most journals cover politics or international studies. The first and the fifth journals in this category cover security, terrorism and extremism in the world: The journals *Studies in Conflict and Terrorism* and *Terrorism and Political Violence* with 132 and 82 citations to BBC News, respectively. This might be due to exclusive aspects of the BBC coverage of the above incidents.

In the Arts and Humanities, the citing journals are mostly media, cultural studies, and performing arts, such as *Contemporary Theatre Review* and *Continuum: Journal of Media & Cultural Studies* with 27 and 25 citations to BBC News, respectively. In natural and environmental sciences, the journals with the most citations to BBC News are related to energy, climate change and environmental issues (e.g., *Environment and Planning A* or *Energy Policy*).

In Medical and Health Sciences, some journals with the most citations to BBC News stories are prestigious medical journals such as *BMJ Online*, *The Lancet*, and *The British Medical Journal*. Most other journals are about health care issues (e.g., *The British Journal of Midwifery*, *The British Journal of Nursing* and *BMC Public Health*) or medical ethics and law (e.g., *Journal of Medical Ethics*, *Bioethics* or *Medical Law Review*). In Computer Science and Engineering most journals are about computer security and law (e.g., *Computer Law and Security Review*).

RQ4: Relationships between Scopus, Google Books and Wikipedia Citations to BBC News stories

There are statistically significant moderate Spearman correlations between Scopus citations and Google Books citations to BBC News stories across all news sections and year ranges, suggesting that the impact of BBC News is similarly reflected in both academic articles and books (Table 1). The correlations between Scopus and Google Books citations are higher for the BBC News in the *Middle East* (ranging from 0.334 to 0.475) and *Africa* (from 0.338 to 0.456) sections. Nevertheless, the low correlations between Scopus and Wikipedia citations across most news sections suggest that the value of news for academic articles is only loosely related to its value for encyclopedia entries. The correlations between Google Books and Wikipedia citations to BBC News stories are low to moderate but the highest is for the *Middle East* section published within 2009-2012 (0.524) and 2013-2016 (0.676), perhaps due to the importance of the news in publicising important events and political statements related to this area (Lih, 2004).

Table 1. Spearman correlations between Scopus citations and Google Books or Wikipedia citations to BBC News stories by year range and section. Only articles with at least one Scopus citation are included (main sample).

BBC Sections	Correlations between Scopus citations and Google Books citations (bold) <i>Correlations between Scopus citations and Wikipedia citations (italic)</i> Correlations between Google Books citations and Wikipedia citations				
	1997-2000	2001-2004	2005-2008	2009-2012	2013-2016
Africa	.456** <i>.093</i> .296**	.398** <i>.080</i> .169**	.338** <i>.154**</i> .196**	.395** <i>.188**</i> .267**	.357** <i>.286**</i> .363**
Arts, Enter. and Education	.170 <i>.115</i> .019	.307** <i>.025</i> .219**	.276** <i>.103*</i> .226**	.436** <i>.060</i> .229**	.192 <i>.114</i> .186
Asia	.262** <i>.136</i> .207**	.356** <i>.129**</i> .174**	.354** <i>.252**</i> .217**	.354** <i>.155**</i> .208**	.330** <i>.197**</i> .408**
Business	.260* <i>.129</i> .322**	.346** <i>.168**</i> .279**	.378** <i>.224**</i> .251**	.315** <i>.145**</i> .230**	.425** <i>.235**</i> .165**
Europe	.357** <i>.248**</i> .315**	.423** <i>.200**</i> .263**	.333** <i>.122**</i> .252**	.352** <i>.155**</i> .239**	.267** <i>.191**</i> .437**
Health	.318** <i>.148</i> .154	.249** <i>.179**</i> .192**	.240** <i>.088*</i> .204**	.217** <i>.029</i> .195**	.176* <i>.161*</i> .200**
Middle East	.415** <i>.168</i> .346*	.475** <i>.207**</i> .339**	.334** <i>.175**</i> .250**	.439** <i>.266**</i> .524**	.434** <i>.384**</i> .676**
America	.216 <i>.084</i> .276*	.328** <i>.110</i> .207**	.344** <i>.126**</i> .249**	.312** <i>.201**</i> .294**	.275* <i>.427**</i> .428**

Other news and reports	.340** .196* .112	.491** .145* .228**	.320** .208** .252**	.273** .205** .281**	.404** .257** .467**
Science	.344** .119 .207*	.360** .233** .287**	.386** .199** .294**	.349** .166** .210**	.397** .253** .240**
UK	.331** .138* .241**	.338** .124** .276**	.377** .108** .237**	.352** .128** .203**	.349** .223** .280**

* Significant at the p = 0.05 level. ** Significant at the p = 0.01 level.

RQ5: Content Analysis of BBC News stories cited in Scopus

Following the Landis and Koch (1977) guidelines, inter-coder agreement rates for the narrow categories are all moderate (0.41-0.60) and the values for the broad categories are either moderate or substantial (0.61-0.80) (Table 2). These are high enough to be able to meaningfully interpret the categories but suggest that the categories require a degree of subjective human judgement to interpret.

Table 2. Cohen’s kappa agreement statistics for the 14 narrow and 4 broad content analysis categories.

BBC News sections	Arts, Entertainment & Education			Health & Medicine			Science, Environment & Technology		
	A-B	A-C	B-C	A-B	A-C	B-C	A-B	A-C	B-C
Narrow	0.515	0.694	0.555	0.410	0.479	0.481	0.595	0.486	0.463
Broad	0.528	0.691	0.633	0.477	0.536	0.520	0.688	0.563	0.521

About half of the sampled BBC News stories with at least one Scopus citation in Science, Environment & Technology (53%) and Health & Medicine (47%) were based upon published research (e.g., journal articles, conference papers, research reports, non-academic research), but only quarter (24%) in Arts, Entertainment & Education (Table 3). In contrast, over two thirds (68%) of the BBC News stories with at least one Scopus citation in the Arts, Entertainment & Education had non-research content (e.g., news about an event or incident), although there were fewer in Science, Environment & Technology (33%) and Health & Medicine (46%). Of BBC News stories with research content, a minority in Science, Environment & Technology (32%) and Health & Medicine (24%) were based on research results published in journal articles or conference papers compared with only 3% in Arts, Entertainment & Education (Table 3).

Table 3. A content analysis of a random sample of 100 BBC News stories with at least one Scopus citation from each of three BBC News sections (n=300).

Broad category	Narrow categories	BBC News Sections		
		Arts, Enter. & Edu.	Health & Medicine	Sci., Envir. & Tech.
a) BBC News about research results	Journal article	1%	17.6%	18.6%
	Conference paper	2%	6%	13%
	Research report	4.3%	5%	4.7%
	Non-academic Res.	15.7%	15.3%	11.6%
	Other research	0.7%	2.7%	4.7%

	% of news about research	23.7%	46.6%	52.6%
b) BBC News based on an interview	Interview with an academic	1.3%	1.7%	3%
	Interview with an expert	0%	0.3%	2.7%
	Interview with a med. prof.	0%	1.7%	0%
	Interview with a politician	0.3%	0.3%	0.7%
	Other interviews	6.6%	2.7%	1%
	% of news based on interviews	7.6%	6.7%	7.4%
c) BBC News about legislation, a published statistic, an event or incident	Gov. policy and legis.	13.7%	11.7%	17%
	Non-res. statistics	5.7%	8%	1.3%
	News event or incident	49%	26.7%	20%
	% of news about legislation, statistics, events or incidents	68.4%	46.4%	38.3%
d) Not clear	Not clear	0.3%	0.3%	1.7%

RQ6: Highly cited BBC News stories

The sections below describe the five news stories that are most frequently cited from each source (see Tables B to D in the online appendix for lists, <https://dx.doi.org/10.6084/m9.figshare.4003359.v1>). This gives more detailed insights into the nature of important news stories. Politics of various forms is a common factor for many of them. In addition, whilst several are major policy initiatives or political pronouncements with wide significance, two are individual tragic incidents that serve as examples rather than being major events in their own right.

Frequently cited BBC News stories (Scopus)

State multiculturalism has failed says David Cameron (Scopus citations: 48; Google Books citations: 36; Wikipedia citations: 9): This contains text and a video of David Cameron's speech criticising "state multiculturalism" in the UK, arguing that a new vision to strengthen national identity was needed to prevent radicalization and extremism. Most citing articles cover political science, religion, or ethnic and cultural studies and the citing articles discuss cultural integration and multicultural policies in Europe. > Politics/Government policy change

Over 5 billion mobile phone connections worldwide (Scopus: 45; Google Books: 22; Wikipedia: 1): This gives statistics about mobile phone penetration in different regions. Most citing articles mention the headline statistic, using it as background context within computing or telecommunication journals. More than three-quarters (77%) of the citing publications were from 2011 and 2012, indicating that the news tended to attract most citations within two years after its BBC publication in July 2010, subsequently becoming obsolete. > Technology statistics

China overtakes Japan as world's second-biggest economy (Scopus: 37; Google Books: 15; Wikipedia: 3): Most citing articles were about economics, business, management and marketing, citing the news to discuss the high economic growth rate of China in recent years. > Economic statistic

S Korean dies after games session (Scopus: 35; Google Books: 18; Wikipedia: 6): This covers a young South Korean man dying after playing an online computer game for over two days. Most citing articles were

about psychological aspects of gaming addiction, disorder and behaviour, mainly published in psychology or psychiatry journals. The news was used as an extreme example to highlight the health consequences of long gaming sessions. > *Social example*

Merkel says German multicultural society has failed (*Scopus: 34; Google Books: 34; Wikipedia: 8*): This includes both text and a video of Chancellor Angela Merkel arguing that building a multicultural society in Germany has been completely unsuccessful and stating that immigrants should learn to speak German. Twelve Scopus publications had cited this news story alongside a similar argument by David Cameron (see above). > *Politics/Government policy change*

Frequently cited BBC News stories (Google Books)

Iraq war illegal, says Annan (*Google Books citations: 40; Scopus citations: 33; Wikipedia citations: 21*): This is a BBC interview with the former UN Secretary-General, Kofi Annan, admitting that the Iraq invasion was an illegal act that contravened the UN charter. Most citing books were about the Iraq war, international law, or the United Nations. Most books used this as an example of a violation of international law (e.g., *"The Cambridge Companion to International Criminal Law"*). > *Politics*

London bomber: Text in full (*Google Books: 36; Scopus: 33; Wikipedia: 2*): This includes the full transcript of a videotape of one of the 7 July London bombers giving reasons for the attack. Most citing books cover terrorism and radicalism, quoting some parts of his talk. Most citing articles used the news in same context as books and were published in terrorism journals (e.g., *Terrorism and Political Violence: 3 citations; Studies in Conflict and Terrorism: 2 citations*). > *Politics*

Rwanda: How the genocide happened (*Google Books: 34; Scopus: 11; Wikipedia: 33*): This reports that 800,000 people were killed in Rwanda's genocide between April and June 1994. This was mainly cited in books with international law, genocide, and human rights topics, referring to the scale of the mass murder. A few law journals also cited this, such as *Harvard Law Review, Journal of International Criminal Justice* and *Journal of International Humanitarian Legal Studies*. > *Genocide statistics*

Outrage at 'old Europe' remarks (*Google Books: 32; Scopus: 12; Wikipedia: 4*): This is about the reactions of French and German leaders when both countries were described as "problems" and representing "old Europe" in the Iraq war by a former US Defence Secretary. Most citing books were about European policy, the European Union, international relations and politics and cited this story as evidence of a European split over the Iraq war and in particular German and French opposition to the invasion of Iraq. > *Politics*

Saudi police 'stopped' fire rescue (*Google Books: 31; Scopus: 5; Wikipedia: 16*): This describes religious police preventing schoolgirls from exiting a burning school building because they were not wearing proper Islamic dress, causing 15 deaths. It was mainly cited in a human rights context, concerning Islamic sharia law and women's rights. > *Politics/religion example*

Frequently cited BBC News stories (Wikipedia)

Muslims in Europe: Country guide (*Wikipedia citations: 85; Google Books citations: 8; Scopus citations: 20*): This reports a census of Muslims across Europe, indicating that Islam is Europe's fastest growing religion. Most Wikipedia articles cited this to document the Muslim population or to reflect religious information across European countries in general entries about topics such as Islam in Europe or Islam in Denmark. However, this story was more specifically cited to discuss Muslim immigration and even Halal food consumption and marketing in food science journals. > *Religion statistics*

Russians left behind in Central Asia (Wikipedia: 74; Google Books: 9; Scopus: 4): This gives census records showing that there was a significant decline in the numbers of ethnic Russians living in Central Asian states after independence in 1991. Most Wikipedia articles cited this story to report census information in entries about the Central Asia region, countries and ethnic Russians. > *Geographic statistics*

Kosovo MPs proclaim independence (Wikipedia: 71; Google Books: 6; Scopus: 2): This covers Kosovo's independence from Serbia after Kosovo's parliament approved a declaration of independence. Most Wikipedia citing articles about Kosovo used this story to give historical information about the declaration. > *Political event*

Man dies after setting himself on fire in Saudi Arabia (Wikipedia: 46; Google Books: 2; Scopus: 2): This describes a Saudi man imitating the suicide protest of a Tunisian man, Mohamed Bouazizi, that triggered of the Arab Spring anti-government protests. Most citing Wikipedia articles cover the Arab Spring, protests in Saudi Arabia and Mohamed Bouazizi. > *Politics/religion example*

The girl who named a planet (Wikipedia: 41; Google Books: 3; Scopus: 1): This is an interview with 87-year-old Venetia Phair, who proposed the name "Pluto" for the ninth planet in the solar system in 1930 when she was 11 years old. The interview was mostly cited in entries about Venetia Phair, the planet Pluto and other astronomy contexts. > *Science social context*

Limitations

The main limitation of this article is its restriction to a single news source for most of the results. Additional data from other news organisations would have given better information about the extent of news use in academic research. The selection of the BBC also introduces an obvious bias towards UK news that would presumably not occur for non-UK news sources. Hence, a future study might assess citations to other news organisations, such as CNN and Reuters. Newspapers, both offline and in print editions, are also important sources of news information and may well be cited in a different way to the BBC News website. An additional important limitation is that most of the findings only concern broad disciplines and reveal little about the specific fields that cite the BBC.

A statistical limitation is that the correlation tests were conducted on a sample of articles with at least one Scopus citation. This was therefore a biased sample. Presumably, if a complete sample of BBC News stories had been used instead then the correlations would have been much higher due to many human interest stories of little lasting relevance that would not be cited outside of the BBC.

Discussion and Conclusions

The continued growth in the annual number of Scopus articles citing online BBC News stories (RQ1), approaching 3% of Arts and Humanities articles, shows that whilst citing news is not the norm, it is becoming accepted. It is therefore now an important scholarly communication phenomenon, although substantially less so in science and medicine.

The most cited area of the BBC News website was the UK section set, presumably because of the substantial UK coverage of the BBC as well as the importance of political news, which would fall within this group (RQ2). In the Science, Environment and Technology section, computing, the internet, climate change, global warming, astronomy and space tended to receive the most Scopus citations from academic publications. In Health, many Scopus citations were to stories about the health care system, especially in

the UK, although there were also citations to child health issues, drugs, medication, and cancer. Overall, however, a wide range of news topics are cited by academic research and it is neither restricted to a narrow area, nor to domains in which the press might be expected to have particular expertise or unique access to primary information. Thus, news websites seem to be sources of a wide range of types of information. Nevertheless, the themes within the frequently citing journals (RQ3) suggest that some topics, such as terrorism and theatre, are more important than others.

Surprisingly, half of the news in the Science and Health sections cited by Scopus were based upon previously published research (RQ5). Although it is not clear why scholars cite journalists reporting research instead of the original scientific articles or reports, the visibility of science news in the mass media and the simplified language of news stories about complex scientific topics might explain why researchers from other areas cite them (Brechman, Lee, & Cappella, 2009). Many science news articles do not mention their sources or link to them and this could prevent researchers from identifying the original studies (Mellor, Webster, & Bell, 2011; Taylor et al., 2015). In contrast, over two thirds of the Scopus-cited BBC News stories in the Arts, Entertainment & Education had non-research topics and the most frequently cited news in Scopus, Google Books and Wikipedia had non-scientific topics, such as incidents, legislation, published statistics/census, social examples and political speeches (RQ6).

The predominantly moderate positive correlations between Scopus and Google Books citations to BBC News stories in most of the news sections and year ranges (RQ4) suggests that books normally use news sources in similar ways to journal articles. This is perhaps surprising given the longer gestation period of books, which therefore seem likely to be interested mainly in news stories of longer term significance. The higher correlations in the Middle East and Africa sections suggests that book-based and article-based scholarship are particularly similar for these themes. The low correlations between Scopus and Wikipedia citations to BBC News in most news sections and years may be a statistical artefact of the low level of citation from Wikipedia, although it also seems possible that there are genuine differences.

In conclusion, news stories are used in different contexts to support discussions or claims in articles, books and encyclopaedia entries. Their value is in containing information about published statistics, events or exclusive interviews of lasting importance. Surprisingly, some news also seems to be useful for translating published research for other researchers, presumably from different areas. In this case, news stories may have added value by reporting additional sources, interviews or evidence that may give valuable supporting information. In this context, a concern is that the reporting accuracy standards of news organisations may be below those of academia, introducing different types of error (Chang, 2015). Even though many types of citations are not essential to the robustness of the citing work, particularly in the social sciences and humanities (Hyland, 2004), this is not true for all of them and so there is an urgent need to assess whether the apparent influx of journalistic information into the social sciences, arts and humanities is a cause for concern.

References

- Bartlett, C., Sterne, J., & Egger, M. (2002). What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers. *British Medical Journal*, 325(7355), 81-84.
- Begum, M., Pallari, E. & Lewison, G. (2016). European cancer research: from bench to bedside and to breakfast table. *ecancer*, 10(ed60). DOI: 10.3332/ecancer.
- Borgman, C. & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, Medford, NJ: Information Today Inc., pp. 3-72.

- Brechman, J., Lee, C., & Cappella, J. N. (2009). Lost in translation? A comparison of cancer-genetics reporting in the press release and its subsequent coverage in the press. *Science Communication*, 30(4), 453-474.
- Carlson, J. (2006). An examination of undergraduate student citation behavior. *Journal of Academic Librarianship*, 32(1), 14-22.
- Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51(7), 635-645.
- Chang, C. (2015). Inaccuracy in health research news: A typology and predictions of scientists' perceptions of the accuracy of research news. *Journal of Health Communication*, 20(2), 177-186.
- Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423-441.
- Creaser, C., Oppenheim, C., & Summers, M. A. C. (2011). What do UK academics cite? an analysis of references cited in UK scholarly outputs. *Scientometrics*, 86(3), 613-627.
- Cullars, J. (1992). Citation characteristics of monographs in the fine arts. *Library Quarterly*, 62(3), 325-342.
- Davis, P. M. (2002). The effect of the web on undergraduate citation behavior: A 2000 update. *College and Research Libraries*, 63(1), 53-60.
- Davis, P. M., & Cohen, S. A. (2001). The effect of the web on undergraduate citation behavior 1996-1999. *Journal of the American Society for Information Science and Technology*, 52(4), 309-314.
- De Tiratel, S. R. (2000). Accessing information use by humanists and social scientists: A study at the universidad de Buenos Aires, Argentina. *Journal of Academic Librarianship*, 26(5), 346-354.
- Entwistle, V. (1995). Reporting research in medical journals and newspapers. *BMJ*, 310(6984), 920.
- Fetahu, B., Anand, A., & Anand, A. (2015). How much is Wikipedia lagging behind news? In Proceedings of the ACM Web Science Conference (WebSci '15). ACM, New York, NY, USA, Article 28. DOI: <http://dx.doi.org/10.1145/2786451.2786460>
- Hammarfelt, B. (2012). Harvesting footnotes in a rural field: Citation patterns in Swedish literary studies. *Journal of Documentation*, 68(4), 536-558.
- Hanney, S., Grant, J., Jones, T., & Buxton, M. (2005). Categorising citations to trace research impact. In Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics. Stockholm: Karolinska University Press.
- Heinzkill, R. (2007). References in scholarly English and American literary journals thirty years later: A citation study. *College and Research Libraries*, 68(2), 141-153.
- Hyland, K. (2004). *Disciplinary Discourses, Michigan Classics Ed.: Social Interactions in Academic Writing*. Ann Arbor, MI: University of Michigan Press.
- Iroaganachi, M. A., Itsekor, V., & Osinulu, I. (2014). Citation analysis of social science research: A case study of bachelor degree research project reports of a Nigerian university 2009-2013. *Library Philosophy and Practice*, paper 1096, <http://digitalcommons.unl.edu/libphilprac/109>
- Jones, S. (2011). BBC Trust review of impartiality and accuracy of the BBC's coverage of science. *BBC Trust*. http://www.bbc.co.uk/bbctrust/our_work/editorial_standards/impartiality/science_impartiality.html.
- Kannappanavar, B., & Banateppanavar, K. (2013). The pattern of using information by research scholars in zoology: A bibliometric study. *Journal of Media and Communication Studies*, 5(9), pp. 137-154.
- Knievel, J. E., & Kellsey, C. (2005). Citation analysis for collection development: A comparative study of eight humanities fields. *Library Quarterly*, 75(2), 142-168.
- Kousha, K., & Thelwall, M. (2014). An automatic method for extracting citations from Google Books. *Journal of the Association for Information Science and Technology*, 66(2), 309-320.

- Kousha, K., & Thelwall, M. (2015). Web indicators for research evaluation. Part 3: books and nonstandard outputs, *El profesional de la información*, 24(6), 724-736.
- Kousha, K., & Thelwall, M. (2016). Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology*. DOI: 10.1002/asi.23694
- Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lewison, G. (2008). The reporting of the risks from severe acute respiratory syndrome (SARS) in the news media, 2003-2004. *Health, Risk and Society*, 10(3), 241-262.
- Lewison, G., & Turnbull, T. (2010). News in brief and features in new scientist magazine and the biomedical research papers that they cite, august 2008 to July 2009. *Scientometrics*, 85(1), 345-359.
- Lewison, G., Tootell, S., Roe, P., & Sullivan, R. (2008). How do the media report cancer research? A study of the UK's BBC website. *British Journal of Cancer*, 99(4), 569-576.
- Lih, A. (2004). Wikipedia as Participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a news resource. In *Proceedings of the 5th International Symposium on Online Journalism*. <http://www.ufrgs.br/limc/participativo/pdf/wikipedia.pdf>
- MacDonald, M. M., & Hoffman-Goetz, L. (2002). A retrospective study of the accuracy of cancer information in Ontario daily newspapers. *Canadian Journal of Public Health*, 93(2), 142-245.
- MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1-12.
- Mellor, F., Webster, S., & Bell, A. R. (2011) Content analysis of the BBC's science coverage. Science Communication Group, Imperial College London. http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/our_work/science_impartiality/appendix_a.pdf
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago, IL: University of Chicago Press.
- Moed, H. F., & Visser, M. S. (2008). Appraisal of citation data sources. Centre for Science and Technology Studies, Leiden University. <http://www.hefce.ac.uk/media/hefce/content/pubs/indirreports/2008/missing/Appraisal%20of%20Citation%20Data%20Sources.pdf>
- Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social studies of science*, 5(1), 86-92.
- Oermann, M. H., Nordstrom, C. K., Wilmes, N. A., Denison, D., Webb, S. A., Featherston, D. E., Bednarza, H., & Striz, P. (2008). Information sources for developing the nursing literature. *International Journal of Nursing Studies*, 45(4), 580-587.
- Ofcom (2015). News consumption in the UK: 2015 report. Ofcom <https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/news-consumption-2015>
- Taylor, J. W., Long, M., Ashley, E., Denning, A., Gout, B., Hansen, K., . . . Newton, P. M. (2015). When medical news comes from press releases - A case study of pancreatic cancer and processed meat. *PLoS ONE*, 10(6).
- Ucak, N. O., & Al, U. (2009). The differences among disciplines in scholarly communication. A bibliometric analysis of theses. *Libri*, 59(3), 166-179.
- Wainer, J., Przebisczki De Oliveira, H., & Anido, R. (2010). Patterns of bibliographic references in the ACM published papers. *Information Processing and Management*, 47(1), 135-142.
- Yang, S., Qiu, J., & Xiong, Z. (2010). An empirical study on the utilization of web academic resources in humanities and social sciences based on web citations. *Scientometrics*, 84(1), 1-19.

Zainab, A. N., & Goi, S. S. (1997). Characteristics of citations used by humanities researchers. *Malaysian Journal of Library and Information Science*, 2(2), 19-36.