How are encyclopedias cited in academic research? *Wikipedia*, *Britannica*, *Baidu Baike*, and *Scholarpedia*

Xuemei Li; **Mike Thelwall**; **Ehsan Mohammadi**

How to cite this article:

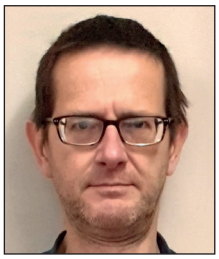
Li, Xuemei; **Thelwall, Mike**; **Mohammadi, Ehsan** (2021). "How are encyclopedias cited in academic research? *Wikipedia*, *Britannica*, *Baidu Baike*, and *Scholarpedia*". *Profesional de la información*, v. 30, n. 5, e300508.

https://doi.org/10.3145/epi.2021.sep.08

Manuscript received on 11th June 2020
Accepted on 02th August 2021

Xuemei Li ✉
https://orcid.org/0000-0002-3384-1987

York University Libraries
Open Scholarship Department
4700 Keel Street, Toronto, Ontario,
Canada M3J 1P3
lixuemei@yorku.ca



Mike Thelwall
https://orcid.org/0000-0001-6065-205X

University of Wolverhampton
Statistical Cybermetrics Research Group
Wulfruna Street
Wolverhampton WV1 1LY, UK
m.thelwall@wlv.ac.uk

Ehsan Mohammadi
https://orcid.org/0000-0003-3481-6991

University of South Carolina
College of Information and Communications
1501 Greene Street, Columbia
SC 29208, USA
ehsan2@mailbox.sc.edu

Abstract

Encyclopedias are sometimes cited by scholarly publications, despite concerns about their credibility as sources for academic information. This study investigates trends from 2002 to 2020 in citing two crowdsourced and two expert-based encyclopedias to investigate whether they fit differently into the research landscape: *Wikipedia*, *Britannica*, *Baidu Baike*, and *Scholarpedia*. This is the first systematic comparison of the uptake of four major encyclopedias within academic research. *Scopus* searches were used to count the number of documents citing the four encyclopedias in each year. *Wikipedia* was by far the most cited encyclopedia, with up to 1% of *Scopus* documents citing it in Computer Science. Citations to *Wikipedia* increased exponentially until 2010, then slowed down and started to decrease. Both the *Britannica* and *Scholarpedia* citation rates were increasing in 2020, however. Disciplinary and national differences include *Britannica* being popular in Arts and Humanities, *Scholarpedia* in Neuroscience, and *Baidu Baike* in Chinese-speaking countries/territories. The results confirm that encyclopedias have minor value for academic research, often for background and definitions, with the most suitable one varying between fields and countries, and with the first evidence that the popularity of crowdsourced encyclopedias may be waning.

Keywords

Wikipedia; *Britannica*; *Baidu Baike*; *Scholarpedia*; Encyclopedias; *Scopus*; Citation analysis; Open access; Scholarly communication; Multidisciplinary; Countries.

1. Introduction

Crowdsourced encyclopedias are sometimes cited in academic research. This practice can be controversial because, unlike traditional scholarly encyclopedias (e.g., **Tomaszewski**, 2018), crowdsourced sites such as *Wikipedia* are publicly editable and dynamic and so do not provide the permanence and authority usually required for references in academic publications. Other encyclopedias may also not be subject to full academic rigor, even if subject to editorial review and written by invited experts. Nevertheless, since *Wikipedia* and other major encyclopedias seem to be usually accurate, it is arguably reasonable to cite them to guide article readers to useful background reading that does not underpin the logic of an article. There is no recent information about whether citations to *Wikipedia* or other encyclopedias are increasing, however, and whether they are drawn upon in different ways. This information is needed to help librarians, authors, and referees to understand the research contribution that a major encyclopedia can make, if any.

Wikipedia has become popular for both education and research since it started in 2001 (**Kousha**; **Thelwall**, 2017; **Mesgari** *et al.*, 2015; **Okoli** *et al.*, 2014), although the authority of its content has been repeatedly scrutinized for accuracy and coverage (**Chesney**, 2006; **Giles**, 2005; **Holman-Rector**, 2008; **Jullien**, 2012; **Messner**; **DiStaso**, 2013; **Okoli** *et al.*, 2012; **Samoilenko**; **Yasseri**, 2014). One advantage of *Wikipedia* is its ability to react quickly to new issues, such as Covid-19 (**Colavizza**, 2020). Since *Wikipedia* summarizes knowledge for a general audience, often supported by references, it has been used as evidence of the wider impact of academic research to complement traditional citation-based indicators (**Jemielniak**; **Masukume**; **Wilamowski**, 2019; **Kousha**; **Thelwall**, 2017; **Lin**; **Fenner**, 2014; **Priem**; **Piwowar**; **Hemminger**, 2012). *Wikipedia* is usually cited to provide general information or a definition (**Tohidinasab**; **Jamali**, 2013). While citations from *Wikipedia* to academic publications may reflect knowledge transfer from academia to a wider public domain, or can be used to verify its information, citations in the reverse direction from academic publications to *Wikipedia* are more controversial (**Fallis**, 2008). Despite this, the number of academic citations to *Wikipedia* increased annually until at least 2015, including from reputable publications and traditional non-OA articles (**Tomaszewski**; **MacDonald**, 2018).

A few studies have investigated citations to *Wikipedia* from academic publications, starting with an investigation of the quality of *Wikipedia* articles that had been cited in online news stories (**Lih**, 2004). Some research has had a subject focus, investigating *Wikipedia* citations from law reviews (**Baker**, 2011; **Shoyama**, 2014), chemistry journals from three major publishers (**Brazzeal**, 2011), and health science journals in *Medline*, *PubMed*, or *Embase* (**Bould** *et al.*, 2014). These studies have found that citations to *Wikipedia* were increasing over time, but there is disagreement over whether it is reliable and whether it should be cited by prestigious journals, including *Nature*, *Science*, and the *BMJ*. Both academic publications about *Wikipedia* and citations to *Wikipedia* in general increased over time shortly after its appearance (**Huggett**, 2012; **Park**, 2011), although it is not known if this has increased in the last decade.

In contrast to *Wikipedia*, *Encyclopedia Britannica* is a well-known, centuries-old English-language encyclopedia that seems to have a reputation for scholarly authority. *Scholarpedia*, a peer-reviewed free online encyclopedia that started in 2006 (**Izhikevich**, 2006), was the closest rival to *Wikipedia* in 2011 but received only 1/20 as many citations (**Huggett**, 2012). *Baidu Baike* is the most popular online Chinese general encyclopedia. It started in 2006 and a few studies have investigated citations from Chinese academic articles to it using the Chinese article index databases *CNKI* (2019) or *CSSCI* (*Chinese Social Sciences Citation Index*, 2015) (**Ding**; **Zhang**; **Liu**, 2013; **Wang**, 2016). Nevertheless, no studies have investigated how *Baidu Baike* is cited in *Scopus*-indexed articles.

This study compares citations from *Scopus*-indexed documents to *Wikipedia*, *Britannica*, *Baidu Baike* and *Scholarpedia*, with the objective of assessing whether changes over time and differences between them can give insights into the different roles that encyclopedias can play in academic research. Whilst there are many other encyclopedias that could have been included, these four have differences that may be illuminating. *Wikipedia* and *Baidu Baike* are crowdsourced encyclopedias. Whilst *Wikipedia* is multilingual, *Baidu Baike* is in Chinese. *Britannica* is an English peer-reviewed proprietary general encyclopedia with some content freely available online. *Scholarpedia* is an English peer-reviewed free online encyclopedia with substantial coverage of Astrophysics, Celestial mechanics, Computational neuroscience, Computational intelligence, Dynamical systems, Physics and Somatosensory systems. This article compares the numbers of citing documents to these four encyclopedias, breaking down the results by subject, language, publication type and author country.

2. Research questions

This project assesses trends in the uptake of four major encyclopedias in formal scholarly communication, and whether their uses are affected by characteristics of the citing documents and their authors. The following questions drive the investigation.

- Is the level of academic citing of *Wikipedia*, *Britannica*, *Baidu Baike* and *Scholarpedia* increasing over time?
- Which fields most cite the four encyclopedias?
- Are there differences between encyclopedias in citing document subjects, types (e.g., open access, journal articles, or books), publication languages, or author characteristics (e.g., national affiliations)?

3. Methods

The evidence used to address the above questions is taken from explicit mentions of the four encyclopedias in academic literature reference lists. *Scopus* was chosen to count how many documents cite the four encyclopedias because *Scopus* covers more publications than does the *Web of Science* (*WoS*) and also it allows more comprehensive searches within the cited reference fields (**Kousha**; **Thelwall**; **Abdoli**, 2012; **Li**; **Thelwall**; **Kousha**, 2015).

Since *Wikipedia* was launched in 2001, citations to *Wikipedia* and *Britannica* were counted from 2002 to 2020. Both *Baidu Baike* and *Scholarpedia* were launched in 2006 and so citations to these two encyclopedias were counted from 2007 to 2020. These searches were run on 7 July 2021 so that the full set of documents from 2002 to 2020 should be included. Appendix I lists the *Scopus* search syntax used for the four encyclopedias.

The example below is the *Scopus* syntax for searching the citing documents to *Wikipedia* in Computer Science from 2002 to 2020.

> (REF("*wikipedia.org/w*") OR REFSRCTITLE("*wikipedia*")) AND ((PUBYEAR > 2001) AND (PUBYEAR < 2021)) AND SUBJAREA(COMP)

The queries sometimes gave a few irrelevant matches, based on a check of 100 random citing documents to each of the four encyclopedias. The *Wikipedia* queries generated one false match, *Baidu Baike* got one while *Britannica* got three – All the false matches were caused because the citing documents cite articles with the relevant encyclopedia names in titles. The *Scholarpedia* queries returned one false match ("Perception of surface stickiness in different sensory modalities: An functional MRI study") using the original syntax:

> (REF("*scholarpedia.org/article*") OR REFSRCTITLE(scholarpedia)) AND ((PUBYEAR > 2006) AND (PUBYEAR < 2019))

This false citing document cites one article that was published in an encyclopedia book "Scholarpedia of Touch". We amended the syntax as below:

> (REF("*scholarpedia.org/article*") OR REFSRCTITLE(scholarpedia) AND NOT REFSRCTITLE("scholarpedia of touch")) AND ((PUBYEAR > 2006) AND (PUBYEAR < 2021))

The new *Scholarpedia* search syntax did not generate any false matches in the subsequent 100 random citing documents checking. The search syntax for the other three encyclopedias remains unchanged to ensure that valid results are not excluded.

Scopus classifies articles into broad or narrow fields based on the journal in which they are published, except for general journals. This is a limitation because an article may be published within an interdisciplinary journal or an out of field journal and receive an inappropriate subject classification. This is not expected to be a substantial problem at the aggregate level reported here.

4. Results and discussion

Altogether *Wikipedia* was cited by 141,991 *Scopus* indexed documents (2002 to 2020), *Britannica* 15,929 (2002 to 2020), *Baidu Baike* 2,934 (2007 to 2020) and *Scholarpedia* 8,399 (2007 to 2020). The proportions of documents citing each of the four encyclopedias have increased over time, with *Wikipedia* being by far the most cited (Figure 1). The proportions of articles citing the crowdsourced encyclopedias, *Wikipedia* and *Baidu Baike*, have stabilized since 2012, however, with a slight decreasing trend from 2013 that seems to have accelerated in 2020. The decrease might be due to stricter editorial policies or a wider recognition of the dangers of citing unstable, editable sources. *Wikipedia* is now substantially more cited than the other three encyclopedias, having overtaken *Britannica* in 2005.

The continuing popularity of *Britannica*, despite its open access competitors, is partly due to citations to old editions for established procedures (e.g., the 2018 article, "Distribution of runs of homozygosity in Chinese and Western pig breeds evaluated by reduced-representation sequencing data" cites a 1948 edition for a genetic formula) and for historical references (e.g., "The Chiropractic Vertebral Subluxation Part 3: Complexity and Identity From 1908 to 1915" cites the 1902 edition to help explain where a scientist got his knowledge from). It is presumably also useful as a relatively scholarly source of definitions or background information (e.g., the first sentence of the introduction of "Ingestion of microplastics by some commercial fishes in the lower Gulf of Thailand: A preliminary approach to ocean conservation" is "Plastic pollution is the gathering of plastic substances in the environments which have several hostile effects on wildlife, wildlife habitat as well as on human beings" with a citation to the *Britannica* entry on plastic pollution). *Britannica* references also have a more scholarly structure, half of the time including the contribution authors (which never occurs for *Wikipedia*), but usually also a *Britannica* online URL. For example, a *Journal of Aesthetics and Art Criticism* discussion of satire cited the *Britannica* article on the topic, mentioning its author, R.C. Elliot, and URL *https://www.britannica.com/art/satire*

> *Wikipedia* was the most cited encyclopedia, with a maximum of 1% of *Scopus* documents citing it in Computer Science

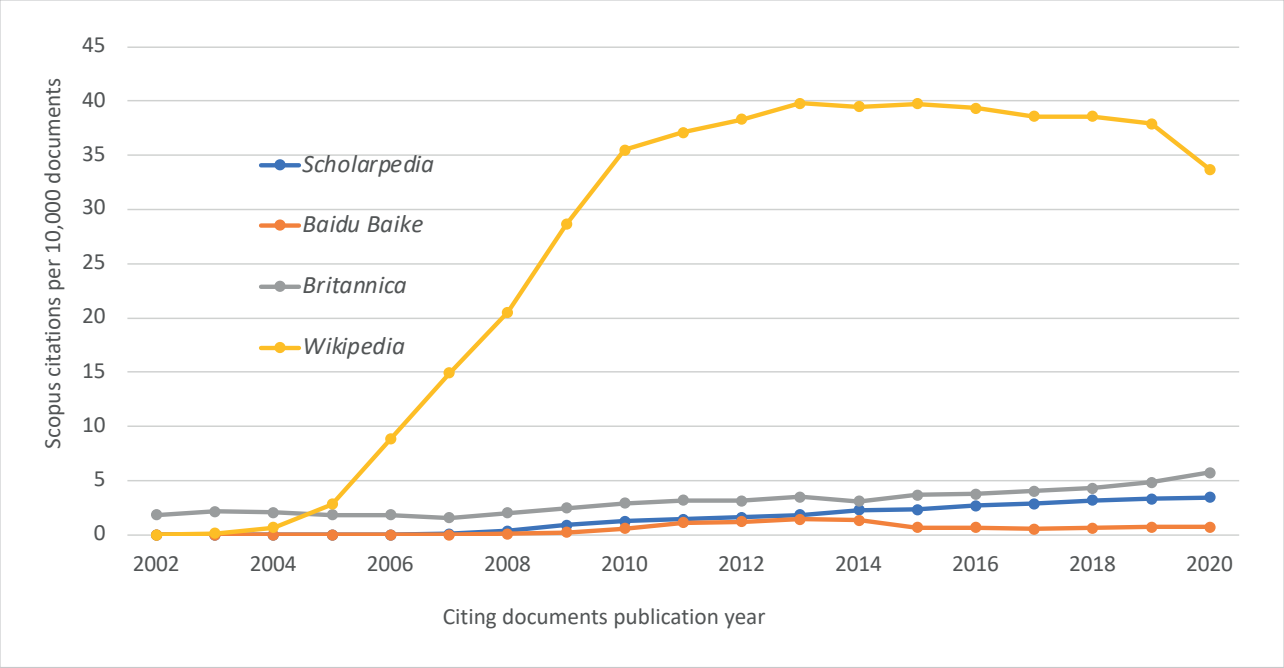


Figure 1. *Scopus* documents (per 10,000) citing *Wikipedia*, *Britannica*, *Baidu Baike* and *Scholarpedia*

Both expert-authored encyclopedias are increasing their relative number of citations. Whilst this is perhaps unsurprising for the relatively new *Scholarpedia*, the increase is unexpected for *Britannica*. The increase for *Britannica* may be partly due to reluctance to cite *Wikipedia*, which it may be starting to replace. It may also reflect *Britannica*'s decision to be online only in 2012 or the success of specific sales or marketing initiatives.

4.1. The main citing fields for each encyclopedia

Despite the overall dominance of *Wikipedia*, there are substantial disciplinary differences in citing the four encyclopedias (Table 1 and Figure 2). From the uneven bar sizes for many fields and the top 10 most cited fields, it is clear that the encyclopedias have substantially different rates of use. Some notable examples are singled out here for comment.

- *Scholarpedia* is very highly used in Neuroscience, highly used in Mathematics and also in Physics and Astronomy. In these three areas it attracts disproportionately many citations from *Scopus* articles compared to the other encyclopedias (as a % share of their cited documents). Thus, it has clear fields in which it is a highly used resource.
- *Britannica* is highly used in Arts and Humanities and Social Sciences, compared to the other encyclopedias. Articles in these two subject areas may reference facts or definitions for historical or cultural topics, such as a citation to a *Britannica* article on the Palmer raids 1919-22 in the USA.
- *Baidu Baike* attracts relatively many citations from Engineering and *Wikipedia* is not disproportionately cited in any particular field.

Some fields have disproportionately high or low encyclopedia use.

- Medicine is the main field in which all the encyclopedias are comparatively rarely cited. The related field of Biochemistry, Genetics and Molecular Biology also rarely cites encyclopedias. These areas might have fewer definitions to cite, may need more scholarly sources of citations, or may include less background information in their articles. This finding is surprising given that *Wikipedia* is known to have good coverage of these areas (**Arroyo-Machado** *et al.*, 2020).
- Materials Science comparatively rarely cites encyclopedias, although it is not clear why.
- Computer Science has relatively many citations to *Wikipedia*, *Baidu Baike* and *Scholarpedia*. Presumably this is for standard definitions or background for computing terminology. For example, both *Baidu Baike* and *Wikipedia* pages on the Internet of Things (IoT) were frequently cited. The more traditional *Britannica* would presumably be less able to keep up with modern information technology changes. Its 2015 IoT article was not cited, although its internet article from 1998 received one citation.
- Decision Sciences is a relatively heavy citer of all four encyclopedias, suggesting that this subject has a particularly strong need to cite reference works, perhaps for standard mathematical or statistical formulae or definitions. For example, the *Baidu Baike* article on affine transformations was cited four times.
- Social Sciences and Business, Management and Accounting are high citing areas for *Wikipedia*, *Britannica* and *Baidu Baike*. Articles in these two subject areas may reference facts or definitions for topics not covered well by *Scholarpedia*. For example, the topics of the *Wikipedia* pages cited included the demographics of Russia, Hurricane Maria, and Rohingya refugees in Bangladesh.

Table 1. The top 10 subjects with the highest proportions of documents citing the four encyclopedias (per 10,000 *Scopus* indexed documents)

Wikipedia		***Britannica***		***Baidu Baike***		***Scholarpedia***	
Computer Science	100.2	Arts and Humanities	20.0	Computer Science	1.8	Neuroscience	14.6
Decision Sciences	83.2	Social Sciences	12.3	Social Sciences	1.5	Computer Science	6.7
Social Sciences	54.6	Economics, Econometrics	9.2	Engineering	1.5	Mathematics	6.7
Business, Management, & Accounting	49.2	Business, Management, & Accounting	7.5	Decision Sciences	1.4	Psychology	4.8
Mathematics	46.8	Psychology	5.1	Business, Management, & Accounting	1.4	Decision Sciences	4.3
Arts and Humanities	43.6	Decision Sciences	4.3	Environmental Sciences	1.0	Physics and Astronomy	4.1
Engineering	43.4	Environmental Sciences	4.2	Arts and Humanities	1.0	Multidisciplinary	3.4
Energy	39.7	Earth and Planetary Sciences	3.6	Energy	0.9	Engineering	2.4
Economics, Econometrics	36.6	Agricultural and Biological Sciences	3.1	Economics, Econometrics	0.9	Arts and Humanities	1.9
Environmental Sciences	25.7	Computer Science	2.9	Mathematics	0.9	Biochemistry, Genetics and Molecular Biology	1.6

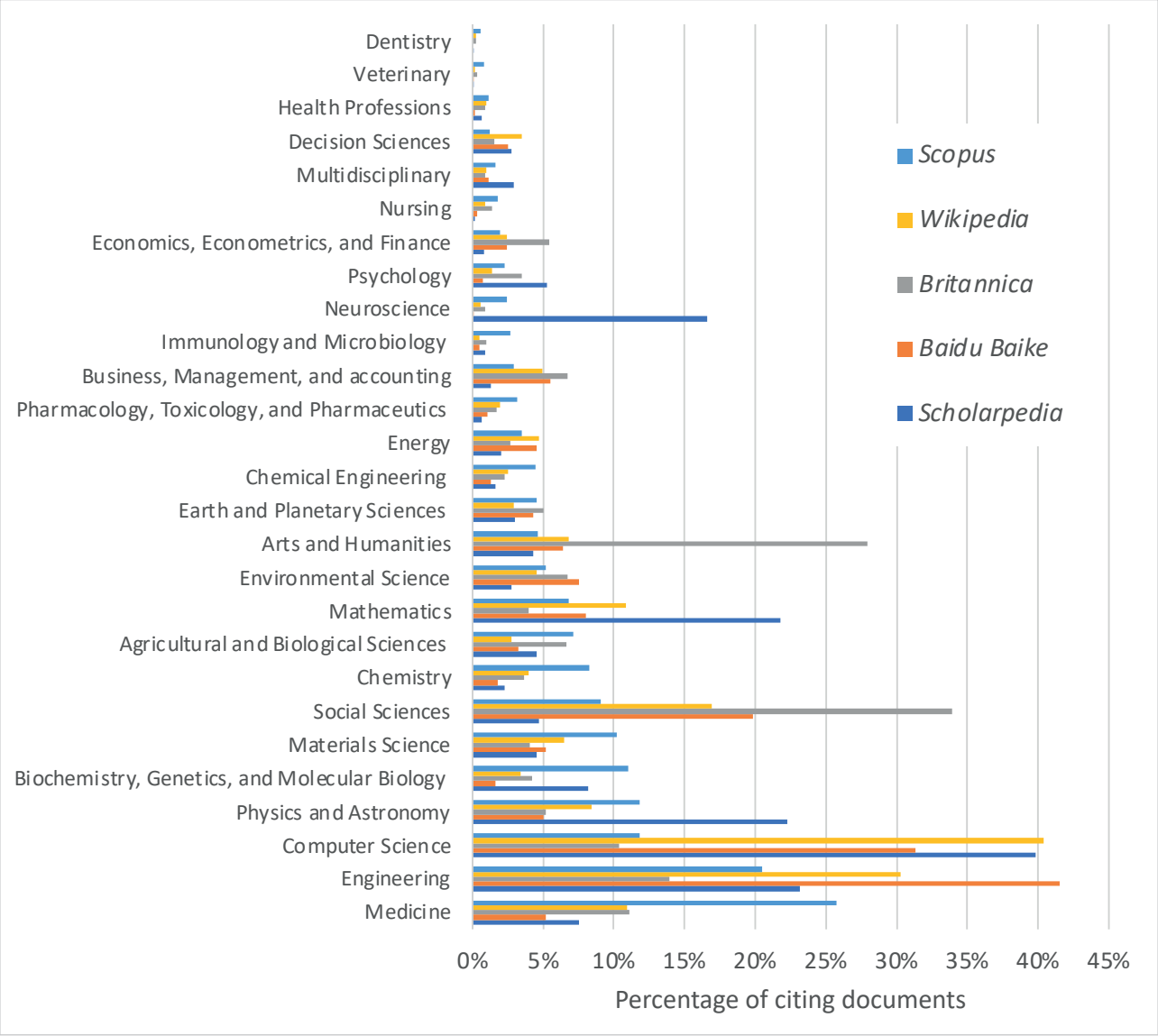


Figure 2. *Scopus* citations to *Wikipedia*, *Britannica*, *Baidu Baike* and *Scholarpedia* against total number of *Scopus* documents indexed (2002-2020). Fields (n=27) are listed in ascending order of size in *Scopus*. The percentages are calculated out of the total data for each source.

4.2. Types of documents citing each encyclopedia

Information about the types of documents citing encyclopedias may give deeper insights into who cites them and why. Open Access (OA) citing documents in *Scopus* are more likely to cite *Scholarpedia* but less likely to cite the other three encyclopedias (Figure 3), perhaps surprisingly given the open access credentials of three of them. Just under half of the citations are from Gold or Hybrid Gold journals, with the remainder from authors publishing their versions online (Green OA) or the publishers making a version of the manuscript available temporarily (Bronze OA). These figures are estimates from *Scopus* since articles can fit in multiple categories. For the purposes of the diagram, Green OA articles that are also Gold or Hybrid Gold or Bronze are not included in the Green figures (see Appendix II for the search syntax).

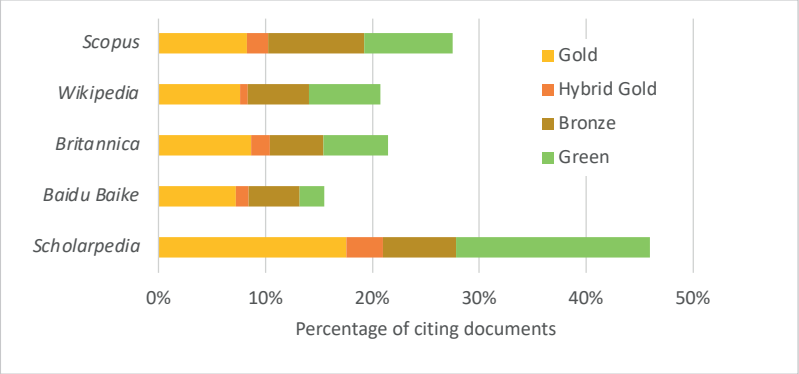


Figure 3. Open access percentages for all *Scopus* indexed and citing documents

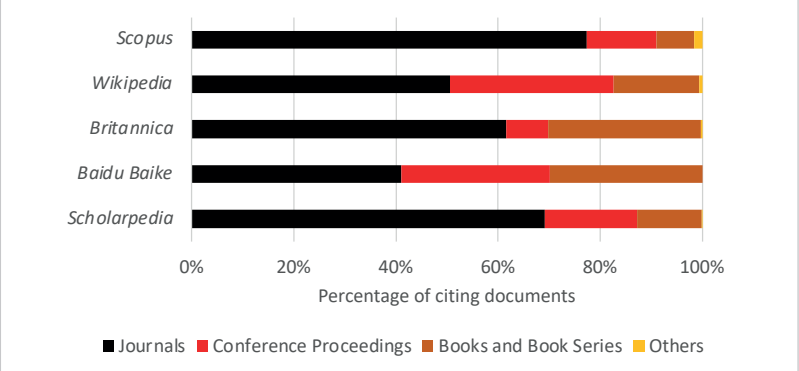


Figure 4. Percentages of publication types for all *Scopus* indexed and citing documents

As illustrated in Figure 4, books and book series in *Scopus* are more likely to cite all encyclopedias. Conference papers published in proceedings are more likely than other *Scopus* documents to cite all encyclopedias except *Britannica*. Since conference papers are important in fast-moving subjects like computing, the low proportion of *Britannica* citations is unsurprising. Nevertheless, this suggests that journal articles (and other documents published in journals) are less likely to cite encyclopedias than other academic document types. This may reflect different attitudes of journal editors/peer reviewers/authors towards encyclopedias or more stringent peer review for journal articles.

4.3. Languages and countries citing each encyclopedia

English language documents in *Scopus* are more likely to cite the three non-Chinese encyclopedias than are non-English documents in *Scopus* (Table 2). Chinese documents in *Scopus* are substantially more likely to cite Chinese language encyclopedia *Baidu Baike*. Even though *Wikipedia* is multilingual (alone of the four encyclopedias), it is most cited in English and with no other common language citing it much. The slight tendency for Portuguese language documents to cite English-language *Britannica* (nearly twice as much as *Wikipedia*) is an anomaly, although there are historical connections between the UK and Portugal.

Table 2. The most common 10 languages for documents citing the four encyclopedias. Bold languages are the main ones above the *Scopus* average

Scopus	**%**	***Wikipedia***	**%**	***Britannica***	**%**	***Baidu Baike***	**%**	***Scholarpedia***	**%**
English	90.8	**English**	95.4	**English**	95.9	English	88.4	**English**	98.5
Chinese	3.5	German	1.1	Spanish	1.0	**Chinese**	11.1	Chinese	0.8
German	1.3	Spanish	0.8	Portuguese	0.5	German	0.1	Russian	0.2
French	1.1	Chinese	0.7	German	0.5	Russian	0.1	Spanish	0.1
Spanish	1.0	French	0.5	French	0.4	French	0.1	Turkish	0.1
Russian	0.6	Russian	0.3	Croatian	0.4	Spanish	0.1	German	0.1
Japanese	0.6	Portuguese	0.3	Russian	0.3	Bulgarian	0.0	Portuguese	0.1
Portuguese	0.5	Turkish	0.2	Italian	0.2	Italian	0.0	French	0.1
Italian	0.3	Polish	0.2	Slovenian	0.2	Korean	0.0	Italian	0.0
Polish	0.2	Croatian	0.2	Chinese	0.2	Polish	0.0	Japanese	0.0

Documents from the USA are more likely to cite *Wikipedia* and *Britannica* (Table 3). China dominates *Baidu Baike* citations, presumably because of its language, and other Chinese-speaking countries/territories also disproportionately cite it. *Scholarpedia* citations are disproportionately from Germany, despite its origins in the USA. This may be a topic issue, if Neuroscience is a German specialty.

Surprisingly, the UK does not cite *Wikipedia* disproportionately often. India, South Korea, Australia and Taiwan testify to the international credibility of *Wikipedia*, although all four countries presumably publish most academic research in English (indexed in *Scopus*) and have historical connections to the UK or USA. *Britannica*, originally from the UK but published in the USA since 1901 (*Encyclopædia Britannica*, 2019), is relatively highly cited only in countries with a historical connection to the UK.

Table 3. The most common 10 countries/territories for documents citing the four encyclopedias. Countries/territories are bold when they are the main country above the *Scopus* average

Scopus	%	*Wikipedia*	%	*Britannica*	%	*Baidu Baike*	%	*Scholarpedia*	%
USA	23.9	**USA**	25.5	**USA**	29.1	**China**	86.9	USA	26.4
China	14.7	China	12.1	UK	9.6	USA	8.8	China	13.1
UK	7.1	India	10.9	India	4.7	Hong Kong	2.4	**Germany**	11.3
Germany	6.1	UK	5.7	Canada	4.6	Australia	2.3	UK	10.6
Japan	5.1	Germany	4.0	Germany	4.4	UK	2.3	France	7.5
France	4.2	South Korea	3.7	Australia	3.7	Canada	1.4	Italy	5.5
India	4.1	Canada	3.4	China	3.0	Taiwan	1.3	India	5.2
Italy	3.7	Australia	3.1	Italy	2.7	Singapore	0.7	Canada	4.4
Canada	3.6	Taiwan	2.4	France	2.1	South Korea	0.6	Japan	3.8
Australia	3	Japan	2.2	Spain	1.9	Japan	0.6	Spain	3.7

5. Limitations

This study relies on *Scopus* advanced search facilities for data gathering. As a result, it is limited by *Scopus*'s coverage, classification schemes and search efficiencies. For example, *Scopus* covers more publications than does the *Web of Science*, but it does not cover all scholarly publications, and it may miss book chapters and Chinese journals which are important sources that cite encyclopedias. This may explain why *Scopus* citations to *Baidu Baike* are much fewer than those counting from Chinese article index databases (**Ding** *et al.*, 2013; **Wang**, 2016).

> Peer-reviewed encyclopedia *Britannica* citations continue to be valuable in the Arts and Humanities, and *Scholarpedia* citations in Neuroscience

Although in collaboration with *Impactstory*, *Scopus* may miss open access documents in its search results as it only sourced open access documents that are harvested by *Unpaywall* –a database run by *Impactstory* (*Scopus: Access and use Support Center*, 2021).

Finally, it is important to be cautious when interpreting the results of citing patterns by subjects because *Scopus* does not index all scholarly publications and its comprehensiveness varies between fields.

6. Conclusions

Citing encyclopedias is rare in all academic subjects. Although *Wikipedia* is the most cited encyclopedia, according to *Scopus* data, in the 27 subjects investigated it occurred in a maximum of 1% of *Scopus* documents citing it in Computer Science.

In answer to Question 1:

For the two free publicly editable encyclopedias: citations to *Wikipedia* increased exponentially until 2010, when the rate of increase slowed down and then started to decrease in 2020 (updating a previous study showing citations continuing to increase: **Tomaszewski** & **MacDonald**, 2018) while citations to *Baidu Baike* decreased substantially after 2013.

For the two peer-reviewed encyclopedias: citations to *Britannica* relatively stabilized over the years while citations to *Scholarpedia* increasing gradually since it started in 2006. Thus, with the possible exception of Neuroscience, citations to major encyclopedias should continue to be rare in the future and librarians/authors/reviewers should not expect to see or use them other than in exceptional circumstances.

They may also see a partial reversion from crowdsourced encyclopedias to expert-written versions. In the context of apparently increasing public scrutiny of academic research (e.g., during the Covid-19 pandemic), it seems particularly important to ensure that all citations are robust. In this context, a citation to a crowdsourced encyclopedia may be a weak point in an article that may be exploited, particularly if the article covers a controversial topic.

In answer to Question 2:

There are substantial disciplinary differences in the uptake of the four encyclopedias, and they have particularly little value in Medicine. The four encyclopedias seem to be particularly useful in mathematical areas, such as

> Unsurprisingly, *Baidu Baike* is disproportionately cited by Chinese-speaking countries/territories

Decision Sciences, presumably for definitions, and in fast moving technological areas, such as Computing, presumably also for definitions as well as explanations of new technological developments. In contrast, *Britannica* citations continue to be valuable in the Social Sciences and Arts and Humanities. Its relative lack of citations in other areas may reflect its targeting of a general audience for which detailed scientific explanations and coverage would be inappropriate and not cost-effective to curate. This seems like a niche that crowdsourced encyclopedias will continue to fill. *Scholarpedia* provides a partial exception to this, by providing detailed peer reviewed expert articles on highly scientific topics, although its coverage is limited and it is not clear if the model is sustainable across academia. Scholars and librarians that (occasionally) need to cite encyclopedias may therefore consider first checking the one most used in their fields.

> Encyclopedias have minor value for academic research, often for background and definition purposes

In answer to Question 3:

Books and book series in *Scopus* are more likely to cite the four encyclopedias than for general *Scopus*-indexed documents. This may be due to less strict refereeing for books or more encyclopedia-like content in books (e.g., handbooks). Open access citing documents are more likely to cite *Scholarpedia* but less likely to cite the other three counterparts. This updates a previous study of *Wikipedia* that did not find OA publications to be the main sources of scholarly citations to *Wikipedia* (**Tomaszewski**; **MacDonald**, 2018). Chinese documents are more likely to cite *Baidu Baike* while English documents are more likely to cite the other three non-Chinese encyclopedias. Unsurprising, *Baidu Baike* is highly cited by Chinese-speaking countries/territories, US documents are more likely to cite the three non-Chinese encyclopedias while *Scholarpedia* is more cited by nearly all the major publishing nations except China and Japan. Thus, the rate of citing encyclopedias varies between countries and document types, suggesting that scholars choose sources that are known to them rather than selecting the best source for any particular citation.

In summary, encyclopedias are continuing to play a minor role in formal scholarly communication, in the form of references. National factors play a role, with authors being more likely to select encyclopedias that are popular in their countries. Despite the criticism of the open editing formats of *Wikipedia* and *Baidu Baike*, each of the four major encyclopedias investigated seems to have found a niche. It is not clear whether the open encyclopedias are cited by scholars that are aware of and accept their limitations, however, or whether in the case of the two unrefereed sources, there are mistaken attempts to underpin research with unstable sources. This is an important issue for the scholarly community, and one that editors and reviewers should monitor.

> Open-access citing documents in *Scopus* are more likely to cite *Scholarpedia* but less likely to cite the other three encyclopedias

7. References

Arroyo-Machado, Wenceslao; **Torres-Salinas, Daniel**; **Herrera-Viedma, Enrique**; **Romero-Frías, Esteban** (2020). "Science through Wikipedia: A novel representation of open knowledge through co-citation networks". *PloS one*, v. 15, n. 2, e0228713.
https://doi.org/10.1371/journal.pone.0228713

Baker, Daniel J. (2011). "A Jester's promenade: Citations to Wikipedia in law reviews, 2002-2008". *Journal of law and policy for the information society*, v. 7, n. 2, pp. 361-404.
https://doi.org/10.2139/ssrn.1525619

Bould, M. Dylan; **Hladkowicz, Emily S.**; **Pigford, Ashlee-Ann E.**; **Ufholz, Lee-Anne**; **Postonogova, Tatyana**; **Shin, Eunkyung**; **Boet, Sylvain** (2014). "References that anyone can edit: Review of Wikipedia citations in peer reviewed health science literature". *BMJ*, v. 348, g1585.
https://doi.org/10.1136/bmj.g1585

Brazzeal, Bradley (2011). "Citations to Wikipedia in chemistry journals: A preliminary study". *Issues in science and technology librarianship*, v. Fall.
https://doi.org/10.5062/F4057CV7

Chesney, Thomas (2006). "An empirical examination of Wikipedia's credibility". *First Monday*, v. 11, n. 11.
https://doi.org/10.5210/fm.v11i11.1413

Chinese Social Sciences Citation Index (2015). *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Chinese_Social_Sciences_Citation_Index&oldid=682574308

CNKI (2019). *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=CNKI&oldid=913654910

Colavizza, Giovanni (2020). "Covid-19 research in Wikipedia". *Quantitative science studies*, v. 1, n. 4, pp. 1349-1380.
https://doi.org/10.1162/qss_a_00080

Ding, Yudong; **Zhang, Chunfeng**; **Liu, Ying** (2013). "Statistics and analysis of wiki-based Chinese online-encyclopedia cited by journal papers". *Journal of intelligence*, v. 32, n. 3.
http://en.cnki.com.cn/Article_en/CJFDTOTAL-QBZZ201303019.htm

Encyclopædia Britannica (2019). *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Encyclop%C3%A6dia_Britannica&oldid=931322238

Fallis, Don (2008). "Toward an epistemology of Wikipedia". *Journal of the American Society for Information Science and Technology*, v. 59, n. 10, pp. 1662-1674.
https://doi.org/10.1002/asi.20870

Giles, Jim (2005). "Internet encyclopaedias go head to head". *Nature*, v. 438, pp. 900-901.
https://doi.org/10.1038/438900a

Holman-Rector, Lucy (2008). "Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles". *Reference services review*, v. 36, n. 1, pp. 7-22.
https://doi.org/10.1108/00907320810851998

Huggett, Sarah (2012). "The influence of free encyclopedias on science". *Research trends*, n. 27.
https://www.researchtrends.com/issue-27-march-2012/the-influence-of-free-encyclopedias-on-science

Izhikevich, Eugene (2006). "Scholarpedia". *Scholarpedia*, v. 1, n. 2, pp. 1.
https://doi.org/10.4249/scholarpedia.1

Jemielniak, Dariusz; **Masukume, Gwinyai**; **Wilamowski, Maciej** (2019). "The most influential medical journals according to Wikipedia: Quantitative analysis". *Journal of medical internet research*, v. 21, n. 1, e11429.
https://doi.org/10.2196/11429

Jullien, Nicolas (2012). "What we know about Wikipedia: A review of the literature analyzing the project (s)". *SSRN*, 86 pp.
https://doi.org/10.2139/ssrn.2053597

Kousha, Kayvan; **Thelwall, Mike** (2017). "Are Wikipedia citations important evidence of the impact of scholarly articles and books?". *Journal of the Association for Information Science and Technology*, v. 68, n. 3. pp. 762-779.
https://doi.org/10.1002/asi.23694

Kousha, Kayvan; **Thelwall, Mike**; **Abdoli, Mahshid** (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*, v. 63, n. 9, pp. 1710-1727.
https://doi.org/10.1002/asi.22717

Li, Xuemei; **Thelwall, Mike**; **Kousha, Kayvan** (2015). "The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication". *Aslib journal of information management*, v. 67, n. 6, pp. 614-635.
https://doi.org/10.1108/AJIM-03-2015-0049

Lih, Andrew (2004). "Wikipedia as participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a news resource". In: *5th International symposium on online journalism.*
http://www.ufrgs.br/limc/participativo/pdf/wikipedia.pdf

Lin, Jennifer; **Fenner, Martin** (2014). "An analysis of Wikipedia references across PLOS publications". In: *Altmetrics14: Expanding impacts and metrics an ACM web science conference 2014 workshop*, pp. 23-26.
https://figshare.com/articles/journal_contribution/An_analysis_of_Wikipedia_references_across_PLOS_publications/1048991

Mesgari, Mastafa; **Okoli, Chitu**; **Mehdi, Mohamad**; **Nielsen, Finn-Arup**; **Lanamäki, Arto** (2015). "'The sum of all human knowledge': A systematic review of scholarly research on the content of Wikipedia". *Journal of the Association for Information Science and Technology*, v. 66, n. 2, pp. 219-245.
https://doi.org/10.1002/asi.23172

Messner, Marcus; **DiStaso, Marcia W.** (2013). "Wikipedia versus Encyclopedia Britannica: A longitudinal analysis to identify the impact of social media on the standards of knowledge". *Mass communication and society*, v. 16, n. 4, pp. 465-486.
https://doi.org/10.1080/15205436.2012.732649

Okoli, Chitu; **Mehdi, Mohamad**; **Mesgari, Mastafa**; **Nielsen, Finn-Arup**; **Lanamäki, Arto** (2012). "The people's encyclopedia under the gaze of the sages: A systematic review of scholarly research on Wikipedia". *SSRN*.
https://doi.org/10.2139/ssrn.2021326

Okoli, Chitu; **Mehdi, Mohamad**; **Mesgari, Mastafa**; **Nielsen, Finn-Arup**; **Lanamäki, Arto** (2014). "Wikipedia in the eyes of its beholders: A systematic review of scholarly research on Wikipedia readers and readership". *Journal of the Association for Information Science and Technology*, v. 65, n. 12, pp. 2381-2403.
https://doi.org/10.1002/asi.23162

Park, Taemin-Kim (2011). "The visibility of Wikipedia in scholarly publications". *First Monday*, v. 16, n. 8.
https://doi.org/10.5210/fm.v16i8.3492

Priem, Jason; **Piwowar, Heather A.**; **Hemminger, Bradley M.** (2012). "Altmetrics in the wild: Using social media to explore scholarly impact". *ArXiv.*
http://arxiv.org/abs/1203.4745

Samoilenko, Anna; **Yasseri, Taha** (2014). "The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics". *EPJ data science*, v. 3, n. 1.
https://doi.org/10.1140/epjds20

Scopus (2021). *How do I find Open Access journals and articles in Scopus?*. Scopus: Access and use Support Center.
https://service.elsevier.com/app/answers/detail/a_id/11268/supporthub/scopus/kw/open+access

Shoyama, Rex (2014). "Citations to Wikipedia in Canadian law journal and law review articles". *Canadian law library review*, v. 39, n. 12, pp. 11-15.
https://ssrn.com/abstract=2578678

Tohidinasab, Fariba; **Jamali, Hamid R** (2013). "Why and where Wikipedia is cited in journal articles?". *Journal of scientometric research*, v. 2, n. 3, pp. 231-238.
https://doi.org/10.4103/2320-0057.135415

Tomaszewski, Robert (2018). "A comparative study of citations to chemical encyclopedias in scholarly articles: *Kirk-Othmer Encyclopedia of Chemical Technology* and *Ullmann's Encyclopedia of Industrial Chemistry*". *Scientometrics*, v. 117, n. 1, pp. 175-189.
https://doi.org/10.1007/s11192-018-2844-1

Tomaszewski, Robert; **MacDonald, Karen I.** (2018). "A study of citations to Wikipedia in scholarly publications". *Science & technology libraries*, n. 35, n. 3, pp. 246-261.
https://doi.org/10.1080/0194262X.2016.1206052

Wang, Zhihong (2016). "Research on characteristics of online encyclopedia cited by LIS journal articles in China". *Library and information service*, v. 60, n. 19, pp. 99-107.
http://en.cnki.com.cn/Article_en/CJFDTotal-TSQB201619016.htm

Appendixes

Appendix I

Scopus syntax for searching citing documents to the four encyclopedias

Query for documents citing *Wikipedia* (2002 to 2020):

(REF("*wikipedia.org/w*") OR REFSRCTITLE("*wikipedia*")) AND ((PUBYEAR > 2001) AND (PUBYEAR < 2021))

Query for documents citing *Britannica* (2002 to 2020):

(REF("*britannica.com*") OR REFSRCTITLE(("*Encyclopædia Britannica*" OR "*Encyclopaedia Britannica*" OR "*Encyclopedia Britannica*" OR "*Britannica Online*"))) AND ((PUBYEAR > 2001) AND (PUBYEAR < 2021))

Query for documents citing *Baidu Baike* (2007 to 2020):

(REF("*baike.baidu.com*") OR REFSRCTITLE("baidu baike" OR "百度百科")) AND ((PUBYEAR > 2006) AND (PUBYEAR < 2021))

Query for documents citing *Scholarpedia* (2007 to 2020):

(REF("*scholarpedia.org/article*") OR REFSRCTITLE(scholarpedia) AND NOT REFSRCTITLE("scholarpedia of touch")) AND ((PUBYEAR > 2006) AND (PUBYEAR < 2021))

Appendix II

Scopus syntax for searching Green OA documents that are not Gold, Hybrid Gold or Bronze

OA(repository) AND NOT (OA(publisherfullgold) OR OA(publisherhybridgold) OR OA(publisherfree2read)