



# OCLC library holdings: assessing availability of academic books in libraries in print and electronic compared to citations and altmetrics

Ashraf Maleki<sup>1</sup>

Received: 19 April 2021 / Accepted: 23 November 2021  
© The Author(s) 2021

## Abstract

Although library holding data is constructed upon work format of books, it is less known how much print and electronic books in libraries contribute to the library holding counts. In response, this research is an attempt to explore the distribution of library holding data across work formats and investigate the availability of library print holdings and library electronic holdings for books as constituents of the library holding metric across fields and over time and compared with other book metrics. ISBNs, titles and author names of 119,794 Scopus-indexed book titles across 26 fields were examined for fourteen variables including OCLC Library Holdings, Scopus Citations, Google Books Citations, Goodreads engagements, and Altmetric indicators. There are three major findings: (a) library holdings are a more comprehensively available metric for books (over 97%) than any other metric and could be useful after short time after first edition publication, followed by Google Books, Goodreads and Scopus, respectively; (b) on average electronic holdings are seven times (median three times) more numerous than print holdings and their ratio is growing considerably for more recent books; (c) there is consistent downward trend in average print book holdings, suggesting that library print holding data are cumulative in nature and statistically comparable to formal citations; however, acquisition of electronic books in libraries is inconsistent in distribution plot as well as over time. In sum, the differences between print and electronic holding data are broad making them distinct metrics, suggesting that further research is needed for understanding their implications for book impact assessment.

**Keywords** Scholarly books · Book impact · Library holding · Library print holding · Library eholding · Library electronic holding · Citation distribution · Altmetric · Google Books · Syllabus mentions · OCLC · WorldCat · Twitter · Mendeley readership · Facebook · Wikipedia · Blogmentions · News stories

---

✉ Ashraf Maleki  
ashraf.maleki@utu.fi

<sup>1</sup> University of Turku, Turku, Finland

## Introduction

Library Holding (LH) which is also called *Libcitation* (White et al., 2009) or *Catalog Inclusion* (Torres-Salinas & Moed, 2009) is introduced as an indicator of books' acquisition worth from librarians' point of view, as they presumably choose titles based on institutional and community requests. Although library holdings have been for over a decade identified as a metric worth exploring for book impact assessment, there is limited information about the type of impact information they can offer. Technically Library Holdings (LH) as a metric refers to the number of libraries that have acquired a certain book title in their collection. Previous studies have acquired LH through Library Catalogue Analysis for a selection of different libraries (Torres-Salinas & Moed, 2009; White et al., 2009). However, LH count as statistics representing considerable number of world libraries providing a book have been available from OCLC's WorldCat and researched later on (see for example Zuccala & Guns, 2013) after the introduction of the metric itself (Torres-Salinas & Moed, 2009; White et al., 2009). LH is shown as *Total Holdings* in the web pages of books in Online Computer Library Centre (OCLC.org). This count mainly represents holding statistics of U.S. libraries (about 63%) and slightly more prominently academic libraries (5804 libraries, 38%) rather than public libraries (4441 libraries, 29%) or other libraries (4950 libraries, 33%) (Torres-Salinas & Arroyo-Machado, 2020). Nevertheless, an important underlying factor for calculation of library holding is work format of books, which is a central service for Library Management and is offered in a broad range of formats in OCLC.

The feature of work format of books, refers to the physical or virtual accessibility of books. There is a breakdown in OCLC for 'Total Holding' statistics to '*holdings*', which refers to extent of libraries with print holdings, and '*eholdings*', which refers to the number of libraries with electronic holdings of books. Therefore, the original contribution of this study is to utilize this technical possibility and distinguish library holdings across print and electronic collections and provide a comparison of availability of books in libraries with a broad range of other book indicators. This research, thus, is centered on statistics of book format in libraries, since it is important to understand how this can influence our understanding and application of library holding counts in book impact assessment. However, one major challenge with library holding is that despite studies on general LH statistics, it is not clear to what extent work format of books acquired by libraries in print and electronic can be relevant in bibliometric and scientometric assessments. There has been a growing interest in massive collection of ebooks in libraries in the past decade, as libraries' budget has increased for them (Romano et al., 2015; Wells & Sallenbach, 2015; Romano, 2016), whilst the impact of this budget rise has not been very much investigated from research assessment perspective. Therefore, current research is an attempt to investigate how libraries' tendency to electronic over print influence the statistical significance of LH compared to other book indicators.

Although library holdings are introduced for more than a decade now (White et al., 2009; Torres-Salinas & Moed, 2009), the distributional characteristics of data, the variations between print and electronic holdings counts, the benefits of library holding data over other book indicators in terms of quantity and the rate at which library holding data can be accrued are less known and will be explored in this research. This research also investigates the highly acquired books in libraries in print and electronic in order to help identify their distinguishing characteristics in relation to each other and other research indicators.

## Research questions

The aim of this study is to investigate the extent of book availability in WorldCat libraries across their publication formats and compared to other book metrics. In order to address the research issue, this study investigated distribution pattern of print and electronic book holdings in libraries across disciplines and over time. Thus, the following questions are drivers of current research:

1. Are there differences in terms of library holdings based on the format of books (print vs. electronic)?
2. How do library holdings compare with other metrics when considering book format?
3. How do the differences in above questions evolve over time?
4. What are the general characteristics of books with the highest library print and electronic holdings?

## Background

### Coverage of books in libraries vs. other book metrics

Various studies have reported broader coverage of books in WorldCat for English language books and not as much attention to non-English books such as books in Spanish (Torres-Salinas et al., 2020). Table 1 gives a general comparison of metrics for their uptake of books in seven different studies. Since proportions were numerous and across a variety of disciplines, for a broader view of the uptake of metrics the total coverage in the studies is reported. As with the Library Holdings, the proportions reported are very close to each other, all above 97% (Kousha et al., 2011; Kousha et al., 2017; Halevi et al., 2016; Torres-Salinas et al., 2017). WorldCat Holdings were available for more titles than EBSCO abstract views (91.5%) and saves (78.2%), Goodreads pages (min=53.7%), and EBSCO PDF views (min=65.64%).

Regarding formal citations, whenever the initial data were from Scopus or Web of Science the coverages were substantially high (for instance, 72% BKCI coverage in Kousha and Thelwall, 2011 and 55% Scopus coverage in Kousha & Thelwall, 2016), but the Scopus citation coverages reported for data from other databases such as EBSCO (Torres-Salinas Robinson-García & Gorraiz, 2017) and Levy Ebrary (Halevi et al., 2016) were minor (both about 4%). Other metrics were much less available such as Mendeley readers (max=43.1%), Goodreads reviews (max=25%) and Wikipedia citations (max=34.6%) (more in Table 1. These percentages suggest the significance of library holdings as a comprehensively available metric for books.

### Library book collection: print vs. electronic

Financial reasons are central to the substantial increase of ebooks and reduction in print books in academic libraries. This can be seen in a single library, such as Curtin University Library in Australia that increased ebook budget from 52% of all monographs in 2010 to 90% in 2014 (Wells & Sallenbach, 2015), as well as at a national level like in the U.S. where ebooks represented 7.4% of total academic library budget in 2010 to 9.3% in 2016 (Romano, 2016). A similar trend is apparent to U.S. public libraries, with increased budget

**Table 1** The overall coverage of books in different metrics based on previous studies

Metrics	Kousha et al. (2011)	Kousha and Thel-wall (2015)	Kousha and Thel-wall (2016)	Kousha et al. (2017)	Halevi et al. (2016)	Torres-Salinas et al. (2017)	Erfanmanesh et al. (2019)
WorldCat Holdings	100%			99.3%	98.8%	97.8%	
Scopus Citations		55%	28.6%		4.04%	4.25%	
BKCI* Citations	72%	77%		73.4%			
Google Books Citations	84%		62.3%	73.8%			
Syllabus mentions		38%	25.9%				
Amazon reviews	29%						
Mendeley captures	7%				43.1%	24.9%	
Goodreads uptakes				64.6%	69.3%	53.7%	
Goodreads reviews				14.1%	25.1%	19.1%	
Wikipedia Citations			33.2%	34.6%	25.5%	16.6%	
EBSCO Saves					78.2%	81.8%	
EBSCO Abstract views					91.5%	95.1%	
EBSCO PDF views					65.6%	68.3%	
Altmetric coverage							20.8%
Springer Bookmarks							90.25%

\*BKCI: Book Citation Index

Kousha, Thelwall, and Rezaei (2011) 2,739 monographs indexed in BKCI from 2008

Kousha and Thelwall (2015) about 12,000 Scopus- and 2300 BKCI-indexed monographs in 2005–2010

Kousha and Thelwall (2016) 18,735 Scopus monographs in 2005–2012, multidisciplinary

Kousha, Thelwall and Abdoli (2017) 15,928 BKCI-indexed books, multidisciplinary and over 2008–2010

Halevi, Nicolas and Bar-Ilan (2016) 71,443 Levy Ebrary e-books in PlumX metrics, multidisciplinary

Torres-Salinas, Robinson-García, and Gorraiz (2017) 263,210 EBSCO e-books

Erfanmanesh et al. (2019) 1,116 e-books Springer Bookmatrix, multidisciplinary

share from 1.7% in 2010 to 6.3% in 2015 (Romano, Grimscheid & Genco, 2015; see also Gu & Berger, 2018). A 2016 survey of 346 U.S. academic libraries showed that money spent on ebooks in about half of the libraries is likely to detriment print book budget (61%) more than any other area and this detraction more frequently occurs in libraries with a medium-ranged budget (52%) and graduate libraries (52%) where students inevitably use ebooks more than prints (Romano, 2016).

Libraries also turned to ebooks in order to save library physical space (Kont, 2016), meet studying habits of millennial generation (Maceviciute et al., 2017; Casselden & Pears, 2020), and keep up with technology whenever software features are the reasons for faculty members to request the ebook versions (Romano, 2016). This attraction to ebooks originates in technical features such as search capability in full text of books (Parkes, 2007) and dynamic cross-linking to other online content (Bunkell & Dyas-Correia, 2009), and also changed studying habits such as bite-sized reading (Zhang et al., 2017; Casselden & Pears, 2020), and research and leisure reading (Majid et al., 2019). A survey of American academic libraries showed that top influential factors for purchasing *specific ebook titles* are 'faculty request' (82%), 'price' (79%), and 'curriculum assignment' (76%), but much less because of 'students request' and 'reviews' (both 38%), although the effect of these factors can be undermined when prepared publisher packages include additional unrequested titles for the libraries.

Another important issue with ebooks is that they are not a perpetual part of libraries and always accessible. 'Subscription' access and short-term loans are prevalent in majority of the U.S. academic libraries (79% in 2016), since they sound more economical in exchange for accessibility of a larger number of titles (Romano, 2016). Furthermore, although majority of academic libraries are willing to provide interlibrary loans of ebooks, the possibility of full ebook lending is frequently bound by license issues and technical problems such as single-user access (34%) rather than simultaneous access (55%), and thus, libraries move to "chapter lending" as a safer alternative (Zhu, 2018), suggesting complexity of libraries' ebook access plans.

### Library-related book usage: print vs. electronic

Monitoring book usages is a major method for libraries to assess and modify library print and ebook collections (Cox, 2011). However, book usage statistics from libraries are not as broadly available as library holdings. Print book usages are assessed through library circulation system, while ebook usages are followed by most American librarians (87%) via protocols such as COUNTER (Counting Online Usage of Networked Electronic Resources) (Cox, 2011) or SUSHI (Standardized Usage Statistics Harvesting Initiative) (Conyers et al., 2017). In U.S. academic libraries, 'number of full-text retrievals' (53%), 'number of downloads' (48%), 'number of sessions' (38%) and 'number of page views' (35%) are reported for ebooks (Romano, 2016), suggesting less than half of ebooks in libraries are actually useful. A study on format preference on e-Duke Scholarly Collection between 2011 and 2013 showed that 73% of ebooks received user interest (Goodwin, 2014) with 12% significantly more in use and slightly more preferred in print format. A survey in U.S. suggested that in only 16% of academic libraries, students prefer only print books, but the general 'preference for print books' have risen over time from 40% in 2010 to 60% in 2016 (Romano, 2016). Despite this, another study of 10,006 titles' print circulation and ebook sessions in library of Graduate School of Education in New York City showed that a substantial proportion of ebooks (82.5%) were viewed more than once, although both print and

electronic versions showed usage over time, suggesting the necessity of their coexistence (Haugh, 2016).

Reports indicate that in general, use of ebooks is more prevalent among social science and humanities and medicine students than other fields. In some of the Library of Congress Classes including A (General Works), F (History of the Americas), H (Social Sciences) and Q (Math and Science) ebooks are used more than in other fields (Kohn, 2018). A study on 7880 titles in the Duke University Library showed that the percentage of ebooks accessed in a subject were either equal or larger than the percentage used in print. The percentages of titles used in ebook in Computers (66%), Psychology (57%), Medicine (51%), and Education (38%) held the broadest, yet only small differences from ones used in print (53, 49, 42 and 31%, respectively) (Littman & Connaway, 2004). Correspondingly, American adults showed a greater tendency to read ebooks in “Business” (48% of U.S. academic libraries), followed by Nursing (44%), Psychology (43%), History (37%) and Education (36%) (Romano, 2016). A 2009 survey at the University of Portsmouth showed 1111 students’ use of ebooks prevailed in Humanities and Social Science (about 70% of respondents) and international students (over 50%), in contrast to Information Technology (just over 40%) (Worden and Collinson, 2011, p. 238, see also 39% of IT students in a Malaysian University in Ismail & Zainab, 2007). ‘Library’ played a key role in initial discovery of ebooks for Humanities and Social Science and international students (about 60%), whereas lecturers’ recommendations (36%) was more important for Technology students (Worden & Collinson, 2011), suggesting that educational use of ebooks also prevails in humanities and social sciences.

Studies across various academic grades have also shown varying preferences for print books and ebooks. One survey of students in American universities suggested that there is a definite preference for e-textbooks at graduate level, as undergraduate students also used less print (42%) and more ebooks (58%) (Romano, 2016). Another survey of 1571 respondents at University of Liverpool showed that students (90%) compared to researchers and lecturers (77%) are more aware of ebooks and use them more (Springer, 2010). However, the extent of using ebooks for research purposes (70 and 85%, respectively) and educational purposes (80% and 48%) significantly vary for both students and researchers. Another survey of 979 users of Health Science Library System at the University of Pittsburgh suggested that over 55% of users used ebooks, while just above one-fifth of faculty members would assign ebooks for class readings (Folb et al., 2011). Overall, these findings suggest that usages of ebooks by students for education are often voluntary, while graduate students and faculty members in general are more inclined to use ebooks.

## Research method

### Dataset

The research design was to compare library holdings of books across two formats of print and electronic. The main aim was to investigate the extent of books with library print and electronic holdings, in a comparable manner with other metrics, over time, and across fields. Data were drawn from seven data sources: (1) Scopus, (2) OCLC (classify.oclc.org); (3) WorldCat (worldcat.org); (4) Altmetric.com; (5) Open Syllabus Project (opensyllabus.org); (6) Google Books (books.google.com); and (7) Goodreads (goodreads.com):

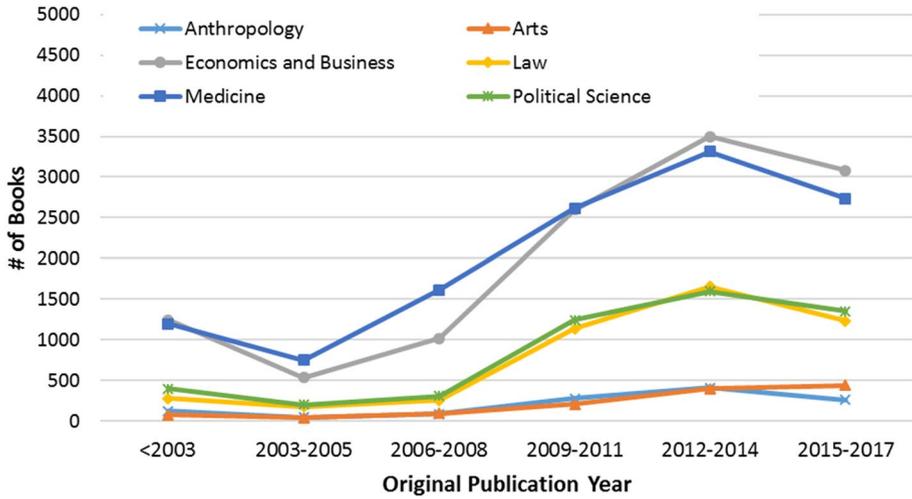


Fig. 1 Number of unique book titles in each time period across six fields

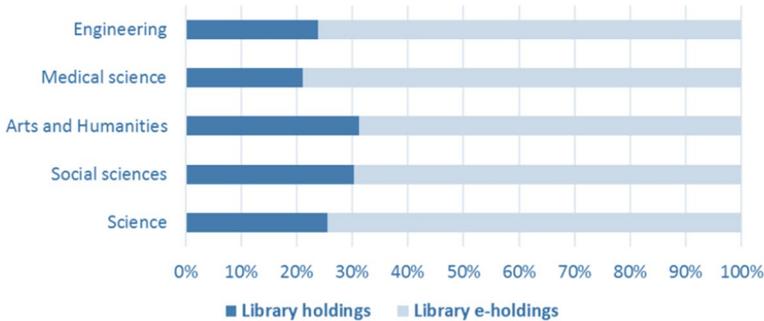


Fig. 2 Proportion of libraries print and electronic holdings across five broad disciplines

- Dataset 1 consisted of all available Scopus-indexed book titles (168,866) in 2018, which was, then, narrowed down to 119,794 unique titles with ISBN or E-ISBN and then to 110,603 titles with a Library of Congress Subject Class (LCC) in OCLC (Tables 3 and 4).
- Dataset 2 involved all 36,493 titles within six broad subject categories of Anthropology, Arts, Business and Economics, Law, Medicine and Political Science as in Dataset 1. This sample is selected for further analyses by the first edition’s publication date of books in WorldCat. Books are important in these selected fields, but are relatively less important in Medicine than fields selected from Social Science and Humanities. However, Medicine makes an interesting case, because uptake of book publishing technologies in this field both for print and electronic is relatively swift and makes a good case of speculation.

**Table 2** Work format of scopus book titles as assigned in OCLC API results

OCLC work format	No. of titles	Sum of holdings (%)	Sum of E-holdings (%)
Book	34,139	9,513,637 (70)	4,138,246 (30)
eBook	101,960	16,629,835 (21)	64,117,071 (79)
Thesis/dissertation	285	91,780 (85)	16,701 (15)
Microform	6	1041 (78)	290 (22)
Computer file	13	514 (92)	46 (8)
Continually updated resource	1	214 (79)	56 (21)
Audiobook on LP	1	149 (100)	0
Electronic article	1	99 (70)	42 (30)
ePeriodical	1	86 (85)	15 (15)
eAudiobook	4	78 (100)	0
Image	1	4 (100)	0
Website	1	1 (100)	0
Total	136,414	26,237,438	68,272,467

*Library print and electronic holdings* OCLC offers Total Holdings, as sum of holdings and e-holdings, in its summary results, but OCLC extended results that were obtained via API (in January 2020) and simultaneously included two counts of ‘holding’ and ‘e-holding’ for each book, suggesting summing of all print and electronic versions, respectively, stored in WorldCat libraries. OCLC’s extended holding statistics were gathered using a custom software to harness book data via ISBN as the handle of primary searches, but a second automatized search was designed for unfound results based on title and first author’s surname, as below:

Primary search:

<http://classify.oclc.org/classify2/Classify?isbn=9781843345213&summary=true>.

Secondary search:

[http://classify.oclc.org/classify2/Classify?author=Adair&title="TheHistoryofAmericanIndians"&summary=true](http://classify.oclc.org/classify2/Classify?author=Adair&title=).

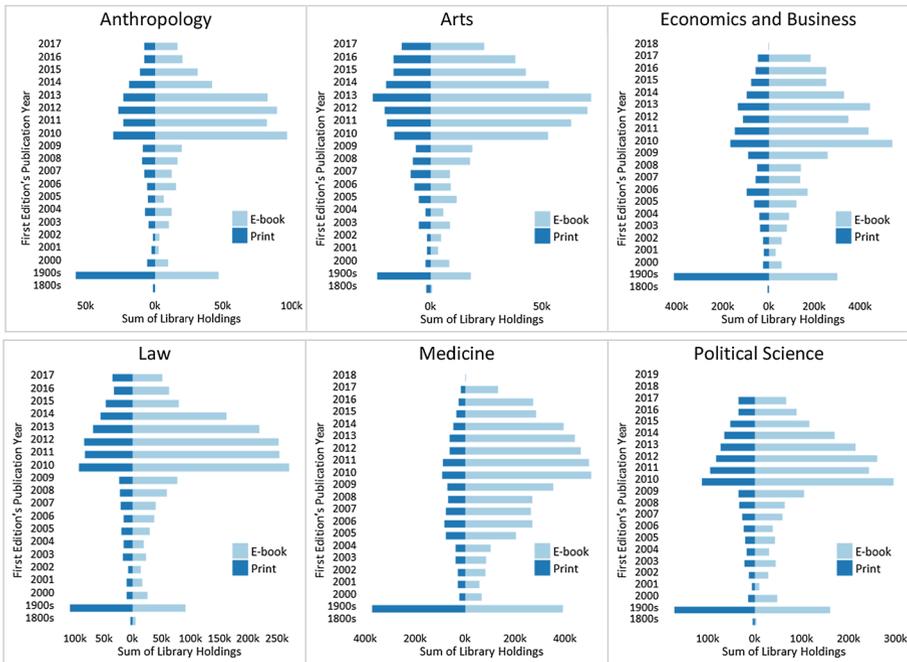
*Work formats and editions* Whenever several editions<sup>1</sup> of a title exist, it would automatically be identified under one ID in OCLC that had amalgamated holdings of all editions in all work formats and classifications. Multi-edition books create duplication problem and this is identified in other studies on scholarly books (e.g., Zuccala & White, 2015; and Zuccala et al., 2018) and tackled for instance through selecting single edition monographs in the database (Kousha & Thelwall, 2015). About 3.5% of book titles in Scopus (3,862) included a mention to edition number and a mere 0.25% of titles (260) represented repeated works. In this research, only one version of the Scopus titles was kept to prevent duplication. In the OCLC extended results, it was also seen that one of twelve default work formats were attributed to each book, although OCLC supports assigning documents to about 50 material format types.<sup>2</sup> As in Table 2, the most assigned work format was eBook

<sup>1</sup> In OCLC, collation of count of book editions and various cataloguing of the same work in different subjects and publication format types morphs into “all editions” of a title (Torres-Salinas and Arroyo-Machado, in press), and therefore OCLC ‘editions’ does not translate into revised publications.

<sup>2</sup> [https://help.oclc.org/Library\\_Management/WorldShare\\_Collection\\_Evaluation/My\\_Library/Filter](https://help.oclc.org/Library_Management/WorldShare_Collection_Evaluation/My_Library/Filter).

(101,960 titles, 74.7%) that is apparently assigned when the proportion of library holdings in the electronic format (79%) actually prevailed print holdings (21%), while ‘Book’ as a work format was assigned to one-third of titles (34,139) that were mostly available in print (70%) rather than electronic (30%), presumably suggesting that books are attributed to the work format that libraries provide the most. Perhaps strangely, some obviously electronic formats, such as eAudiobook and Website, only had holding counts rather than e-holdings. A very negligible proportion (314, 0.2%) was identified across other formats (more in Table 2).

*Publication year of first edition* Publication year of books are important to normalize the effect that time has on citation accumulation and to compare the results across different fields (Thelwall & Fairclough, 2015). Nevertheless, many book titles are duplicated because of numerous republications or revisions in different years and by various publishers. For instance, a book dating originally back to 1775 is republished in 2018 by a different publisher with the same content and title of “*History of the American Indians*” by James Adair; it is shown in Scopus by one of its republication dates of 2005. Although Scopus publication date of books are used to indicate the distribution of print and electronic holdings, in order to deal with the problem of publication year, the ‘Publication Year of First Edition’ in WorldCat is used. This date is extracted for Dataset 2 and results are checked and modified because sometimes the oldest publication year of a book title would be one year before the actual publication date and in some cases the first edition year in OCLC showed a wrong date such as 1900 or 1800 which could both be identified by oddly low frequency of libraries assigning books to that date. Furthermore, publication years are



**Fig. 3** Population of print and ebook holdings of libraries over first edition's publication year across six fields (as of January 2020)

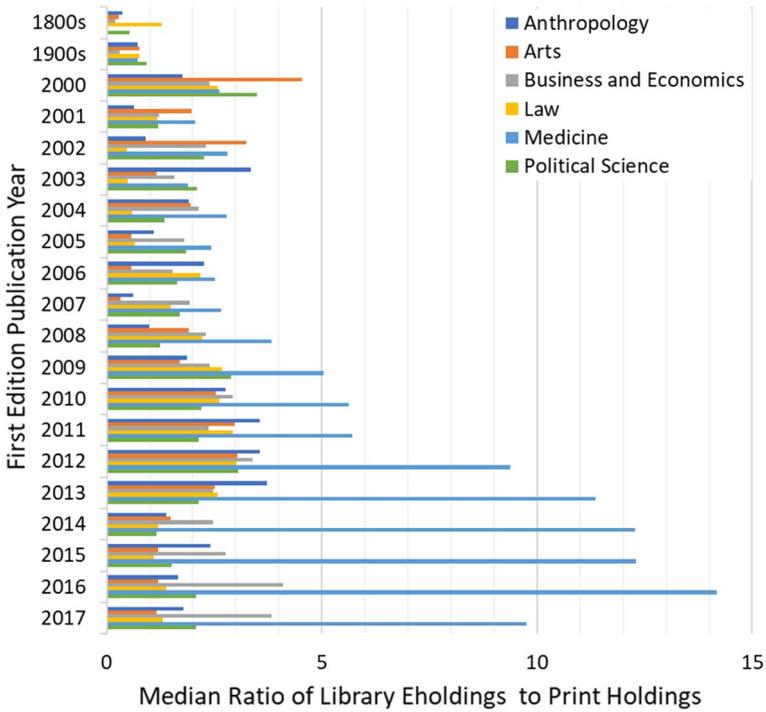


Fig. 4 Median Ratio of LEH to LPH counts across six fields over years

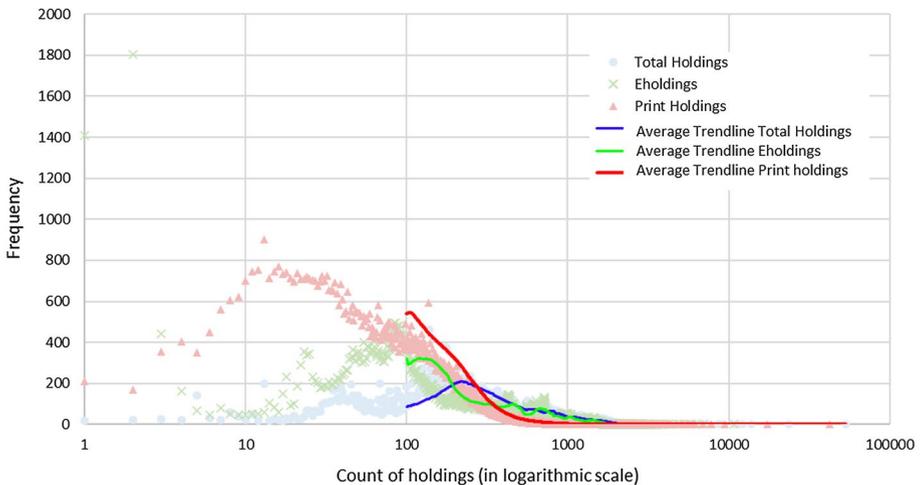
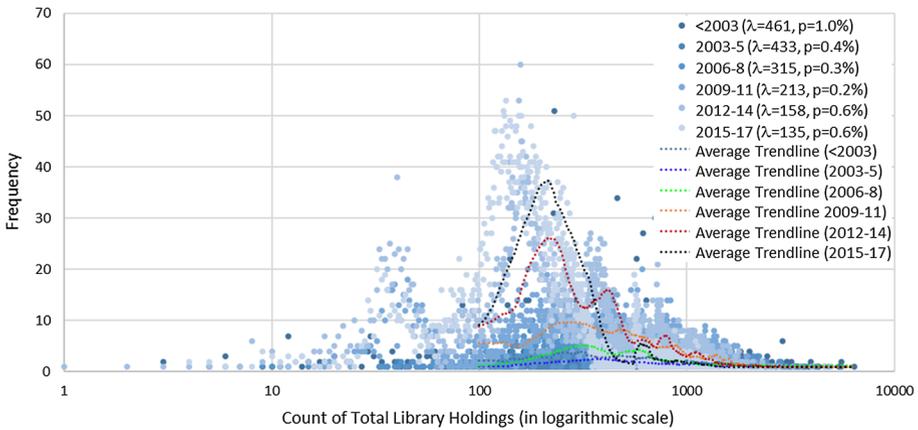
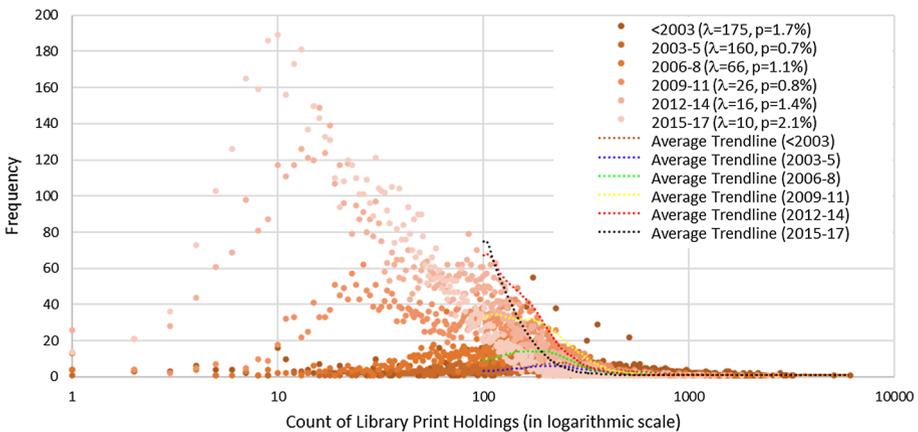


Fig. 5 Trendline of distribution of LPH, LEH, and TLH in Dataset 1 (n = 110,603)

grouped in six consecutive time periods (Fig. 1) in a way that each time period includes more than 30 titles, so that they can produce statistically significant results when analyzing



**Fig. 6** Trendline of distribution of TLH for books in Dataset 2 over time (n of <2003=3,307; n of 2003–5=1,748, n of 2006–8=3,366; n of 2009–11=8,074; n of 2012–14=10,872, n of 2015–17=9,091)

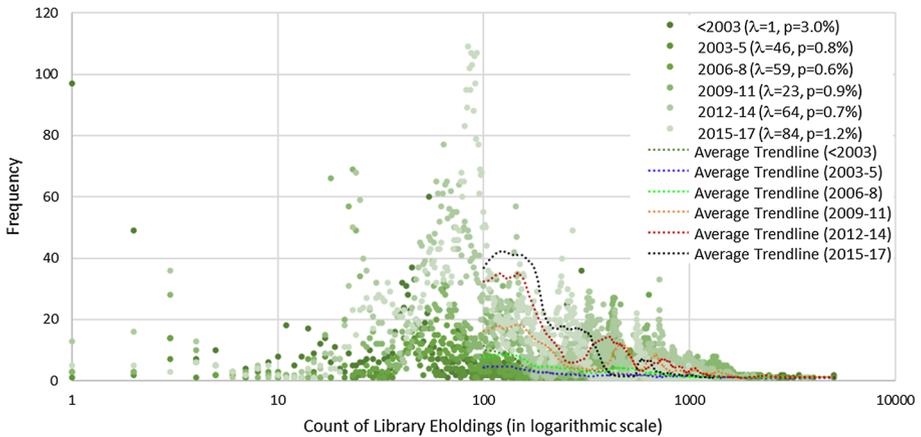


**Fig. 7** Trendline of distribution of LPH for books in Dataset 2 over time

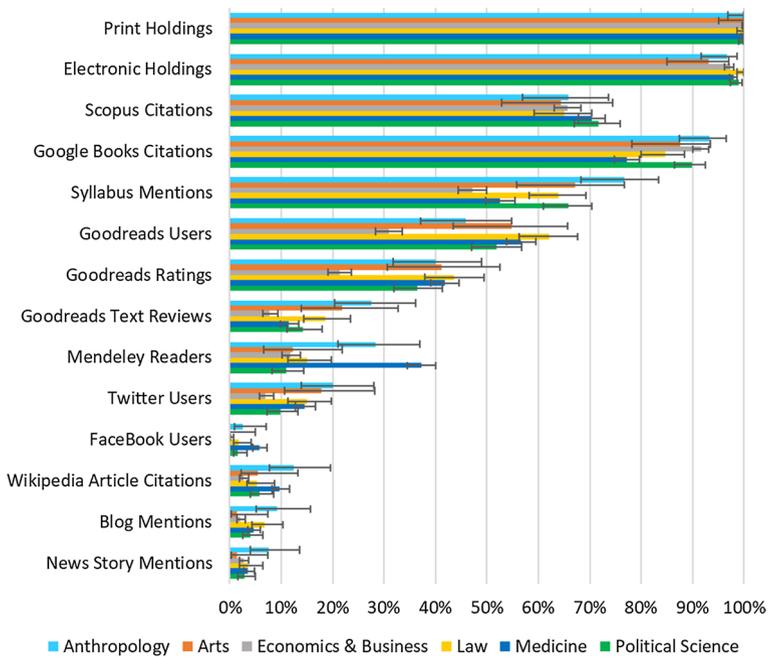
the confidence intervals of error bars in the trend analysis of normalized proportion cited across different indicators.

Figure 1 shows that the number of books in each time period are substantially different across fields. However, some pairs of fields are showing parallel changes and similar frequencies over time periods: there are changes on a range of 40 (2003–2005) to 437 (2015–2017) for Arts and Anthropology, on a range of 179 and 1656 for Law and Political Science, and on a range of 532 (2003–2005) and 3499 (2012–2014) for Medicine and Economics and Business.

*Google Books Citations (GB)* ‘Book-to-Book citations’ are extracted from Google Books via API using the method proposed by Kousha & Thelwall (2015), running searches with Webometric Analyst (lexiurl.wlv.ac.uk) during December 2019. The least word count of titles was set to four, therefore, the possibility of false positive retrievals in search results could be reduced. The maximum word count in titles was set to six and up to three authors’



**Fig. 8** Trendline of distribution of LEH for books in Dataset 2 over time

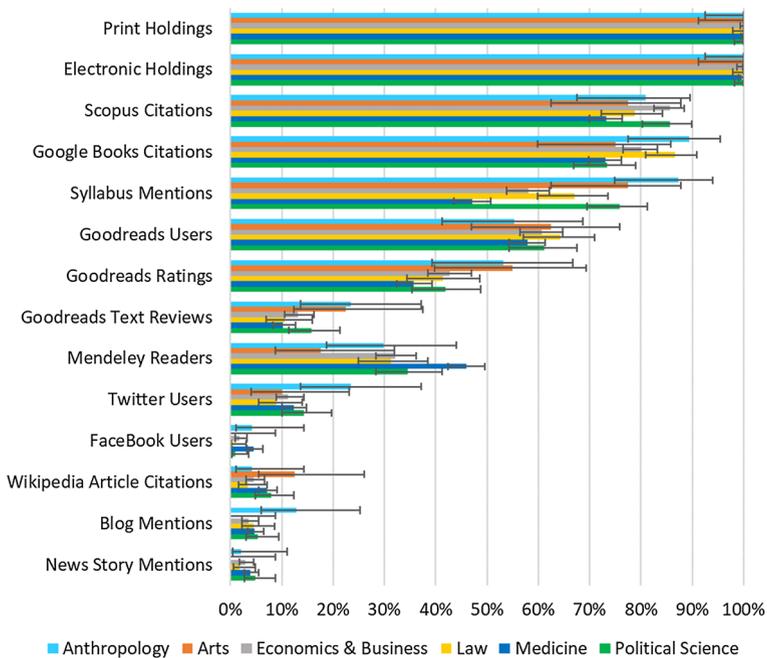


**Fig. 9** Normalized Proportion Cited of books published <2003 across all metrics and six fields

last names were included in the searches. Query below exemplifies the search conducted for the book entitled “*Advanced test methods for SRAMs: Effective solutions for dynamic fault detection in nanoscaled technologies*” and authored by “Bosio A., Dilillo L., Girard P., Pravossoudovitch S., and Virazel A.”.

Example Google Books search query:

Bosio Dilillo Girard "Advanced test methods for SRAMs Effective"



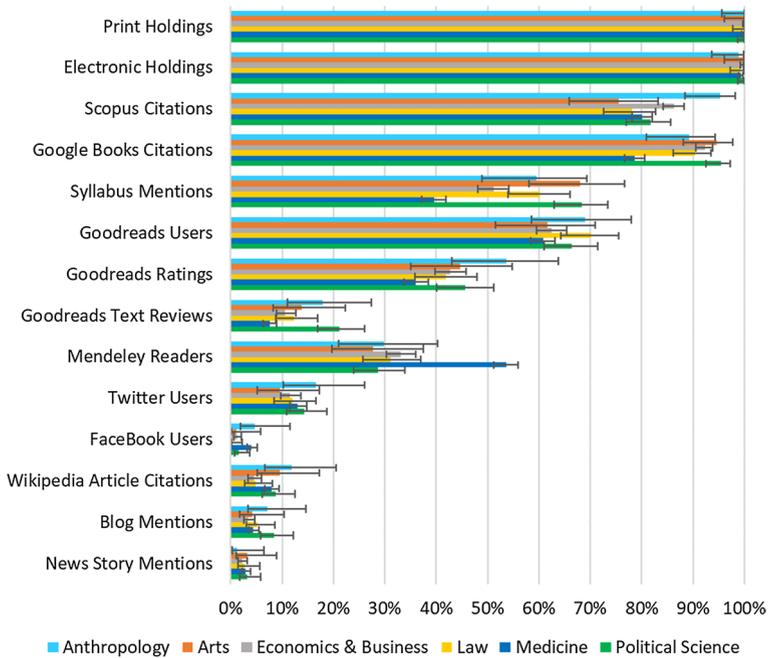
**Fig. 10** Normalized Proportion Cited of books published 2003–2005 across all metrics and six fields

Publication year is not included in the search so that all citations given to all editions and prints could be collected.

*Syllabus Mentions (SM)* Counts of university courses referring to book titles are obtained from *opensyllabus.org* (in March and April 2020) through the explorer rather than *altmetric.com*, since they indicated significant variation with more comprehensive results viewed at *opensyllabus.org*. This metric is only obtained for selected six fields (Dataset 2) due to extensive manual checking required in the Open Syllabus explorer. *Opensyllabus.org* is a website collecting and offering open free university curricula worldwide and is currently expanding its database to improve its recommendation on most important co-assigned book titles/journal articles to aid the learners.

In current research, all 36,493 titles were searched in “TITLES” search feature and whenever there were several results matching a title the “AUTHORS” search feature was also used. In rare cases, several titles were retrieved, all indicating the same book; in those cases, sum of the assigning course descriptions was used, because their details showed different universities offering the same book for their courses. For instance, for the book “*Debates in art and design education*” authored by Addison N. and Burgess L. two results were retrieved in *opensyllabus.org*, one indicating 84 and the other 6 curriculum appearances. Although the separation did not suggest different books, the listed authors for one book entry are partial and the listed curricula within each entry were linked to distinct universities. Majority of other multi-edition books, however, are already collected under one entry, suggesting that the issue is only sporadically present.

*Goodreads Engagements* Public engagement and reviews of books are extracted from *Goodreads.com* (October 2018) as four indicators of (a) *Goodreads User (GU)s*, which indicate the number of users who showed interest in a book or interacted with it in Goodreads, (b)



**Fig. 11** Normalized Proportion Cited of books published 2006–2008 across all metrics and six fields

*Goodreads Ratings (GR)*, that show the number of star ratings, (c) *Goodreads Text Reviews (GTR)*, that show the number of Amazon and Goodreads text comments about the content of book, and (d) *Goodreads Average Rating (GAR)*, that is the mean ratings given to books by users on a five star assessment. Books were primarily searched in Goodreads through submitting ISBNs via API, running searches in Webometric Analyst. About half of the titles were not retrieved with ISBN or E-ISBN and therefore another custom software was developed to search via API by submitting title and author of books. As there can be numerous pages for some books in Goodreads, *goodreads.com* itself offers two user engagement statistics (work and page statistics) for each result, of which the results representing the total work engagements were used.

*Altmetric.com* All Altmetric indicators were extracted from *altmetric.com* via API using ISBN in January 2020. However, only five indicators are mentioned in this research as the primary research conducted recently (Maleki, 2020) showed that observations for other metrics were frequently low and thus negligible: (a) *Mendeley readers*, which indicate the number of users that have saved a book’s bibliographic information in the reference management service of Mendeley (Henning & Reichelt, 2008); (b) *Twitter*, the unique counts of users posting a book as found by altmetric.com (Adie & Roe, 2013; Eysenbach, 2011); (c) *Facebook*, the unique count of users posting books in Facebook; (d) *Blog Mentions*, the unique count of blogs including citations to books in blog posts; (e) and *News Posts*, the unique count of News outlets including citations to books.

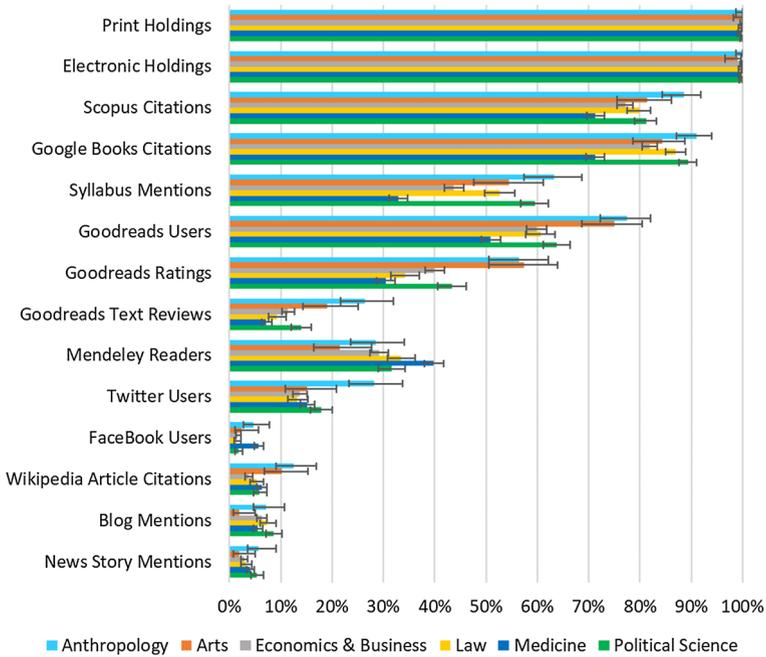


Fig. 12 Normalized Proportion Cited of books published 2009–2011 across all metrics and six fields

**Analysis**

*Normalized proportion cited of books in different metrics* Normalized Proportion Cited (NPC) is used to assess the possible variation in uptake of books in different metric, over years and across fields. Thus, NPC of books from the years with fewer number of books might show larger error bars and vice versa. The experiment was run using the software Webometric Analyst based on the formula proposed in Thelwall (2017) and illustrated via bar charts (Figs. 9, 10, 11, 12, 13 and 14).

*Average of metrics* For the same reason to prevent from mean skewing towards a few remote and highly cited works as in arithmetic mean, the average impact of books in different metric counts was calculated with the Geometric Mean of log-transformed citation (c) scores with  $\ln(c + 1)$  that helps to include books with zero citation by adding one to all records and then contracting it after log-transformation (Thelwall, 2017). Thus, it should be noted that geometric mean is a “density” measure, meaning that it calculates the average citations/mentions for all documents (including zero-counts), in contrast to “intensity” measure that only takes account of average uptakes for non-zero cases.

**Findings**

**Library holdings of print books vs. ebooks across subjects**

In answer to the first research question, Fig. 2 shows the relative sum of print and electronic library holdings compared across broader disciplinary levels in Dataset 1. On

**Table 3** Proportion Cited of books in eight (out of 14) metrics across 26 fields

Fields\Metrics	LPH	LEH	Scopus	GB	SM (OS)	SM (alt)	GU	GR	# of Books
Agriculture	99.9%	99.0%	70.2%	70.3%		1.9%	43.8%	23.5%	1,775
Anthropology*	100.0%	99.6%	77.3%	83.6%	44.9%	9.8%	64.6%	46.0%	1,198
Arts*	100.0%	98.6%	61.8%	71.0%	33.4%	6.8%	64.9%	42.5%	1,245
Business and Economics*	99.8%	99.5%	66.2%	74.6%	30.5%	5.0%	50.7%	31.5%	11,987
Chemistry	99.4%	98.5%	80.1%	67.7%		1.8%	29.1%	10.8%	1,651
Education	99.9%	99.9%	63.3%	77.8%		2.9%	50.5%	32.2%	4,475
Engineering and Technology	99.5%	93.2%	69.5%	62.8%		1.4%	33.4%	16.3%	14,140
Ethics and Religion	99.6%	99.9%	69.4%	84.9%		9.0%	72.3%	54.5%	4,354
Geography	99.6%	98.0%	72.4%	78.1%		5.5%	48.8%	51.7%	1,367
History	99.5%	99.6%	70.7%	83.0%		9.9%	71.4%	53.2%	8,798
Languages and Literature	99.8%	99.8%	71.5%	79.7%		10.9%	68.1%	47.4%	9,798
Law*	99.8%	99.8%	67.0%	79.4%	35.8%	9.8%	52.6%	28.5%	4,734
Library Science and Bibliography	100.0%	100.0%	63.5%	70.7%		6.0%	62.9%	46.8%	563
Mathematics	99.5%	98.8%	73.4%	65.1%		4.0%	47.8%	29.9%	4,093
Medicine*	98.9%	99.5%	62.2%	62.1%	25.6%	2.8%	45.7%	25.8%	12,227
Military Science	99.4%	99.5%	58.2%	69.7%		6.8%	52.6%	34.5%	822
Music	99.8%	100.0%	68.3%	79.3%		10.2%	68.3%	51.7%	1,161
Natural History and Biology	99.8%	98.3%	73.3%	69.1%		4.8%	48.7%	29.1%	1,640
Philosophy	99.9%	100.0%	72.3%	85.0%		8.3%	72.6%	52.5%	1,935
Physics	99.8%	97.7%	72.4%	63.7%		4.5%	44.2%	22.6%	2,628
Physiology	99.4%	99.4%	66.9%	60.5%		2.6%	41.0%	20.2%	1,716
Political Sciences*	99.9%	99.9%	71.8%	82.4%	42.6%	9.1%	56.8%	35.5%	5,098
Psychology	99.8%	99.8%	67.5%	79.4%		5.6%	63.1%	45.2%	1,996
Recreation and Leisure	99.7%	99.7%	69.8%	79.6%		6.7%	67.6%	51.2%	777
Social Sciences	99.8%	99.8%	71.1%	79.8%		8.4%	58.3%	39.4%	9,645
Zoology	99.6%	99.1%	74.0%	67.9%		5.0%	54.4%	37.0%	758
Unknown	99.6%	99.1%	64.2%	69.1%		2.8%	37.7%	23.4%	9,190
<b>Average of 26 Fields (excluding Unknown)</b>	<b>99.70%</b>	<b>99.11%</b>	<b>69.39%</b>	<b>74.12%</b>		<b>6.7%</b>	<b>55.17%</b>	<b>35.98%</b>	<b>110,603</b>
<b>Average of Six* Fields</b>	<b>99.74%</b>	<b>99.47%</b>	<b>67.72%</b>	<b>75.52%</b>	<b>35.46%</b>	<b>7.2%</b>	<b>55.90%</b>	<b>34.97%</b>	<b>36,493</b>

\*: six sample fields in Dataset 2; LPH: Library Print Holdings; LEH: Library Electronic Holdings; Scopus: Scopus Citations; GB: Google Books Citations; SM (OS): Syllabus Mentions gathered from OpenSyllabus.org; SM (alt): Syllabus Mentions gathered from Altmetric.com; GU: Goodreads Users; GR: Goodreads Ratings; Gray cells: Not Checked.

average counts of books consist of 72% library electronic holding (LEH) and approximately 28% library print holdings (LPH). The highest proportions of LPH were seen in Arts and Humanities (31%) and Social Sciences (30%) and the lowest ratio was in Medical Science (21%). In sub-category level, the highest proportion of book titles with at least one LPH belonged to Psychology (36%), followed by Language and Literature and Music (both 32%), while the highest proportions of book titles with at least one LEH were in Medicine (79%), Agriculture, and Physiology (both 78%), suggesting that libraries acquiring ebooks prevail in Sciences and Medicine, as in contrast to Social Sciences and Humanities where they are slightly less dominant.

**Population of print and ebooks in libraries over years**

There were 64,395,780 total holdings recorded for 115,706 book titles. Figure 3 gives a comparison of sum of library holdings across print and electronic over time for six sampled fields (Dataset 2). The patterns of library print and ebook holdings are persistent across fields and over publication date of their first edition. However, the charts for

comparison of print and ebook holdings for 26 fields according to “Scopus publication date” (see Online Appendix<sup>3</sup>) rather than first edition publication date are concentrated on publication dates after 2000 (mostly 2010–2013), suggesting that Scopus publication date refers to *republications* rather than *original publication date* (Fig. 3). For books originally published in the twentieth century, the overall counts of library electronic holdings are almost equal to or slightly smaller than that of print books, suggesting that libraries have almost equally covered for ebook collection of older books. In the first decade of twenty-first century purchase of print and ebooks by libraries has undulated. Nevertheless, in 2010, there have been sharp jumps in the number of libraries purchasing ebooks across fields. These sudden increases were similarly present for ‘*print*’ books in Humanities and Social Science fields, but not in Medicine. For more recent books or books first published from 2010 onwards, both print and electronic holdings gradually fell; however, the contractions were more significant for print books compared to ebooks. The pattern is in line with the U.S. academic and public libraries budgeting reports (Romano et al., 2015; Romano, 2016). Medicine and Political Science books make extreme cases for how magnitude of library eholdings have suddenly grown substantially bigger than print holdings and how systematically libraries have scaled their print collections down over years.

Figure 4 shows the median proportion of LEH to LPH in each title for six sample fields. Based on publication date of the first edition of books, the overall ratio of LEH to LPH showed more print holdings in pre-2000 periods (LEH/LPH < 1) and afterwards the average ratio has grown gradually over time to up to four times more eholdings than

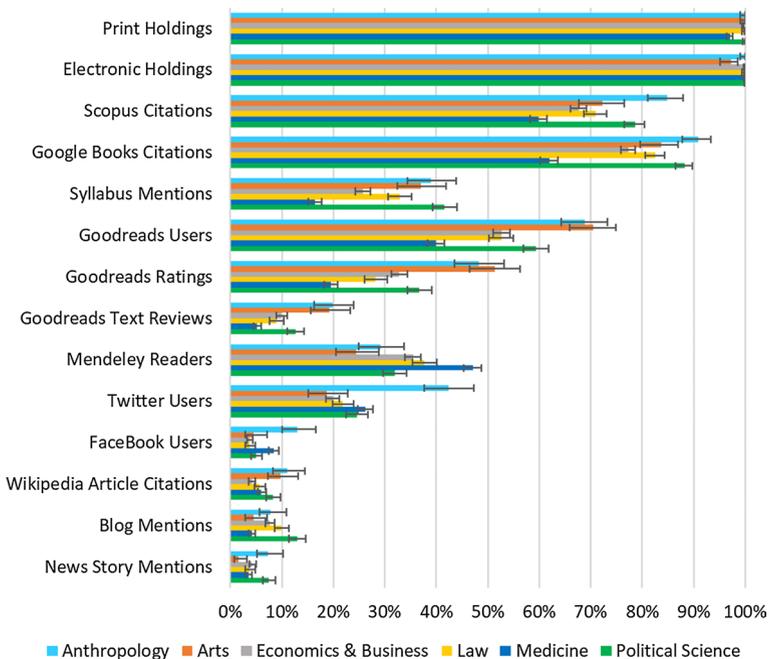


Fig. 13 Normalized Proportion Cited of books published 2012–2014 across all metrics and six fields

<sup>3</sup> - <https://doi.org/10.6084/m9.figshare.16531155>.

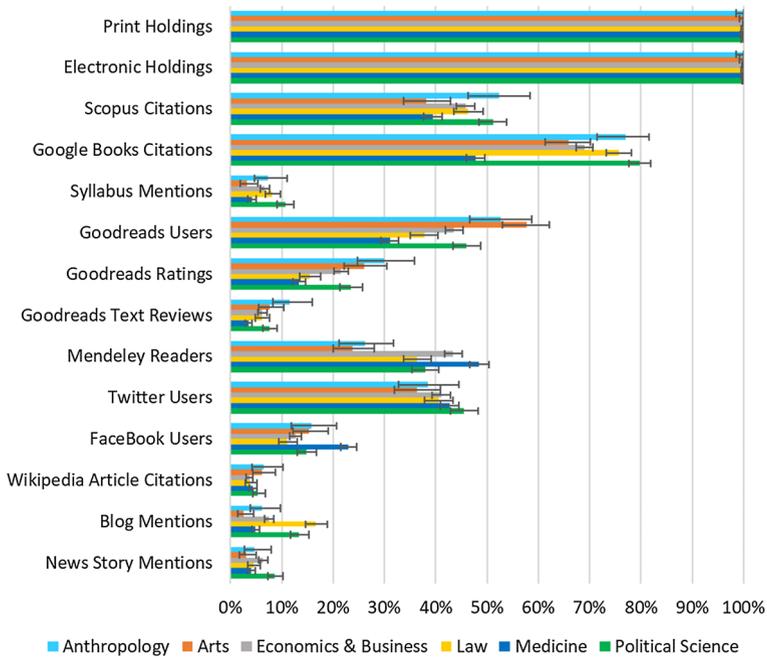


Fig. 14 Normalized Proportion Cited of books published 2015–2017 across all metrics and six fields

Table 4 Proportion Cited of books in six (out of 14) metrics across 26 fields (Continuation for Table 3)

Fields\Metrics	GTR	Mendeley	Twitter	Facebook	Wikipedia	Blogs	News	# of Books
Agriculture	7.8%	36.6%	21.2%	11.4%	6.5%	4.4%	4.2%	1,775
Anthropology*	20.4%	28.4%	33.4%	9.7%	10.3%	7.6%	5.8%	1,198
Arts*	14.9%	23.1%	23.3%	7.3%	8.4%	3.1%	2.2%	1,245
Business and Economics*	9.3%	33.3%	21.5%	4.9%	3.8%	6.2%	4.1%	11,987
Chemistry	1.6%	26.6%	13.8%	7.1%	4.6%	4.7%	2.8%	1,651
Education	10.9%	30.3%	24.8%	7.6%	3.1%	3.8%	3.8%	4,475
Engineering and Technology	3.5%	26.6%	15.7%	6.8%	3.2%	2.1%	1.9%	14,140
Ethics and Religion	28.3%	22.2%	20.3%	4.2%	8.5%	4.0%	3.7%	4,354
Geography	9.4%	33.6%	23.8%	7.6%	5.6%	6.4%	3.3%	1,367
History	27.0%	19.5%	21.1%	4.7%	12.4%	4.9%	4.1%	8,798
Languages and Literature	18.4%	21.4%	18.9%	4.2%	6.5%	3.5%	2.0%	9,798
Law*	9.1%	34.4%	23.2%	4.7%	4.9%	10.4%	3.7%	4,734
Library Science and Bibliography	20.2%	19.0%	25.2%	8.9%	5.3%	6.4%	0.4%	563
Mathematics	9.3%	42.7%	23.1%	6.9%	5.3%	3.3%	1.1%	4,093
Medicine*	6.5%	45.6%	23.8%	10.0%	6.4%	4.7%	3.7%	12,227
Military Science	16.9%	29.0%	23.7%	4.5%	6.7%	6.7%	6.0%	822
Music	25.3%	25.6%	26.2%	6.7%	15.5%	2.9%	3.4%	1,161
Natural History and Biology	10.5%	43.5%	27.8%	10.7%	7.4%	8.9%	3.7%	1,640
Philosophy	22.5%	27.6%	23.1%	6.6%	3.4%	4.3%	1.2%	1,935
Physics	7.6%	35.5%	18.8%	6.3%	5.8%	5.8%	3.0%	2,628
Physiology	6.0%	43.7%	26.6%	10.9%	6.3%	5.3%	3.4%	1,716
Political Sciences*	12.4%	31.7%	26.3%	6.2%	6.7%	10.7%	6.5%	5,098
Psychology	17.9%	33.6%	26.2%	9.0%	4.4%	7.1%	5.4%	1,996
Recreation and Leisure	23.5%	30.5%	30.3%	7.7%	8.9%	6.6%	5.8%	777
Social Sciences	15.9%	31.1%	27.1%	6.6%	6.1%	7.1%	5.8%	9,645
Zoology	15.9%	38.3%	30.3%	12.6%	15.7%	12.0%	7.2%	758
Unknown	8.7%	27.4%	18.1%	6.2%	6.8%	4.6%	3.1%	9,190
<b>Average of 26 Fields (excluding Unknown)</b>	<b>14.27%</b>	<b>31.29%</b>	<b>23.83%</b>	<b>7.45%</b>	<b>6.99%</b>	<b>5.88%</b>	<b>3.77%</b>	<b>110,603</b>
<b>Average of Six* Fields</b>	<b>12.08%</b>	<b>32.74%</b>	<b>25.27%</b>	<b>7.12%</b>	<b>6.78%</b>	<b>7.12%</b>	<b>4.31%</b>	<b>36,493</b>

\*: six sample fields in Dataset 2; GTR: Goodreads Text Reviews; Mendeley: Mendeley Readers; Twitter: Unique Twitter Users; Facebook: Unique Facebook Users; Wikipedia: Wikipedia Articles; Blogs: blog pages; News: news outlets.

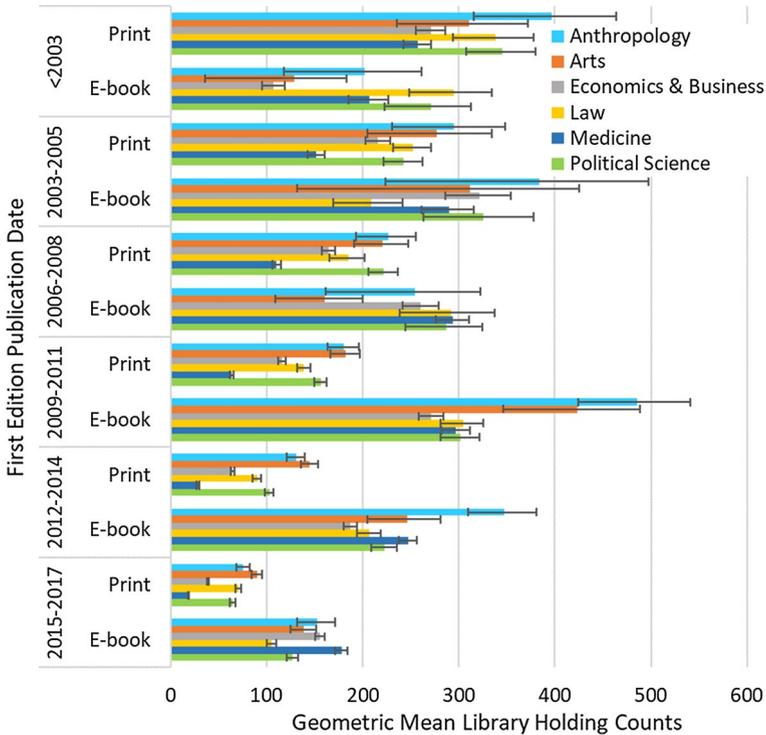


Fig. 15 Geometric mean metrics across six fields and over six time periods

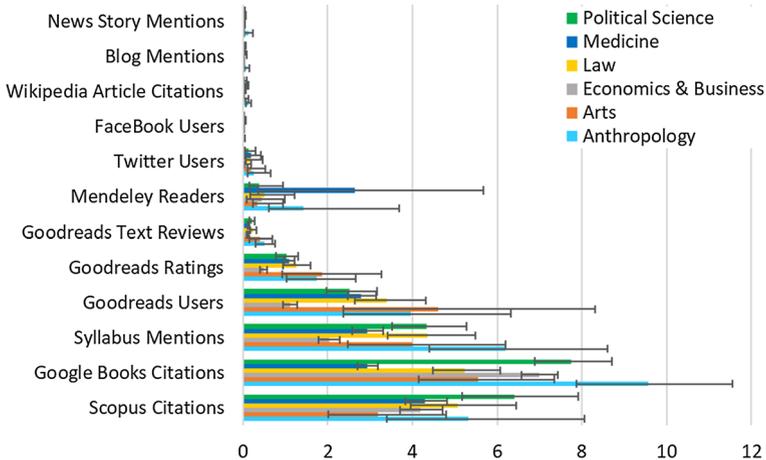


Fig. 16 Geometric mean metrics across six fields for books originally published prior to 2003

print ( $LEH/LPH > 1$ ). The disciplinary differences in year-by-year approach suggest that the median ratios ranged between 0.2 times for eighteenth century Business and Economics books to 14.2 times 2016 Medicine books (Fig. 4), for which there was a

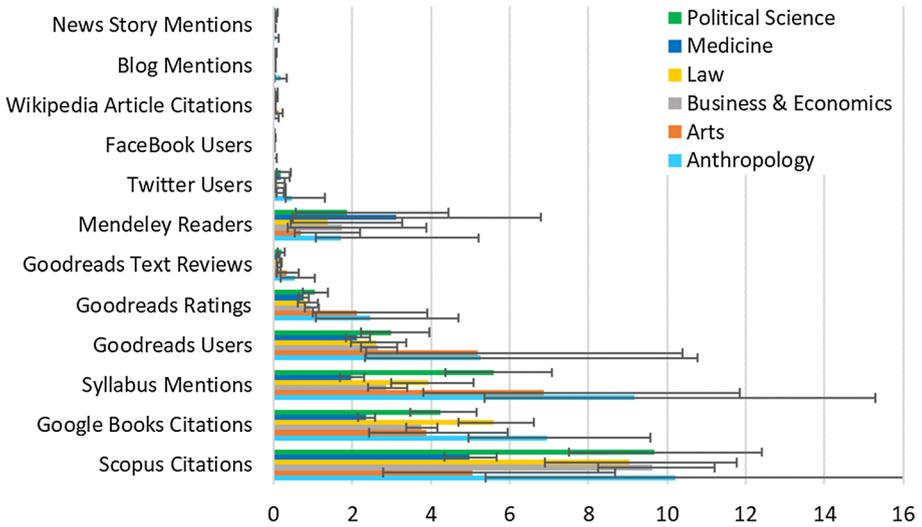


Fig. 17 Geometric mean metrics across six fields for books originally published in 2003–2005

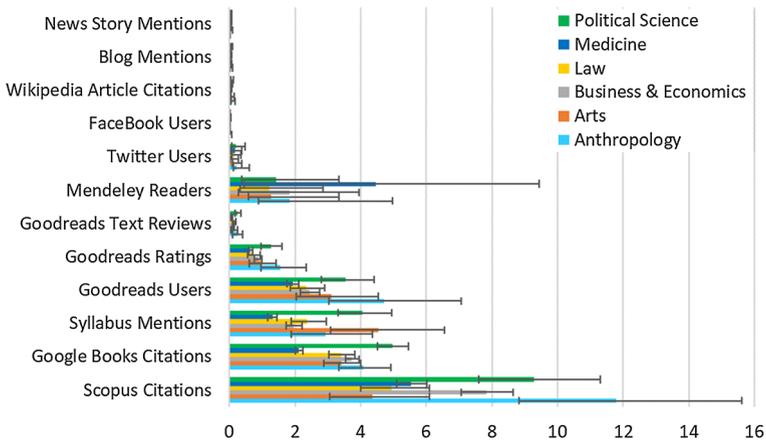


Fig. 18 Geometric mean metrics across six fields for books originally published in 2006–2008

dramatic growth in ebook acquisition. Among 26 fields, the median rate of this relation in disciplinary level was the lowest in Law at 1.7 times and the highest in Medicine at about 6 times. Overall, the average number of libraries with e-versions of books was 7.4 times (median = 2.6 times) that of libraries with print version in the whole Dataset 1.

**Distribution of library print vs. e-book holding counts**

Figure 5 shows distribution of various holding counts in a logarithmic scale for all books in Dataset 1. The overall patterns suggest that the highest frequency TLH counts was at 139

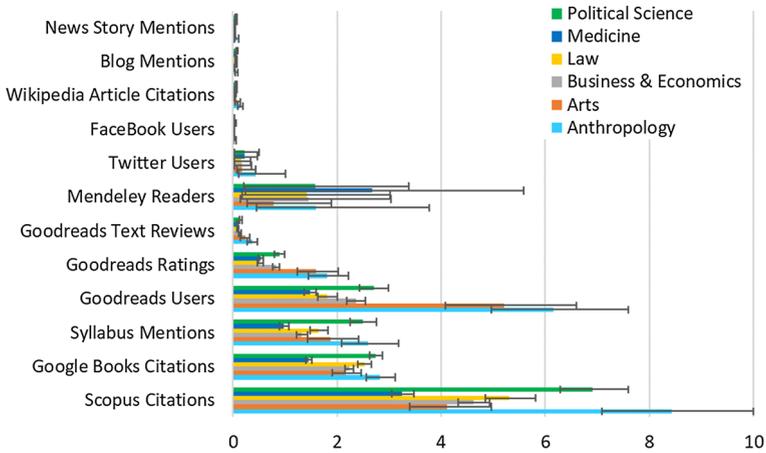


Fig. 19 Geometric mean metrics across six fields for books originally published in 2009–2011

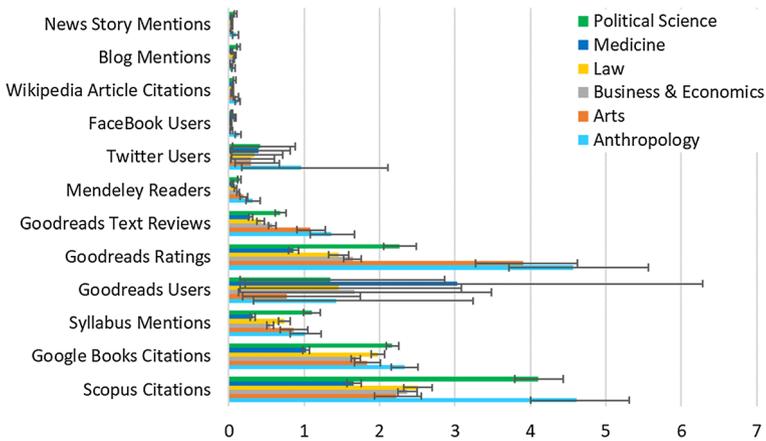


Fig. 20 Geometric mean metrics across six fields for books originally published in 2012–2014

(frequency = 463), LPH counts (904) was at 13, and LEH counts peaked twice at 2 (frequency = 1803) and 85 (frequency = 990).

Figures 6, 7 and 8 show the changes in the distribution pattern of holdings over the six time periods for Dataset 2. Let  $\lambda$  be the LH count with the highest probability of occurrence and  $p$  be the proportion of books with  $\lambda$  occurrence. For print holdings, the distribution patterns indicate that the densest holding count (2.1%) is 10 (2015–17) and by increasing age of books  $\lambda$  gradually increases reaching 175 in final period (175) (Fig. 7). For eholdings, the distribution patterns are less consistent and despite a general reduction in  $\lambda$  from 84 (1.2% of 2015–17 books) to 1 (3% of prior to 2003 books) the pattern is broken at 2009–11 (Fig. 7). Multiple peak points suggest that distribution of TLH is more similar to LEH (Fig. 6).

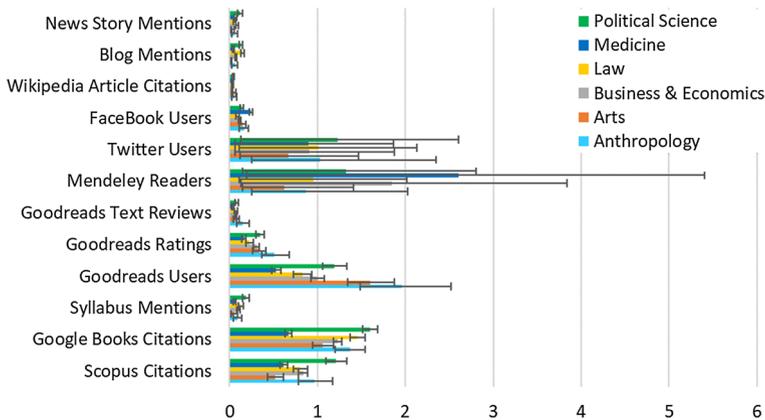


Fig. 21 Geometric mean metrics across six fields for books originally published in 2015–2017

### Proportion cited of books in metrics across fields

In answer to the second research question, Tables 3 and 4 offer a comparison of proportion mentioned of books across metrics. The average proportion of books with at least one library holding count in OCLC was 99.7% for print books and 99.1% for ebooks which corroborate with proportions (Table 1) reported for library holding in previous studies (Kousha, Thelwall and Abdoli, 2017; Halevi et al., 2016; Torres-Salinas et al., 2017). The only prominent exception, for less clear reasons, is *Engineering and Technology* where the percentage of books with at least one LEH was much lower (93.2%) than ones with LPH, suggesting that about 7% of titles are not available in ebook format through libraries (Table 3). From another point of view, the proportions of titles that were exclusively available in one format, either print or electronic, did not exceed 1 to 2%, with more often non-digitized print books rather than non-print electronic books. Perhaps not surprisingly, Library Science and Bibliography was the only class that had neither print nor electronic books without the alternate format.

As with other metrics, the highest average proportions cited of books across 26 fields were seen for Google Books Citations (74.12%), Scopus Citations (69.39%), and Goodreads Reviews (55.17%), which all agree with the previous observations (reviewed in Table 1). The average proportion of books with Syllabus Mentions acquired from *opensyllabus.org* explorer (35.46%) is also consistent with previous studies (Kousha & Thelwall, 2015, 2016), but *altmetric.com* results were much lower (6.7%) (Both in Table 3). Average percentage of books with Mendeley Readers (31.29%) was slightly below Halevi et al., (2016) (43.1%), but above Torres-Salinas (2017) (24.9%), perhaps because of different primary sources of data. Less than one quarter of books had Twitter User mentions (23.83%), and much less Facebook Wall mentions (7.45%), Wikipedia Citations (6.99%), Blogs (5.88%), and News Post mentions (3.77%) for which there was no major reports for comparisons.

In disciplinary level, the highest proportion of book titles with at least one Scopus Citation was 80.1% in Chemistry. Proportions of books with Google Books Citations ranged between 60.5% in Physiology and 85% in Philosophy. Proportion of books with Syllabus Mentions changed between a quarter of books in Medicine (25.6%) to below half of them

in Anthropology (44.9%). For Goodreads Users, Goodreads Ratings and Goodreads Text Reviews the least proportion covered was in Chemistry (29.1, 10.8, 1.6%, respectively) and the most in Ethics and Religion (72.3, 54.5 and 28.3%, respectively). As with Mendeley Readers, the proportions of books in Library Science and Bibliography (19%) were the least and substantially below Medicine (45.6%) as the highest. However, Twitter Users uptake of books ranged between 13.8% in Chemistry and 33.4% in Anthropology. Along with all other metrics, the highest proportion of books in Facebook (12.6%), Wikipedia (15.7%), Blogs (12%) and News Post (7.2%) was seen in Zoology (Table 4).

### Normalised proportion cited of books in metrics

In answer to the second and third research questions, Figs. 9, 10, 11, 12, 13 and 14 gives a comparison of the Normalized Proportion Cited (NPC) of books across 14 different indicators each at one period of time based on the original publication date (Dataset 2). There was an almost comprehensive coverage of books through LPH and LEH across all fields and time periods (over 97%) which is substantially above coverage of other book metrics. It suggests that Library Holdings can be available for majority of books even at a short time after book publication (2015–2017). But the error bars suggest that accuracy of NPCs for older books in Arts and Anthropology can be lower due to smaller sample sizes, although high and not below 90%. Furthermore, normalized proportion holdings of prior to 2003 books were in fact less likely to be as comprehensive as more recent books.

Three trends were observed in the remaining metrics. NPC Google Books Citations, Syllabus Mentions, and Goodreads Text Reviews, and Wikipedia Article Citations consistently dropped over years with fewer more recent books cited. NPC Scopus Citations, Goodreads Users and Goodreads Ratings showed an increase until 2006–2008 and then gradually dropped. In contrast, a steady rising trend occurred for NPC Mendeley readers, Twitter User, Facebook Wall, Blog, and News Post mentions throughout the Dataset 2, suggesting the increasing availability of altmetric indicators in more recent years. Nevertheless, among the variables, Google Books Citations was the most available metric for books after library print and electronic holdings either in the short term (Fig. 14, about 75% for Social Science and Humanities fields in 2015–2017) or the long term (Fig. 9, about 85% in prior to 2003), perhaps suggesting that book-sourced citations tend to appear faster than Scopus Citations which indicates a substantial drop over time (from around 85% in 2006–2008 to about 45% in 2015–2017) or Syllabus Mentions (from 65% in 2003–2005 to 5% 2015–2017). An assessment at subject level provided the following observations.

*Medicine* Books in Medicine consistently showed the lowest NPC ‘Scopus Citations’, ‘Syllabus Mentions’, ‘Google Books’, and ‘Goodreads’ engagements, suggesting substantial disciplinary differences with Social Science and Humanities for books as the reverse can be expected for journal articles. In contrast, Mendeley uptake of books in Medicine fluctuated between about 40% and 60%, in other fields it was often below 40%, suggesting lower usage of Mendeley in book-centered fields (Kousha & Thelwall, 2017). The highest uptake of books in Twitter and Facebook was also in Medicine.

*Political science* In 2015–2017, NPC Scopus Citations, Google Books, and Syllabus Mentions in *Political Science* (44, 67, and 6%, respectively) were higher than those of other fields. It suggests that uptake of *Political Science* books in research and education is faster than the average uptake in other fields. The highest publication uptakes in Blogs and News Posts were in Political Science.

*Anthropology and art* The highest proportion cited of books was seen at 2006–2008 with 95% in Scopus for Anthropology, and in Google Books for Arts, suggesting the delayed effect of citation accumulation in these fields. In three metrics of Goodreads User, Rating and Text Review, a similar pattern as above persists with *Arts* (64, 46, 17%, respectively) and *Anthropology* (62, 47, 21%, respectively) appearing on top.

*Law* Perhaps the surprising case is *Law*, where NPC of Goodreads Users of books authored in years prior to 2008 (about 66%) had higher possibility of having been shelved by Goodreads Users than those in other fields (57% on average).

### Average of book metrics over years

In response to the second and third RQs, geometric mean library print holdings and library eholdings are compared (Fig. 15). LEH average counts are the most numerous among all metrics, reflecting the massive collection of ebooks in libraries. Geometric mean LEH have undulated over time with a steadier drop from 2009–2011 onwards in all fields. However, geometric mean LPH showed a steady decline over time in all fields. As other metric counts were substantially less numerous than library holding counts in all years, they are embedded separately in Figs. 16, 17, 18, 19, 20 and 21, each figure illustrating one time period.

On average the geometric mean of library print holdings for books originally authored *before 2003* (approximately 320) is considerably larger than their library electronic holdings (approximately 202) in all fields, whereas the reverse is true in the following time periods. For books *After 2003*, geometric mean LEH for *Medicine* undulated between about 178 and 297, yet LPH descended with great disparity from 151 in 2003–2005 to 18 in 2015–2017. After 2003, *Anthropology* makes an extreme case for the highest geometric mean library holdings of both ebooks (irregularly changing between 152.8 and 486) and print books (descending from 295 to 75). These all suggest that library holdings are over time cumulative in nature, both in print and electronic format.

Majority of metrics indicate an overall drop in average metric counts of books over time except in Twitter, Facebook, Blogs and News post mentions, that have consistently increased over years and this pattern is recurrent for the six fields. Geometric mean Scopus Citations decreased from around 10 to about 1, Google Books Citations from around 8 to about 2, Syllabus mentions from around 6 to below 1, Goodreads Users from around 4 to 1, Goodreads Ratings from about 2 to 1, Goodreads Text Reviews from about 0.3 to 0.1, Goodreads Average Ratings from 1 to 0.5, Mendeley Readers from around 2 to below 1 (except *Medicine* from over 4 to slightly below 3), and Wikipedia Articles from about 0.1 to about 0.05. However, social media related metrics, including geometric mean Twitter Users, Facebook Wall Mentions, Blogs and News Posts showed a gradual growth over years ending the period at fields mean of 1.38, 0.96, 0.15, 0.08, 0.07, respectively.

### Highly acquired book titles in libraries: print vs. e-books

Age and subject theme were the two predominating characteristic distinguishing books with extremely high library print holdings (old books) and electronic holdings (recent books). In *Medicine*, most highly acquired library *print* holdings were often in “basic and theoretical aspects” of medicine, specifically for nursing, whereas for *eholdings* they were more focused on “evidence-based textbooks and practical guidebooks”. In *other fields*,

the high acquisition of “handbooks” and “historical books” in *print* version is dominant, whereas highly acquired ebooks represented more modern topics such as “*media in political events*”, “*contemporary theories in economics*”, “*law in medicine*”, “*modern arts in different nations*”, and “*recent narratives of various ethnicities*” (see titles and holding counts in Online Appendix, Tables 1 and 2).

To obtain a comprehensive list of reasons for libraries preferring print books over ebooks or vice versa, there would be a need to conduct survey. Nevertheless, some of the characteristics is tried to be speculated here. Since it is obvious to expect higher library holding counts for older books, the holding data is investigated for a list of top ten print and ebook holdings of book titles with first edition published in 2015 onwards (Online Appendix, Tables 3 and 4) and compared with their rank in the other book indicators. In less frequent cases, it was seen that the print holding counts of books substantially exceeded that of eholdings, for instance, Armitage (2017) “*Civil wars: A history in ideas*” (1095 print holdings vs. 94 eholdings) and Scott (2017) “*Against the grain: A deep history of the earliest states*” (1000 print holdings and 27 eholdings). A general look at the characteristics of these books in WorldCat shows that they are extensively available in different countries and translated quite early (within two years) to at least five other major languages, suggesting at their universal significance.

On the other hand, the highly acquired titles in electronic format (examples below) are often in a single language (mostly English). Some book titles have too many pages and perhaps lower sales rate for print version, for instance Houseman and Mandel (2015) “*Measuring globalization better trade statistics for better policy: Biases to price, output, and productivity statistics from trade*” which has 649 pages and 1181 eholdings compared to only 49 print holdings by the date data was collected. Alternatively, other book titles not only are heavy, but also have audience who are comfortable with ebooks particularly in Medicine, for instance Tyagi and Prasad (2016) “*Gastrointestinal Cancers: Prevention, Detection and Treatment*” which has 415 pages and 1122 eholdings but 4 print holdings. Many illustrations can be another reason, perhaps not always because of the cost of colorful print version but because its audience prefer a flexible screen to view rather than a fix-sized print, especially in Arts such as for Lauzon (2016) *The Unmaking of Home in Contemporary Art* (1119 eholdings and 152 print holdings).

With regards to other metric counts, of the post-2015 book titles for all six fields at the top ten print holdings and eholdings only 28 (47%) and 17 (28%) titles, respectively, appeared at the top ten of at least one other indicator in terms of frequency of mentions, suggesting that broadly available print books in libraries are generally more likely to enjoy popularity of different kinds than electronic ones. For instance, Scott (2017) “*Against the grain: A deep history of the earliest states*” had a considerable number of Goodreads Users (6290), unique Twitter user mentions (44) and Blog citations (8) placing it within top ten books of those metrics in Anthropology, as well. The field, however, with the highest number of titles (6) with at least one title among the top ten of the other indicators was the Political Science. Some of titles in this field have been the top mentions of several social media platforms such as Hopkins (2017) “*Red Fighting Blue: How geography and electoral rules polarize American politics*” (434 print and 52 electronic holdings) with 67 unique Twitter user and 4 unique Facebook user mentions, and 2 Wikipedia citations. But majority of book titles were often enjoying the highest popularity of a limited number of platforms, and it could often only be the Goodreads which attracted the highest attention for altogether 23 titles of top print holdings in contrast to only 6 of top eholdings for six fields. Perhaps, however, interestingly Social Science and Humanities fields are more likely to enjoy the cross platform popularity across social media-related metrics compared to

Medicine that with its titles either on the top ten print or electronic holdings list could only appear among top ten titles in Scopus, Syllabus mentions and Goodreads, suggesting that high library holding counts for books in medicine can barely influence their social uptake.

## Discussions

One general limitation in the studies of book metrics concerns with editions. In books, each title gets a different ISBN across different editions, helping to identify each newer edition from a previous one. Additionally, electronic publishing of each book also results in assignment of a new ISBN, E-ISBN, specified only to identify the electronic version. Overall, this means that one title is specified by a number of different ISBNs and hence owns an ISBN family. This leads to problems when tracking citations, since references often do not indicate edition number and whether the author have used the electronic version or print format. Although new editions can be important for authors in assessment, it is not clear how much citation information at edition or publishing format (i.e., print or electronic) levels will be useful for impact assessment. The overall unified citation impact at title level, however, seems relevant. Zuccala et al., (2018) proposed introducing a unifying ID for each ISBN family or each single title regardless of edition counts. Original publication year is important because older books tend to accrue more citations over time. Nevertheless, using all edition's publication year can create the illusion of citation impact in short period of time for books that have been around for a longer period of time. Although in this research we did not identify ISBN families, we addressed the problem of multiplicities and editions of books by identifying the original or first edition publication year of books.

Another major limitation in this research was the discovered inconsistency between the proportion of syllabus mentions obtained from *altmetric.com* and that of previous studies. It was found that results directly acquired from *Opensyllabus.org* are more comprehensive (Table 3). However, as the author was not successful in getting an API key from this website, manual search results in *Opensyllabus.org Explorer* were used and therefore only six fields were chosen for this analysis instead of all the 26 fields.

In answer to the first research question, in terms of availability, it was seen that despite comprehensive accessibility of books of all ages in libraries either in print or electronic, on average the scale of libraries' ebook holdings is about seven times (median = three times) larger than libraries with print holdings. On the one hand, this disparity in library holdings between print and ebooks is growing disproportionately for more recent books, presumably as a consequence of a budget shift from print to electronic. On the other hand, it tends to suggest availability of books in many libraries and could be the results of reducing costs and maximizing circulation over academics. Unlike in humanities, the ratios show dramatic increase in medicine, suggesting a considerable tendency in medicine toward electronic holding acquisition.

Although the distribution analysis is intended to help the bibliometric analysts, there is a possibility of library managers getting informed and benefiting from that. The distribution patterns of print holdings over time were similar to Poisson distribution with the highest number of libraries with print holdings consistently growing bigger over time up from 10 to 175 despite dropping in peak frequency. However, distribution of library holding counts were inconsistent, recurring in several peak points in the same time period, without surpassing about 90 library holdings. The irregularity of eholdings can be related to complex marketing strategies of publishers as well as library purchase models. In sum, the

variation between print holdings and eholdings suggest that total holdings need to be used more cautiously, since unlike eholdings, print holdings indicate a more stable distribution pattern. Thus, there is a need for further analysis to identify the differences of these metrics in book impact assessment.

In answer to the second and third RQs, the proportion cited of books in different metrics suggested that overall, in line with previous studies (Halevi et al., 2016) library holdings is the most comprehensive metric as a means of assessing books. The availability of this metric is substantially above other metrics even after a short period of time, while Google Books is available for only slightly more than 50% of books and all other metrics for less than half.

The findings in answer to the third RQ showed that library holdings either as print or electronic tend to be accrued over time like formal citations. One important finding in this part is that substantial number of library holding counts could be expected to appear earlier than the appearance of other metrics for books, suggesting that library holdings is a relatively early indicator to capture book impact. It also indicated that ebooks of older books are less likely to be available in libraries than their print, which suggests limited usefulness of electronic holdings. However, for more recent books both formats are likely to be available in lower frequency with still print holding even less frequently available than eholdings, mostly suggesting a clear budget shift from print to electronic for more recent book publications. Both print and electronic book holdings in libraries saw sudden jumps in 2010, especially in Social Science and Humanities. This had influenced the average number of libraries acquiring ebooks but did not have a significant impact on the consistency of the downward trend of mean print book holdings. It suggests that unlike the well-established print holding statistics, the eholdings are more likely to produce erratic holding counts. The broad error bars, however, suggest that Mendeley is not a good early source of book impact compared to other indicators such as Google Books and Goodreads Users. Scopus Citations are also not useful for books, unless a few years have passed from the first edition publication.

Unlike journal articles, books are subject to numerous republications and editions over time, that lead to duplication of books across publication dates. Using initial edition publication date of books is more important than republication or current edition date that is more immediately available. Using first edition publication date helped in this research to see that like formal citations, library holding counts are accrued over time either in print or in electronic. However, given the importance of first edition publication date of books and as the first publication date is not an immediately available data, in local book impact assessment practices there is a need to interpret holding trends with caution particularly if the publication dates are directly used from the publishers as in Scopus or even the local library database. Furthermore, studying a few examples showed that library holding counts can be irregularly changing for each edition of a book, depending on the significance of content contribution, the age of book, and the extent of acquisition for former edition(s). Thus, relative to the purpose of assessment, a book holding statistic available for whole editions of a book or the book family needs to be distinguished from edition level statistics.

In answer to the fourth research question, it was seen that highly acquired print books in libraries are basically considered universally important works, as they tend to be translated to various languages, whereas the highly acquired books in electronic format might have non-comprehensive list of other features such as high cost of print alternatives, large quantity of content (number of pages), benefit of virtual features such as high-quality visual content. In connection with other metrics, it can be expected to see high frequency of social media mentions for books with high library print holding counts, but perhaps not as

much frequently for highly acquired ebooks in libraries. In line with the findings of Kousha et al., (2017), the findings in this part of study also suggest that Goodreads is a more versatile platform attracting usages for variety of different academic, educational and social purposes as it groups with various social media metrics in terms of highly mentioned books. But there also is a clear separation of metrics that need to be noticed in assessment practices as works with the highest Scopus Citations, and Blog mentions indicate different orientation from Twitter, Facebook and Wikipedia.

## Conclusion

This research attempted to unravel library holding as an indicator of scholarly book impact assessment, and found that there is a need to distinguish the differences between electronic holdings and print holdings before using their combined format as Total holding from OCLC. The differences of library holding with other metrics in terms of availability, average uptake and timely uptake also suggest serious implications about the benefits of library holding and particularly print holdings over other metrics. The descriptive findings in this research about the characteristics of library print and electronic holdings suggest that there is a need for another study to identify the relationship of print and electronic holdings with other research metrics to expand our understanding of library holding metric as a significant indicator for book impact assessment.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11192-021-04220-6>.

**Acknowledgements** Author appreciates *Statistical Cybermetrics Research Group (SCRG)* for the software *Webometric Analyst* and *Altmetric.com*, *Goodreads.com* and *OCLC.org* teams for supporting API accesses to data. Author also appreciates anonymous reviewers for contributing in the article with their invaluable comments.

**Funding** Open Access funding provided by University of Turku (UTU) including Turku University Central Hospital.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adie, E., & Roe, W. (2013). Altmetric: Enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26(1), 11–17. <https://doi.org/10.1087/20130103>
- Bunkell, J., & Dyas-Correia, S. (2009). E-books vs print: Which is the better value? *The Serials Librarian*, 56(1–4), 215–219. <https://doi.org/10.1080/03615260802698283>
- Casselden, B., & Pears, R. (2020). Higher education student pathways to ebook usage and engagement, and understanding: highways and cul de sacs. *Journal of Librarianship and Information Science*, 52(2), 601–619. <https://doi.org/10.1177/0961000619841429>

- Conyers, A., Lambert, J., Wong, L., Jones, H., Bamkin, M., & Dalton, P. (2017). E-book usage: Counting the challenges and opportunities. *Insights*, 30(2), 23–30. <https://doi.org/10.1629/uksg.370>
- Cox, L. (2011). Librarians' use of usage statistics for journals and e-books. *Learned Publishing*, 24(2), 115–121. <https://doi.org/10.1087/20110206>
- Erfanmanesh, M., Noorhidawati, A., & Abrizah, A. (2019). What can Bookmetrix tell us about the impact of Springer Nature's books. *Scientometrics*, 121(1), 521–536. <https://doi.org/10.1007/s11192-019-03198-6>
- Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123. <https://doi.org/10.2196/jmir.2012>
- Folb, B. L., Wessel, C. B., & Czechowski, L. J. (2011). Clinical and academic use of electronic and print books: The health sciences library system e-book study at the University of Pittsburgh. *Journal of the Medical Library Association: JMLA*, 99(3), 218–228. <https://doi.org/10.3163/1536-5050.99.3.009>
- Goodwin, C. (2014). The e-Duke scholarly collection: E-book v print use. *Collection Building*, 33(4), 101–105. <https://doi.org/10.1108/CB-05-2014-0024>
- Gu, X., & Berger, P. D. (2018). Who Reads E-Books? An Analysis Of Key Factors In E-Book Reading In The US. *Advances in Social Sciences Research Journal*, 5(1), 126–137. <https://doi.org/10.14738/assrj.51.4108>
- Halevi, G., Nicolas, B., & Bar-Ilan, J. (2016). The complexity of measuring the impact of books. *Publishing Research Quarterly*, 32(3), 187–200. <https://doi.org/10.1007/s12109-016-9464-5>
- Haugh, D. (2016). How do you like your books: Print or digital? an analysis on print and e-Book usage at the graduate school of education. *Journal of Electronic Resources Librarianship*, 28(4), 254–268. <https://doi.org/10.1080/1941126X.2016.1243868>
- Henning, V., & Reichelt, J. (2008). Mendeley-a last. fm for research?. In *2008 IEEE Fourth International Conference on eScience* (pp. 327–328). <https://doi.org/10.1109/eScience.2008.128>
- Ismail, R., & Zainab, A. N. (2007). Factors related to e-books use amongst IT students. In A. Abdullah, et al. (Eds.), *ICOLIS 2007, Kuala Lumpur: LISU, FCSIT* (pp. 265–275). [http://dspace.fsktm.um.edu.my/bitstream/1812/260/1/26MY\\_Roesnita\\_OK.pdf](http://dspace.fsktm.um.edu.my/bitstream/1812/260/1/26MY_Roesnita_OK.pdf)
- Kohn, K. (2018). Using logistic regression to examine multiple factors related to e-book use. *Library Resources & Technical Services*, 62(2), 54. <https://doi.org/10.5860/lrts.62n2.54>
- Kont, K.-R. (2016). Demand-driven E-book program in tallinn university of technology library: The first two years of experience with the EBL platform. *Slavic & East European Information Resources*, 17(1–2), 36–67. <https://doi.org/10.1080/15228886.2016.1171085>
- Kousha, K., & Thelwall, M. (2015). An automatic method for extracting citations from google books. *Journal of the Association for Information Science and Technology*, 66(2), 309–320. <https://doi.org/10.1002/asi.23170>
- Kousha, K., & Thelwall, M. (2016). An automatic method for assessing the teaching impact of books from online academic syllabi. *Journal of the Association for Information Science and Technology*, 67(12), 2993–3007. <https://doi.org/10.1002/asi.23542>
- Kousha, K., & Thelwall, M. (2017). Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology*, 68(3), 762–779. <https://doi.org/10.1002/asi.23694>
- Kousha, K., Thelwall, M., & Abdoli, M. (2017). Goodreads reviews to assess the wider impacts of books. *Journal of the Association for Information Science and Technology*, 68(8), 2004–2016. <https://doi.org/10.1002/asi.23805>
- Kousha, K., Thelwall, M., & Rezaie, S. (2011). Assessing the Citation Impact of Books: The Role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, 62(11), 2147–2164. <https://doi.org/10.1002/asi.21608>
- Littman, J., & Connaway, L. S. (2004). A Circulation Analysis Of Print Books And e-Books In An Academic Research Library. *Library Resources & Technical Services*, 48(4), 256–262.
- Maceviciute, E., Wilson, T. D., GudinaVICIUS, A., & Šuminas, A. (2017). E-books in academic libraries: Results of a survey carried out in Sweden and Lithuania. *Information Research: an International Electronic Journal*, 22(3), n3. Accessible at <http://InformationR.net/ir/22-3/paper762.html>
- Majid, S., Chenqin, Y., Chang, Y., & Zilu, C. (2019). Perceptions and e-book use behavior of university students. *International Journal of Digital Society*, 10(4), 1534–1540. <https://doi.org/10.20533/ijds.2040.2570.2019.0190>
- Maleki, A. (2020). P-Libcitation vs. E-libcitation?: Libraries' Print Book Holdings Resonate with Citations and Altmetrics But E-book Holdings Do Not. The 2020 Online Altmetrics Workshop. Accessible at <http://altmetrics.org/altmetrics20/>

- Parkes, D. (2007). E-books from ebrary at Staffordshire University: A case study. *Program: Electronic Library and Information Systems.*, 41(3), 253–261. <https://doi.org/10.1108/00330330710774138>
- Romano, R., Girmscheid, L., & Genco, B. (2015). Sixth annual survey ebook usage in U.S. public libraries. *Library Journal*. Accessible at [https://s3.amazonaws.com/WebVault/ebooks/LJSLJ\\_EbookUsage\\_PublicLibraries\\_2015.pdf](https://s3.amazonaws.com/WebVault/ebooks/LJSLJ_EbookUsage_PublicLibraries_2015.pdf)
- Romano, R. (2016). Survey of ebook usage in US academic libraries 2016. *Library Journal*. Accessible at [https://s3.amazonaws.com/WebVault/research/LJ\\_2016\\_EbookUsage\\_AcademicLibraries.pdf](https://s3.amazonaws.com/WebVault/research/LJ_2016_EbookUsage_AcademicLibraries.pdf)
- Springer (2010) A Survey of eBook Usage and Perceptions at the University of Liverpool: University of Liverpool eBook Study: part 2. White paper. Retrieved from: [http://www.springer.com/cda/content/document/cda\\_downloaddocument/V7671+Liverpool+White+Paper+Part2.pdf](http://www.springer.com/cda/content/document/cda_downloaddocument/V7671+Liverpool+White+Paper+Part2.pdf). Accessed at: 14 May 2020.
- Thelwall, M. (2017). *Web indicators for research evaluation: A practical guide*. San Rafael: Morgan & Claypool. <https://doi.org/10.2200/S00733ED1V01Y201609ICR052>
- Thelwall, M., & Fairclough, R. (2015). The influence of time and discipline on the magnitude of correlations between citation counts and quality scores. *Journal of Informetrics*, 9(3), 529–541. <https://doi.org/10.1016/j.joi.2015.05.006>
- Torres-Salinas, D., Arroyo-Machado, W., & Thelwall, M. (2020). Exploring WorldCat Identities as an altmetric information source: A library catalog analysis experiment in the field of Scientometrics. arXiv preprint [arXiv:2011.12626](https://arxiv.org/abs/2011.12626).
- Torres-Salinas, D., & Arroyo-Machado, W. (2020). Library Catalog Analysis and Library Holdings Counts: origins, methodological issues and application to the field of Informetrics. In *Evaluative Informetrics: The Art of Metrics-Based Research Assessment* (pp. 287–308). Springer, Cham. [arXiv:1910.03855](https://arxiv.org/abs/1910.03855)
- Torres-Salinas, D., Gumpenberger, C., & Gorraiz, J. (2017). PlumX as a potential tool to assess the macroscopic multidimensional impact of books. *Frontiers in Research Metrics and Analytics*, 2, 5. <https://doi.org/10.3389/frma.2017.00005>
- Torres-Salinas, D., & Moed, H. F. (2009). Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics. *Journal of Informetrics*, 3(1), 9–26. <https://doi.org/10.1016/j.joi.2008.10.002>
- Torres-Salinas, D., Robinson-García, N., & Gorraiz, J. (2017). Filling the citation gap: measuring the multidimensional impact of the academic book at institutional level with PlumX. *Scientometrics*, 113(3), 1371–1384. <https://doi.org/10.1007/s11192-017-2539-z>
- Wells, D., & Sallenbach, A. (2015). Books and ebooks in an academic library. *The Australian Library Journal*, 64(3), 168–179. <https://doi.org/10.1080/00049670.2015.1041216>
- White, H. D., Boe, Yu., et al. (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, 60(6), 1083–1096. <https://doi.org/10.1002/asi.21045>
- Worden, A., & Collinson, T. (2011). Engaging Staff and Students with E-books in a University Setting. *E-books in libraries: A practical guide*, 237–251.
- Zhang, T., Niu, X., & Promann, M. (2017). Assessing the user experience of e-books in academic libraries. *College & Research Libraries*, 78(5), 578.
- Zhu, X. (2018). E-book ill in academic libraries: A three-year trend report. *The Journal of Academic Librarianship*, 44(3), 343–351. <https://doi.org/10.1016/j.acalib.2018.03.006>
- Zuccala, A., & Guns, R. (2013). Comparing book citations in humanities journals to library holdings: Scholarly use versus perceived cultural benefit. In *14th international conference of the international society for scientometrics and informetrics* (pp. 353–360).
- Zuccala, A. A., & White, H. D. (2015). Correlating Libcitations and Citations in the Humanities with WorldCat and Scopus Data. In A. A. Salah, Y. Tonta, A. A. Akdag Salah, C. Sugimoto, & U. Al (Eds.), *Proceedings of the 15th International Society for Scientometrics and Informetrics (ISSI), Istanbul, Turkey, 29th June to 4th July, 2015* (pp. 305–316). Bogazici University.
- Zuccala, A., Breum, M., Bruun, K., & Wunsch, B. T. (2018). Metric assessments of books as families of works. *Journal of the Association for Information Science and Technology*, 69, 146–157. <https://doi.org/10.1002/asi.23921>