Research Paper

JIS

Journal of Information Science
1–15

DOI: 10.1177/01655515241245952
journals.sagepub.com/home/jis


Benefits of open access to researchers from lower-income countries: A global analysis of reference patterns in 1980–2020

Henrik Karlstrøm
Nordic Institute for Studies in Innovation, Research and Education, Norway

Dag W Aksnes
Nordic Institute for Studies in Innovation, Research and Education, Norway

Fredrik N Piro
Nordic Institute for Studies in Innovation, Research and Education, Norway

Abstract
The main objective of the open access (OA) movement is to make scientific literature freely available to everyone. This may be of particular importance to researchers in lower-income countries, who often face barriers due to high subscription costs. In this article, we address this issue by analysing over time the reference lists of scientific publications around the world. Our study focuses on key issues, including whether researchers from lower-income countries reference fewer publications in their research and how this trend evolves over time. We also investigate whether researchers from lower-income countries rely more on the literature that is openly available through different OA routes compared with other researchers. Our study revealed that the proportion of OA references has increased over time for all publications and country groups. However, publications from lower-income countries have seen a higher growth rate of OA-based references, suggesting that the emergence of OA publishing has been particularly advantageous to researchers in these countries.

Keywords
Bibliometrics; information access; knowledge organisation; open access; scientific publishing

1. Introduction

A fundamental objective of the open access (OA) movement is to make scientific literature freely available to everyone [1]. The rise of OA has generated a vast literature corpus, mostly related to the production of OA, for example, studying the costs of producing OA [2], the citation impact of OA papers [3,4] or empirical studies about where OA publishing takes place [5], largely supporting the expectation that OA will be most useful for researchers from low-income countries [6]. These topics are covered extensively in the literature (see Simard et al. [7] for an overview of global and regional differences in OA publishing). Recent studies indicate that the percentage of OA publications indexed in the Web of Science (WoS) is approaching 50%, with 42.9% of WoS publications being OA in the period 2015–2019 [7], a share that has increased steadily over the last two decades [8,9].

In the debate about the benefits of OA, lower-income countries have typically been seen as illustrating the advantages (and disadvantages) of OA. Changing the costs of publishing from subscription fees to author payments [10] has inevitably

Corresponding author:
Henrik Karlstrøm, Nordic Institute for Studies in Innovation, Research and Education, 0653 Oslo, Norway.
Email: henrik.karlstrom@nifu.no

sparked a discussion about whether lower-income countries, as *producers* of research, would be disadvantaged by no longer being able to afford to publish in journals [11]. Our study, however, is situated in the realm of potential benefits of OA from the *consumer* perspective (i.e. in the use of knowledge). From this perspective, it is reasonable to foresee that the greatest use and benefit will be in countries initially characterised by poor access to scientific publications. OA was introduced with the aim of remedying such problems, that is, making scientific literature (and, increasingly, research data) available to everyone. However, to the best of our knowledge, studies are yet to explore the specific benefits of increased accessibility to scientific literature, beyond the fact that it has become freely available for reading.

Several recent studies have explored country and regional differences in OA publishing. In terms of production, lower-income countries have been found to have the highest share of OA publishing [5]. Regarding consumption, they have the highest share of OA citations [7]. The latter finding is particularly interesting. While having access to reading a paper is inherently valuable, assessing the impact of such access requires observable and measurable behavioural changes.

Our study is grounded in the *consumer* perspective. In addition to presenting descriptive findings regarding OA reference patterns, we analyse how and the extent to which OA is beneficial to lower-income countries. In this study, we explore whether the potential increase in the use of OA publications and references in lower-income countries relative to high-income countries covaries with other indicators exhibiting evidence of better research capacity or represents a greater inclination to be updated in the frontiers of science. The main research question in our study is whether increased use of OA publications is also followed by:

(a) Improved access to scientific literature, as expressed in longer reference lists and a higher share of scientific papers being cited in these lists.
(b) Improved access to more recent scientific literature, as expressed in more recent and updated publications cited in reference lists.

2. Background: global differences in research and development (R&D) capacity

The *raison d'être* of the expectation that OA is more beneficial to research systems in lower-income countries has to do with the generally low levels of R&D investments in these countries' (e.g. OECD or World Bank statistics on R&D investments per capita), with several studies documenting very high shares of OA publishing in these countries [5,7]. Access to scientific literature has been limited in several developing countries, especially in Africa [12,13]. Low R&D investments negatively impact the ability to cover the cost of (very expensive) journal subscriptions [14–16]. Although several initiatives have been introduced to remedy this situation [13] – for example, initiatives from the World Health Organization and the United Nations aimed at helping developing countries obtain increased access to more (and more recent) scientific information, thereby providing lagging and developing countries greatly subsidised access to scientific literature [17] – they have been characterised by inherent coordination problems and a lack of consistency in defining eligible beneficiary countries [13]. Moreover, Nobes and Harris [18] emphasised that 'complex systems of authorization portals and systems is a matter of contention for many and makes accessibility more difficult'. Another concern about these initiatives is sustainability. What happens to accessibility once the donorship ends [15,18]? Thus, developing countries' accessibility to scientific knowledge is not only affected by OA but also by donor-funded and negotiated access schemes [19]. Nevertheless, we see OA as the main driver of increased accessibility to science in recent years; thus, we focus on it as a potential driver of changes in reference patterns from lower-income countries.

It is common in OA studies to distinguish between different OA types (e.g. gold, hybrid, green, etc.; see Simard et al. [7] for an overview). In this article, we analyse OA as a unified concept, as the model under which a paper is openly available is irrelevant from the perspective of literature *use*. Thus, it is beyond the scope of this study to provide a comprehensive overview of the literature on OA publishing and OA models, as we are primarily interested in how OA might benefit researchers in their publishing process.

2.1. Added value of our study

OA as a question of global equity has been studied since the emergence of OA as a publication model. Among recent studies, Simard et al. [7] combined an analysis of OA publishing and citation trends (2015–2019) with country data and found that low-income countries had a higher uptake of OA in both trends than high-income countries.[1] Robinson-Garcia et al. [20] found somewhat contrasting findings in their study on the percentages of OA publishing for institutions in the Leiden ranking, where those with a high percentage of OA publishing, with the exception of India and Iran, were all

upper middle- or high-income countries. Here, we observe a noticeable difference between Western and Asian countries. All institutions with OA shares above the world median were European.

Simard et al. [7] examined references to OA literature based on studies assigned to the country of the first and/or last author, with full counts for each country, while Robinson-Garcia et al. [20] used institutional data (aggregated author shares of institutions). Arguably, however, it is important to distinguish between studies in which lower-income countries engage in collaboration with higher-income countries (where we may assume that the scientific literature is more widely available in the authorship groups) and those without such collaboration (where low-income countries cannot rely on infrastructure support from higher-income countries). For example, in a study in which researchers from Burkina Faso collaborated with French and British researchers, we cannot take for granted that a listed reference was also available to the authors from Burkina Faso. Therefore, we limit our study to publications with authors from *one country* at a time, making it more reasonable to claim that the references listed in this article were indeed accessed and then included in the publication by authors from that country, thereby better reflecting their R&D capacity. This means that in comparison with the broad OA literature, we move from the production to the consumption side of OA, which we consider a novel contribution to the OA literature. Therefore, we shed light on how OA may have helped lower-income countries in their ability to make use of scientific literature, thereby going beyond a focus on simply having better access to more literature. We do so by elaborating on the findings of Simard et al. [7] in two ways: (1) specifying three additional dimensions related to reference practices: the average number of references per publication, the share of references indexed in WoS and the age of the cited literature and (2) restricting publications to one country's income level at a time, thus avoiding potential bias from publications where cited references may be assigned to low-income countries when they were the result of access to literature in high-income countries. Below, we describe the three OA 'consumer' indicators analysed in this article.

2.2. Average number of references in reference lists

The first indicator explored here is the change in the *average number of references* in the selected papers. As a result of the gradual change towards OA publishing and the launch of new subscription-free 'gold' journals and repositories for archiving scientific literature, doors have opened to resources and publications and, thus, citable literature, that were previously inaccessible. To the best of our knowledge, no studies have explored country differences for this indicator, which is surprising given how many angles have been used in 'citation studies', for example, studies about how the number of citations in a reference list varies by paper length, research areas/discipline, the number of citations in review papers, etc. (examples taken from Nicolaisen and Frandsen's [21] review of this literature).

The rationale for studying the number of references is based on the logic that accessibility to scientific literature is a determinant of how many relevant papers an author/group of authors can meaningfully use in a paper. In high-income countries, access was unimpeded prior to OA. The reverse ensued in lower-income countries (especially African countries), where there has historically been poor access to Internet and computer facilities, preventing researchers from fully utilising increasingly available online literature [13,17]. Arguably, therefore, with the increasing number of OA studies available, the number of references per paper may have increased more in lower-income countries relative to higher-income countries.

2.3. Higher share of WoS-indexed references

There were country differences in terms of the reference 'style' used, for example, the tendency to cite national papers, which was more pronounced in smaller R&D nations [22]. Basson et al. [23] showed that researchers in developing countries tended to include more non-WoS-indexed OA articles in their reference lists compared with those in wealthier countries. This finding was interpreted as potentially resulting from the inclusion of many smaller national journals in the Dimensions database used. Although there were many reasonable ways to justify publishing in non-WoS-indexed journals (for language issues, potential dissemination issues, etc.), we can also speculate that the use of non-indexed national journals or other types of reference sources (reports, online documents, etc.) has been more prevalent in lower-income countries due to a lack of access to WoS-indexed journals. However, it has been shown globally (at least in the field of ecology) that there is a trend towards fewer citations of non-journal publications [24]. Considering the increasing presence of OA papers, we believe that it is important to study whether there has been an increase in the share of WoS-indexed references from lower-income countries and whether this growth has been more prominent than in higher-income countries, thereby 'closing the gap'.

Didegah, Thelwall and Gazni examined low-income countries' share of references from various journal percentiles (2000–2009) and found that,

> in proportion to their total number of citations given to WoS journals, it seems that less developed countries cite high-quality journals at the same rate as developed countries and so the poorer publishing of less developed countries does not seem to be due to a lack of access to top journals. [25, p. 516]

However, this finding was based on co-publications showing that more than 80% of top-journal *publications* from low-income countries were in collaboration with high-income countries. Again, we argue that it is only when an analysis is restricted to publications with no international collaboration that a genuine country effect can be examined.

2.4. Citations of more recent publications

When the United Nations Food and Agriculture Organization launched its 'Global Online Research in Agriculture' (AGORA) in 2003, eligible developing countries were offered selected journals on CD-ROM [13]. In Adcock and Fottrell's survey in Tanzania [17], some of their respondents mentioned that journal collections classified as either 'new' or 'most recent' could be literature as old as 5 years. Although this has arguably improved since the beginning of the 2000s, it is still relevant to ask whether low-income or developing countries are more exposed to the so-called obsolescence literature [26]. Globally, there is a tendency towards citing more recent publications [27], but does this trend also apply to lower-income countries and can the increased uptake of OA also contribute to more recent research being cited in publications from these countries? We find it reasonable to assume that while OA has enabled easier access to the latest research worldwide, the lagging of developing countries makes it logical to assume that the increase in more recent papers being cited will be much more pronounced there. Therefore, we examine whether lower-income countries have had a relatively stronger reduction than high-income countries in the average number of years of their references since the date of publication.

2.5. The objective of this study

The objective of this study is to explore whether citation behaviour among researchers in lower-income countries has changed throughout the period of OA to become more similar to that of researchers in economically wealthier countries – that is, that the performance indicators of lower-income countries related to use of scientific literature has improved relative to those of wealthier countries.

3. Methods

Developments in OA publishing and reference patterns across regions and country income levels have been carefully analysed in recent studies [7,23]. However, we believe that it is important to continue this line of research by asking what OA growth means from a *benefit* perspective. Does OA really contribute to improved research conditions in countries characterised by lagging R&D? To answer this question, we used a stepwise strategy involving four research questions:

- First, we explored whether researchers from lower-income countries refer to fewer publications when they publish and, if so, whether the gap narrows over time.
- Second, we explored whether researchers from lower-income countries rely more on literature that is openly available through different OA routes than other researchers.

To answer these questions, we analysed OA trends using total and field-specific numbers. In order to study references, especially over time, it is important to take the scientific field dimension into account. Apart from the arts and humanities field, there has historically been a steady increase in the number of references in all fields of science, as illustrated by Nicolaisen and Frandsen for the period 1996–2019 (in a study of 27 million papers in the WoS from seven major scientific domains) [21]. Moreover, the share of OA papers in reference lists is highly field-dependent, with the highest share being in medical and health sciences [7,28]. As there are large differences in the specialisation profile of countries [29], national and regional aggregate numbers may be misleading if the scientific field is not accounted for. Therefore, in our analyses, we also addressed the field dimension in our calculations of the OA numbers.

The second part of the analysis took the benefit perspective into account. Here, we focused on the reference practices of researchers, placing the question of OA literature citations in the analysis of changes in reference practices over a long period of time. This allowed us to better understand how OA references fit into broader developments around reference

practices and provided us with a baseline of comparison for understanding the magnitude of the uptake of OA in the literature. Specifically, we examined the following:

Third, whether lower-income countries experienced relatively stronger growth in the share of WoS-indexed references relative to high-income countries.
Fourth, whether lower-income countries witnessed a relatively stronger reduction in the average age of references than high-income countries.

To put these research questions in context, we included some initial analyses of general developments in reference patterns over time.

The analysis relied on simple descriptive statistics. It followed the development of indicators measuring the *use* of OA literature over a 15-year period (2005–2020) and focused on intuitively observable changes and patterns rather than attempts at drawing statistically significant causal mechanisms. All analyses were carried out using the R programming language [30], specifically the dplyr package for data manipulation [31] and the ggplot2 package for the production of figures [32].

The study is based on data from the WoS, a scholarly publication database that comprises the world's leading scholarly journals, books and proceedings in the sciences, social sciences and arts and humanities. The WoS contains full bibliographical metadata, including the time of publication, authorship, affiliations and the reference lists of its indexed publications. This enables the analysis of reference trends between countries and across time. The WoS Core Collection, which we used in this study, now covers the Science Citation Index Expanded, the Social Sciences Citation Index, the Arts & Humanities Citation Index, the Emerging Sources Citation Index, the Conference Proceedings Citation Indexes and the Book Citation Index. We applied a local version of the WoS maintained by the Norwegian Agency for Shared Services in Education and Research.

The analyses were global in nature. As transnational collaborations make it impossible to identify the source country of a given reference in a publication, the study was limited to regular articles with a single author country and involved a total of 32.7 million articles during the period 1980–2020 and 1.29 billion entries in the reference lists of these articles. The corresponding global total, which included internationally co-authored articles, was 42.0 million. Thus, the analysis covered the majority (78%) of WoS-indexed articles published during the period. This is in accordance with a recent study showing that in 2020, 25% of the world production of articles involved international collaboration [33]. We classified author countries into four income groups (i.e. high, upper middle, lower middle and low income) based on the World Bank's classification of countries using the 2020 gross national income per capita [34].

The number of papers included in the analysis increased from approximately 380,000 in 1980 to 1,702,000 in 2020. The number of papers from the high-income group was by far the highest, with 24,020,000 articles over the entire period. The corresponding figures for the other groups were as follows: 6,554,000 articles for upper middle income, 2,094,000 for lower-middle income and 2941 for low income. Clearly, there were remarkably few publications from low-income countries. This is because these countries have limited national budgets for research activities, and most of the research is carried out in collaboration with researchers from other countries, resulting in internationally co-authored publications [33], which were not included in this study.

Data on the OA status of references were retrieved from Unpaywall, a global, openly licenced publication database identifying the reading access status of around 50 million scientific publications [3]. Unpaywall operates with several aspects of OA. As we were concerned only with the possibility of reading the contents of a publication for use as a reference in a publication and not the access licence indicated by the different types of OA, we considered all forms of OA as one category for the purposes of this study.

Having collected all WoS publications that fit the above-outlined criteria (i.e. published 1980–2020, regular articles, single-country authorship and containing a DOI), the publications were matched against the Unpaywall database to obtain their OA status. The matching was done using a strict string comparison of the DOIs of the publications, a field featuring in both databases. In cases where no match for a reference was found, its status was set to 'not applicable'. The distribution of the access status of the references of a paper containing non-identifiable references was assumed to be identical to that of references whose OA status could be determined. The analyses were carried out at the field level. We used a classification system consisting of 16 broad categories (for further details, see [35]). Noteworthy, the ambiguous field classification 'multidisciplinary' was not used in the analysis.

This approach has some limitations. First, not all publications included in the various reference lists could be assigned an OA status. This was due to several factors: (1) not all references in a publication were from scientific publications; (2) the WoS only indexes a subset of all scientific publications; (3) not all references have a DOI, making matching with Unpaywall impossible and (4) not all scientific publications have an OA status in Unpaywall. Moreover, the domination

of English as a research language in the WoS database should be noted. Many non-Anglophone countries also represent scientific publication spheres among themselves, and for low- or middle-income countries, there are many OA venues for research dissemination that are not indexed in a database, such as the WoS. Therefore, this study only encompasses what was probably the most internationally oriented research from these countries, which may have different reference profiles from research published in more nationally or regionally oriented channels.

In order to mitigate potential biases in the results, two additional key measures were included: first, analysing trends across distinct scientific fields and, second, limiting the analysis to papers without authors from several country income-level groups. While this did significantly reduce the overall number of publications included, we viewed this constraint as a strength, as it allowed us to isolate and scrutinise the impact of country income levels on reference practices.

4. Results

The analysis is divided into five parts. First, we look at the changes in the length of the reference lists over time, focusing overall on the global level. Second, we analyse the differences in reference patterns among the country income groups. Third, we analyse the OA patterns, with the last two sections addressing the changes in the WoS-indexed literature and the age of the cited references over time.

4.1. Overall reference patterns over time

Figure 1 shows that the number of references per article increased considerably during the period 1980–2020. In 1980, an average article contained 19.8 references, compared with 43.2 in 2020.

This general upwards trend was observed in all fields, as shown in Figure 2, which plots the mean number of references per publication and by field for the period 1980–2020. At the same time, there were large differences across fields. For example, business studies and biology had much longer reference lists on average than mathematics and statistics. In 2020, the average mathematics publication included more references than the average biology publication in 1980.

4.2. Differences across country income groups

To what extent are there differences in the length of the reference lists by country income group? Figure 3 shows that articles from high-income countries had more references on average than those from other income groups. This held over the entire 40-year period, but the gap has narrowed in recent years. Interestingly, there were relatively small differences across the groups of low- and middle-income countries. For high-income countries, the average number of references was 21.4 in 1980 and 43.7 in 2020. For upper middle-, lower middle- and low-income countries, the figures varied from 15.1 to 15.7 in 1980 and 36.7 to 38.4 in 2020. This pattern of small differences in the groups of low- and middle-income countries also held for other variables analysed in this article. In such cases, we present figures for the three groups combined in the analyses below.

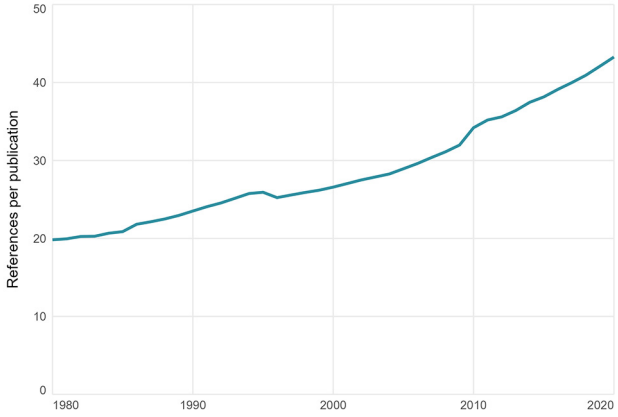

Figure 1. Average number of references per publication, 1980–2020.

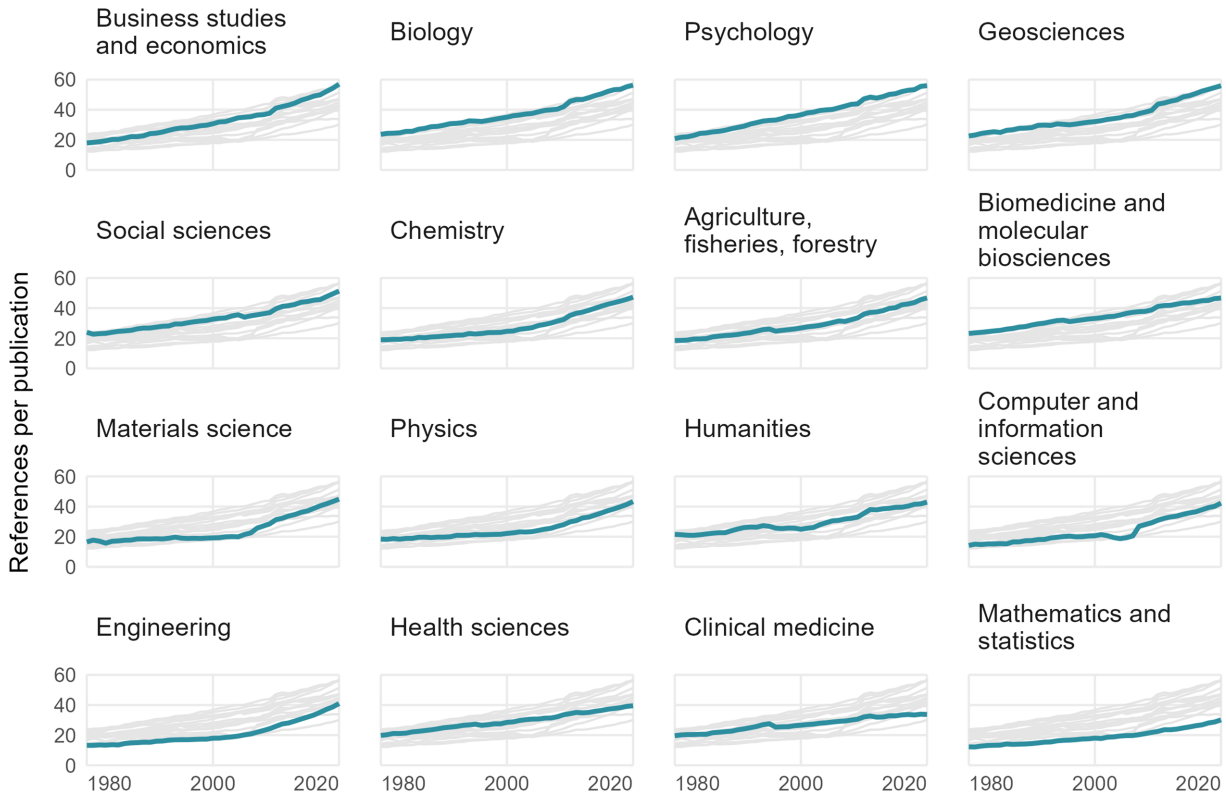

Figure 2. Average number of references per publication by field, 1980–2020.

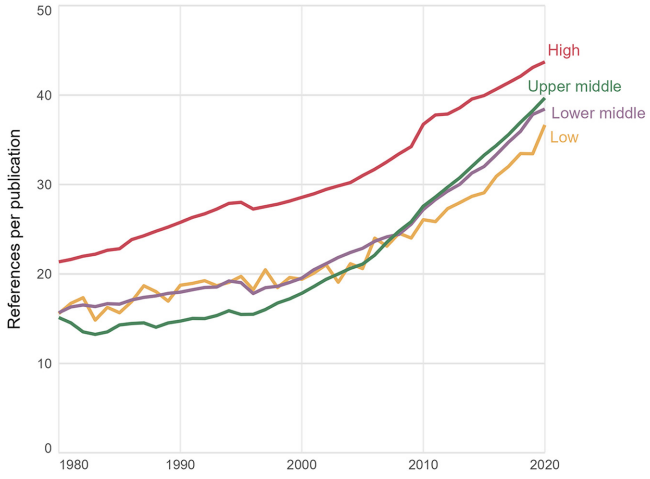

Figure 3. Average number of references per publication by country income group, 1980–2020.

Combining the insights from Figures 2 and 3, we examine whether the mean references per publication varied by field and income group, as shown in Figure 4.

The general trends from the previous figures held across all fields and income groups: the number of references per paper increased steadily in all fields, and there was a gap between the high-income countries and the middle- and low-income countries. However, there were field-specific differences: in engineering and physics, the difference in the average number of references was much less pronounced than in biomedicine and biology.

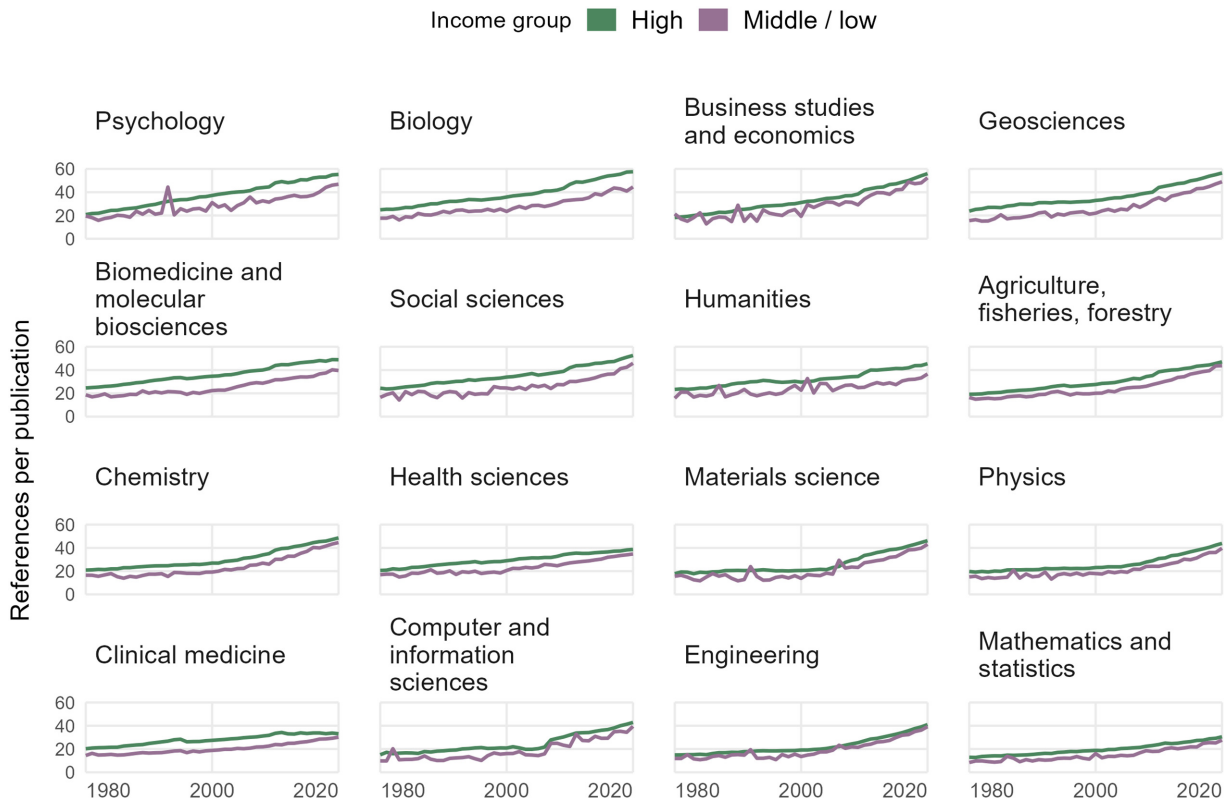

Figure 4. Average number of references per publication by field and country income group, 1980–2020.

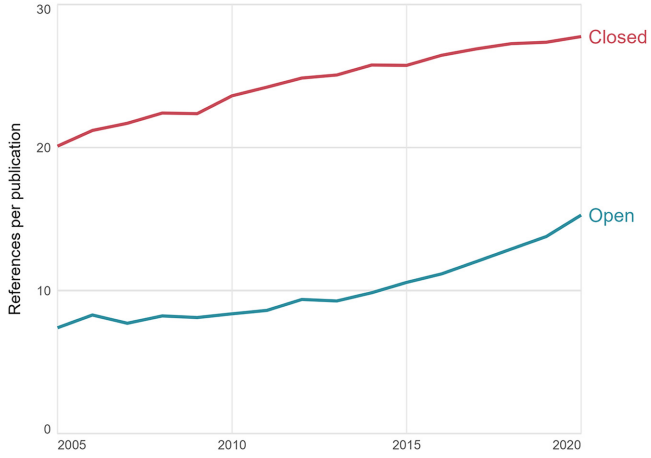

Figure 5. Average number of references per publication by access type, 2005–2020.

4.3. Changes in OA and non-OA reference patterns over time

In the period under consideration, another trend occurred concurrently with the increase in the number of references in papers – the growth of openly available scientific publications. Although virtually unknown before the late 2000s, OA to publications has grown to capture a substantial share of global publications in recent years. Figure 5 shows the number of references per paper based on the OA status of the referenced publications for the period 2005–2020. In 2005, papers cited on average 7.4 OA publications and 20.1 non-OA (closed access) publications, while in 2020, these figures were 15.3 and 27.8, respectively.

Table 1. Average number of references per publication by country income group and access type, 2005–2020.

	High-income		Middle/Low-income	
	Non-OA	OA	Non-OA	OA
2005	24.2	7.5	18.6	7.4
2006	24.7	7.7	20.0	8.5
2007	25.5	7.9	20.4	7.6
2008	26.5	8.2	21.0	8.2
2009	27.0	8.7	20.8	7.9
2010	28.6	9.3	22.0	8.1
2011	29.3	9.7	22.5	8.2
2012	29.5	10.0	23.3	9.2
2013	29.7	10.4	23.5	8.9
2014	30.2	10.9	24.3	9.5
2015	30.2	11.4	24.3	10.3
2016	30.2	12.4	25.2	10.7
2017	30.5	13.3	25.7	11.6
2018	30.4	14.1	26.2	12.5
2019	30.5	15.4	26.3	13.2
2020	30.3	16.5	26.9	14.9

Table 2. Compound average growth rate by income group and access type, 2005–2020.

Income group	Non-OA	OA
Low-income	35%	42%
Lower middle income	28%	33%
Upper middle income	26%	32%
High-income	13%	18%

The corresponding figures for the high- and middle-/low-income groups are shown in Table 1. Overall, the number of OA and non-OA references increased for both income groups. In terms of OA publications, we observed an increase in growth around 2015. In 2005, publications from high-income countries had an average of 7.5 OA references and 24.2 non-OA references, increasing to 16.5 OA and 30.3 non-OA references in 2020. Thus, the growth was stronger for OA than non-OA references in absolute (+ 9 references/ + 6 references) and relative terms (+ 55%/ + 25%). A similar pattern was shown for middle- and low-income countries. The number of OA references per publication increased from 7.4 to 14.9 during the period 2005–2020, while the number of non-OA references increased from 18.6 to 26.9 in the same period, corresponding to a relative growth of 101% for OA and 45% for non-OA. Thus, for both types of publications, the relative increase was stronger than that observed for high-income countries.

The trend towards referencing more OA publications was also evident when calculating the proportion of references going towards OA publications. For high-income countries, an average paper in 2005 referenced 23.7% OA publications. In 2020, this proportion was 35.3%. For middle- and low-income countries, the corresponding proportions were 28.5% and 35.6%, respectively.

The above analysis of the relative share of OA references is based on an analysis of the compound average growth rate (CAGR) of the references in absolute terms for the four income groups for both closed access and OA references during the study period. Table 2 demonstrates that the CAGR was positive for all references, as shown in Figure 3, and that it was higher for OA references for all income groups and highest for the two lowest income groups.

The growth of OA references, particularly among publications from lower-income countries, indicates a convergence of reference practices among researchers from different income groups. Calculations were also performed at the field level (not shown in this article) and revealed that the overall OA patterns identified for the country groups were also observed at the field level. However, in some fields, the patterns did not follow the general tendencies regarding the country groups, thereby making the conclusions less generic.

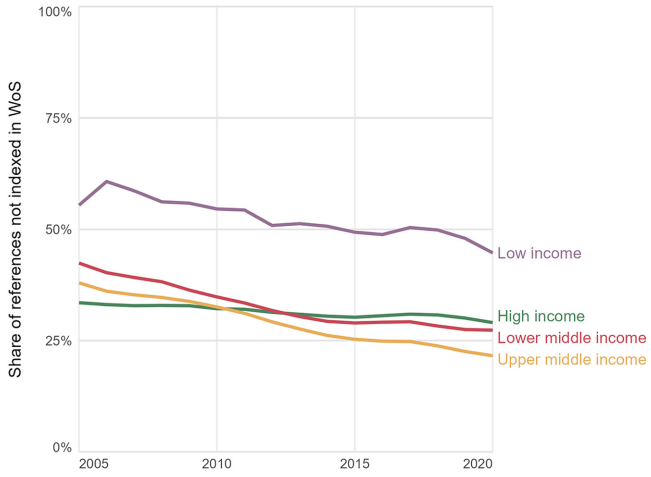

Figure 6. Average proportion of references not indexed in the WoS by country income group.

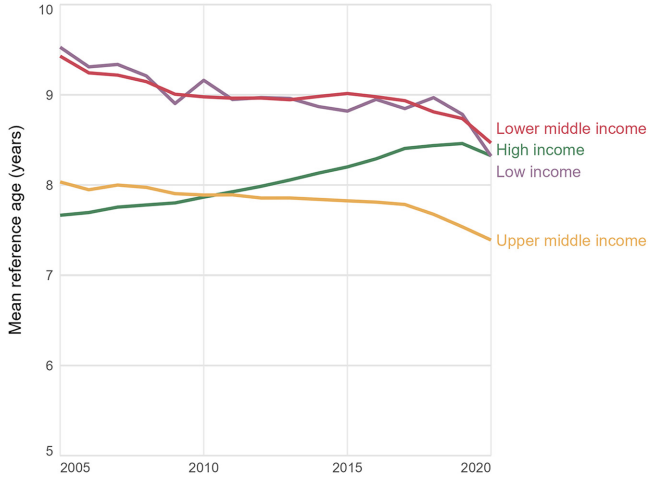

Figure 7. Average age of references per publication by country income group, 2005–2020.

4.4. Proportion of WoS-indexed references

As described in the introduction, we investigated whether low-income countries had stronger relative growth in their share of WoS-indexed references compared with high-income countries. This analysis was carried out for the 2005–2020 period, and the results are shown in Figure 6. We found that low-income country researchers referenced non-indexed publications more often than their counterparts in other country groups. This held over the entire period, but the proportion decreased from 60% in 2006 to 45% in 2020. For high-income countries, the pattern was quite stable, with a marginal decline from 33% in 2005 to 29% in 2020. For the middle-income country groups, there was a marked decline (15 percentage points) during the period. This means that articles are referencing indexed publications more often than in the past. In fact, these groups have surpassed high-income countries during the period and are now referencing indexed publications at a higher rate than those in high-income countries.

4.5. Age of references

Finally, we analysed whether there were differences across the income groups in terms of the average age of the references. This analysis was carried out for publications from the 2005–2020 period using a citation time window limited to 25 years prior (older references are rare) (Figure 7). Thus, for publications from 2005, we included references to works only as far back as 1980 because there are no data on the OA status of references prior to 1980. In 2005, there were

marked differences between publications from high-income countries and those from low-income and lower middle-income countries. Publications from high-income countries cited articles that were, on average, 2 years younger. This suggests that in the past, researchers in lower-income countries were unable to take advantage of new research to the same degree as they are today. However, the gap almost disappeared during the 15-year period, mainly due to the reduced age of the references in papers from the two lowest income groups (from 9.5 to 8.3 years and 9.4 to 8.5 years) and an increase in the reference age of papers from high-income countries (from 7.7 to 8.3 years). The pattern for upper middle countries not only differed but also showed a declining trend. Further analysis revealed that China, which contributed to most of the publications in this group, deviated from the more recent references than other countries in this group as well as the high-income group. Omitting China from the data set resulted in a declining trend for the upper middle countries, with the line being positioned somewhere between high-income countries and low- and lower middle-income countries.

5. Discussion

This study has shown that the evolution of OA has been accompanied by changing citation behaviour, with the main finding being that the reference practices of researchers from lower-income countries have become more similar to those of researchers from more wealthy countries due to a relatively stronger increase in lower-income countries with regard to the number of references. Furthermore, this has been accompanied by an increase in the share of references from WoS-indexed scientific journals, in addition to citing increasingly more recent (i.e. 'younger') literature. In sum, these findings support the notion that OA has played a positive role in lower-income countries' access to scientific literature. Our study does not contradict the little empirical evidence on the way in which OA benefits lower-income countries; it adds important additional information in understanding what changes in scientific behaviour have become possible with the introduction of more openly available literature.

Our findings show a steady growth in the number of references to scientific publications in the last 40 years. Overall, the number of references per article has doubled from 20 in 1980 to more than 40 in 2020, representing a major change in the structure of scientific communication and how new research is embedded in the previous literature. This also has a huge effect on the inverted citation patterns and frequencies of individual publications [24]. Previous research has suggested that the growth is linked to the transformation of scientific communication, such as the availability of online search tools [36] and more easily accessible literature through electronic publication and communication [24,27].

In addition, our study shows that reference patterns differed significantly across fields and country groups. We investigated variations in reference practices according to the economic status of countries, limiting the study to papers without cross-country collaboration in order to isolate the effect of each country's R&D capacity and access to scientific literature. Researchers from low-income countries tend to refer to fewer publications when they write papers compared with those from high-income countries. While this gap persisted over the entire 40-year period analysed, it is narrower today than in the past. In 2020, an article from high-income countries included 44 references, on average, compared with 37 references for articles from low-income countries.

We observed that in the period coinciding with the rise of the OA movement, researchers in lower-income countries have narrowed the gap in terms of the length of their reference lists. Although we did not provide causal evidence of the effect of OA on reference practices, we believe that this development can be partly explained by the larger increase in references to OA publications from lower-income countries relative to high-income countries. Lower-income countries have a much stronger CAGR rate than high-income countries for both closed access and OA literature, but the increase is clearly highest for OA publications.

Moreover, the reference patterns of low-income countries have changed in two additional dimensions. First, researchers from these countries refer more to WoS-indexed literature than they did in the past; second, they refer to more recent literature. Both patterns may be linked to the rise of OA, although further investigations are required to verify this.

Our study shows that the reference patterns of high- and low-income countries differed in general and with respect to OA. Nevertheless, we observed that even today, most of the references by researchers in low-income countries are attributed to closed access literature. One may ask why this is the case and why the OA differences across country groups are not larger. One important factor may be the lack of institutional OA mandates and policies to encourage uptake [37]. In addition, researchers from lower-income countries face similar situations as those from high-income countries when they collect literature, that is, only a few parts of the relevant literature on a particular topic are openly available [38]. Thus, based on such considerations, one cannot expect radically different patterns, although it is argued that both the production and use of OA in some developing countries are dominated by predatory publishers [37], which boosts OA activity in such regions, or they benefit from waiver programmes [37]. Moreover, what appears to be closed access literature in this study may still be openly available. It should be noted that articles available through academic social networks only

(e.g. ResearchGate and Academia.edu) were not counted as OA due to the criteria applied by Unpaywall [3]. Full-text articles from a very large body of literature are available through these websites [39], and some studies point to geographic differences in the use of such sources [40]. For example, Meier and Tunger [41] found that ResearchGate was more actively used in India, Italy and Spain compared with Canada, the Netherlands and the United States, while Wan [42] found that the information pirate website Sci-Hub was more frequently used in South American countries. Another option is to contact the authors directly to get a printed version of articles or PDFs sent by e-mail. Finally, according to some studies [43–45], authors do not always read the full paper they cite. The extent to which such citing practices affect our results is not possible to assess. This points to an interesting issue unresolved in our study: citations of papers based purely on the reading of abstracts [45]. Abstracts are also available in the WoS for closed access literature. There is a possibility that some of the observed increase relates to the expanded coverage of the WoS over time [46], which has increased accessibility to abstracts.

In most of the variables analysed in this study, we did not observe large differences across the groups of low- and middle-income countries; however, the reference patterns of the high-income countries diverged. Therefore, we conclude that researchers from the first group of countries behave quite similarly when it comes to reference practices. This is somewhat surprising considering that middle-income countries have better economies, more developed science systems and assumedly better access to scientific literature. One possible mechanism is that in low-income countries with very low rates of one-country-only papers, authors may represent a very select group of researchers who may still have access to relevant literature due to past and other ongoing international collaborations. It is important to note that low 'productivity' here refers to the world of the WoS and does not consider that outside of North America and Europe, much scientific work is published in channels not indexed in the WoS [23]. Furthermore, in European countries, we may also find a large share of publications from (national language) journals not indexed in WoS [46].

Relatedly, it should be mentioned that some countries moved from one income category to another during the study period. For example, China appeared as 'an upper middle-income country' for the first time in 2012. In this study, the changing economic status of countries was not adjusted, and the current status was superimposed onto previous years. In the case of China, as it has emerged as a large contributor to the world's publication system, it would seem disruptive to remove it from one category to another. Although applying the historical status of countries seemed justified, this would not change the main conclusions of this article due to the fact that the figures for the low- and middle-income countries were similar in several dimensions, and transpositions to the highest group mainly occurred for smaller research jurisdictions [47].

As mentioned in the introduction, many studies have examined the role of OA from a global equity perspective [48]. However, most studies have been limited to the production side, with research on the consumption side remaining limited (i.e. the use of literature, as reflected in reference patterns). One exception is the study of Simard et al. [7], which showed that low-income countries have a higher uptake of OA than high-income countries. Despite similarities to our study, Simard et al. also differed fundamentally in terms of methodological approach. They assigned papers to country groups based on the address affiliation of the first and corresponding authors only. Thus, publications with co-authors from additional countries were also included. This may have affected the findings regarding low-income countries, as their scientific production is limited, and international collaboration is widespread [33]. In the introduction, we explained why such a methodological approach may be problematic. Therefore, our study was limited to articles that did not involve international collaboration. Nevertheless, the overall findings of our study resemble those of Simard et al., as the methodological disparities did not result in contradictory main conclusions.

While Simard et al. found that OA citations were more frequent among researchers in lower-income countries in the period 2015–2019 (combined), we showed that throughout a longer period (2005–2020), the growth rate in the OA citations by researchers in lower-income countries was more than twice the increase among researchers in high-income countries. Although the number of OA citations was still higher in high-income countries in our final year (2020: on average, 16.5 OA citations per paper) compared with lower- and middle-income countries (on average, 14.9 OA citations per paper), the ratio between closed access and OA citations was almost identical: both cite 80 per cent more non-OA papers than OA papers. The difference is that non-OA citations were more frequently subscription-based journal articles in high-income countries and non-indexed (in the WoS) publications in lower-income countries.

In this study, we analysed additional dimensions of reference patterns to provide further evidence of the broader role of OA. One issue was to use the age of references as a parameter to ascertain whether low-income countries were more exposed to the so-called obsolete literature. We found that in the past, publications from lower-income countries did refer more to older literature compared with publications from high-income countries. However, this gap has narrowed in recent years. This is partly a consequence of the increasing age of the references of high-income countries during the study period, potentially reflecting the fact that the average age of all available scientific literature around the world increases with time.

There was also a notable difference in that publications from low-income countries more frequently referenced non-WoS-indexed publications than those from other country groups. This indicates that researchers from these countries rely more on peripheral literature than core international scientific literature. This may be explained by the more restricted availability of the latter literature (perhaps also the greater use of local language publications) and deviant research profiles focused on local and regional issues, which has, for example, been demonstrated in health and medical research [49].

The proportion of WoS-indexed literature has also been decreasing over time, but so are the corresponding figures for the other country groups. This may be partly explained by the expansion of the WoS database over time, with the addition of new citation indexes (e.g. the Proceedings Citation Index), which means that more of the scientific literature is covered by the database [35].

We hope that this analysis will spur further research on the extent to which OA has benefitted research and researchers in countries with limited financial resources. For example, it would be interesting to assess how the patterns are influenced by specific support models that offer free or low-cost access to academic literature, such as the Research4Life programme. Literature categorised as closed access in Unpaywall might still be accessible to researchers in lower-income nations via these models. Similarly, the impact of closed access literature that remains accessible through providers and platforms such as ResearchGate (researchgate.net) was unaccounted for in our analysis and could be investigated.

This article focused on single-country-authored publications. Another fruitful research avenue could be to analyse this publication set against one containing co-authorships between lower- and higher-income countries to ascertain whether collaborating with researchers in higher-income countries incurs the same changes in reference lists as the existence of OA literature.

6. Conclusion

OA has significantly changed the global publishing landscape. The potential benefits and risks of this transformation have been extensively discussed in the scientific literature and other forums. Despite the prevalence of OA in the contemporary literature, a considerable portion of historical publications remains subscription-based. Our investigation has analysed changes in researchers' reference practices during the evolution of OA (the last 20 years). The backdrop is that there has been a significant increase in the number of references in scientific publications in the last 40 years, with a traditionally strong difference between lower-income and higher-income countries in terms of the number of references. Part of this difference may be attributed to country and field differences in citation practices, but we believe that the main explanation is the difference in access to scientific literature, that is, research that can be used for reference purposes. Our study points to OA as a successful tool in reducing the traditional cleft between regions based on economic wealth in terms of (1) number of references, (2) the age of references and (3) the share of references from WoS-indexed scientific papers.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Henrik Karlstrøm https://orcid.org/0000-0002-9942-4326

Data availability

The data analysed in this study were retrieved from the Web of Science (WoS) database, licenced by the Norwegian Agency for Shared Services in Education and Research from Clarivate. The licencing agreement permits distribution of publication metadata from the WoS in aggregated form. The data required to reproduce the analysis in this paper are available from a data repository at the following URL: https://doi.org/10.5281/zenodo.10954539

Note

1. An interesting observation made by Simard et al. [7] is the fact that most publishers provide publication fee waivers for researchers from low-income countries but only partially for middle-income countries, which may explain why OA is used more frequently in low-income countries relative to middle-income countries.

References

[1] Oppenheim C. Electronic scholarly publishing and open access. *J Inf Sci* 2008; 34(4): 577–590.
[2] Zhang L, Wei Y, Huang Y, et al. Should open access lead to closed research? The trends towards paying to perform research. *Scientometrics* 2022; 127: 7653–7679.
[3] Piwowar H, Priem J, Lariviere V, et al. The state of OA: a large-scale analysis of the prevalence and impact of open access articles. *PeerJ* 2018; 6: e4375.
[4] Norris M, Oppenheim C and Rowland F. The citation advantage of open-access articles. *J Am Soc Inf Sci Technol* 2008; 59(12): 1963–1972.
[5] Iyandemye J and Thomas MP. Low-income countries have the highest percentages of open access publication: a systematic computational analysis of the biomedical literature. *PLoS ONE* 2019; 14(7): e0220229.
[6] Evans JA and Reimer J. Open access and global participation in science. *Science* 2009; 323(5917): 1025.
[7] Simard M-A, Ghiasi G, Mongeon P, et al. National differences in dissemination and use of open access literature. *PLoS ONE* 2022; 17(8): e0272730.
[8] Heidbach K, Knaus J, Laut I, et al. Long term global trends in open access. *A data paper*, 2022, https://hdl.handle.net/21.11116/0000-0009-C475-6
[9] Demeter M, Jele A and Major ZB. The international development of open access publishing: a comparative empirical analysis over seven world regions and nine academic disciplines. *Publ Res Q* 2021; 37: 364–383.
[10] Budzinski O, Grebel T, Wolling J, et al. Drivers of article processing charges in open access. *Scientometrics* 2020; 124: 2185–2206.
[11] Kwon D. Open-access publishing fees deter researchers in the global south. *Nature*, https://www.nature.com/articles/d41586-022-00342-w (2022, accessed 13 February 2024).
[12] Abrahams L, Burke M, Gray E, et al. *Opening access to knowledge in Southern African universities (Study series 2008)*. Cape Town, South Africa: Southern African Regional Universities Association (SARUA), 2008.
[13] Chan L and Costa S. Participation in the global knowledge commons. Challenges and opportunities for research dissemination in developing countries. *New Libr World* 2005; 106(3–4): 141–163.
[14] Aronson B. Improving online access to medical information to low-income countries. *N Engl J Med* 2003; 350(10): 966–968.
[15] Dulle FW and Minishi-Majanja MM. Researchers' perspectives on open access scholarly communication in Tanzanian public universities. *S Afr J Inf Manag* 2009; 11(4): a413.
[16] Matheka DM, Nderitu J, Mutonga D, et al. Open access: academic publishing and its implications for knowledge equity in Kenya. *Glob Health* 2014; 10(1): 1–5.
[17] Adcock J and Fottrell E. The North-South information highway: case studies of publication access among health researchers in resource-poor countries. *Glob Health Action*. Epub ahead of print 13 November 2008. DOI: 10.3402/gha.v1i0.1865.
[18] Nobes A and Harris S. Open access in low-and middle-income countries: attitudes and experiences of researchers. *Emerald Open Res* 2019; 1: 17.
[19] Malapela T. Access to scholarly research information in sub-Saharan Africa: a review. *Libri* 2017; 67(1): 1–3.
[20] Robinson-Garcia N, Costas R and van Leeuwen TN. Open access uptake by universities worldwide. *PeerJ* 2020; 8: e9410.
[21] Nicolaisen J and Frandsen TF. Number of references: a large-scale study of interval ratios. *Scientometrics* 2021; 126: 259–285.
[22] Barrantes BSL, Bote VPG, Rodriguez ZC, et al. Citation flows in the zones of influence of scientific collaborations. *J Am Soc Inf Sci Technol* 2012; 63(3): 481–489.
[23] Basson I, Simard M-A, Quangre ZA, et al. The effect of data sources on the measurement of open access: a comparison of dimensions and the Web of Science. *PLoS ONE* 2022; 17(3): e0265545.
[24] Mammola S, Fontaneto D, Martinez A, et al. Impact of the reference list features on the number of citations. *Scientometrics* 2021; 126: 785–799.
[25] Didegah F, Thelwall M and Gazni A. An international comparison of journal publishing and citing behaviours. *Journal of Informetrics* 2012; 6: 516–531.
[26] Dorta-Gonzalez P and Gomez-Deniz E. Modeling the obsolescence of research literature in disciplinary journals through the age of their cited references. *Scientometrics* 2022; 127: 2901–2931.
[27] Evans JA. Electronic publication and the narrowing of science and scholarship. *Science* 2008; 321(5887): 395.
[28] Maddi A. Measuring open access publications: a novel normalized open access indicator. *Scientometrics* 2020; 124: 379–398.
[29] Aksnes DW, van Leeuwen TN and Sivertsen G. The effect of booming countries on changes in the relative specialization index (RSI) on country level. *Scientometrics* 2014; 101: 1391–1401.
[30] R Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2023, https://www.R-project.org/
[31] Wickham H, Francois R, Henry L, et al. dplyr: a grammar of data manipulation_ (R package version 1.1.4), 2023, https://CRAN.R-project.org/package=dplyr
[32] Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag, 2016.
[33] Aksnes DW and Sivertsen G. Global trends in international research collaboration, 1980–2021. *J Data Inf Sci* 2023; 8(2): 26–42.
[34] World Bank. How does the World Bank classify countries? https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries (2022, accessed 13 February 2024).

[35] Piro FN (ed.). Comparing research at Nordic higher education institutions using bibliometric indicators: covering the years 1999–2014. *Nordic Council of Ministers, NordForsk*, 2017, http://norden.diva-portal.org/smash/record.jsf?pid=diva2%3A1115783&dswid=-2712
[36] Gingras Y, Lariviere V and Archambault E. Literature citations in the internet era. *Science* 2009; 323(5910): 36.
[37] Sheikh A and Richardson J. Open access movement in the scholarly world: pathways for libraries in developing countries. *J Inf Sci*. Epub ahead of print 27 October 2023. DOI: 10.1177/01655515231202758.
[38] Laakso M, Welling P, Bukvova H, et al. The development of open access Journal publishing from 1993 to 2009. *PLoS ONE* 2011; 6(6): e20961.
[39] Taylor J. The shady margins of open access: gray OA and academic social network sites. *Tech Serv Q* 2022; 39(4): 428–439.
[40] Singh VK, Piryani R and Srichandan SS. The case of significant variations in gold–green and black open access: evidence from Indian research output. *Scientometrics* 2020; 124(1): 515–531.
[41] Meier A and Tunger D. Survey on opinions and usage patterns for the ResearchGate platform. *PLoS ONE* 2018; 13(10): e0204945.
[42] Wan S. Which nationals use Sci-Hub mostly? *Ser Libr* 2022; 83(3–4): 228–232.
[43] Ball P. Paper trail reveals references go unread by citing authors. *Nature* 2022; 420(6916): 594–594.
[44] Simkin MV and Roychowdhury VP. A mathematical theory of citing. *J Am Soc Inf Sci Technol* 2007; 58(11): 1661–1673.
[45] Nicholas D, Huntington P and Jamali HR. The use, users, and role of abstracts in the digital scholarly environment. *J Acad Librariansh* 2007; 33(4): 446–453.
[46] Aksnes DW and Sivertsen G. A criteria-based assessment of the coverage of Scopus and Web of Science. *J Data Inf Sci* 2019; 4(1): 1–21.
[47] World Bank. World Bank country and lending groups, https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (2022, accessed 13 February 2024).
[48] Chan L, Kirsop B, Costa SMDS and Arunachalam S. Improving access to research literature in developing countries: challenges and opportunities provided by open access, https://archive.ifla.org/IV/ifla71/papers/150e-Chan.pdf (2005, accessed 13 February 2024).
[49] Yegros-Yegros A, van de Klippe W, Abad-Garcia MF, et al. Exploring why global health needs are unmet by research efforts: the potential influences of geography, industry and publication incentives. *Health Res Policy Syst* 2020; 18: 47.