



A comparison of systematic reviews and guideline-based systematic reviews in medical studies

Alexander Schniedermann¹ 

Received: 2 July 2020 / Accepted: 27 October 2021 / Published online: 19 November 2021
© The Author(s) 2021

Abstract

The question of how citation impact relates to academic quality accompanies every decade in bibliometric research. Although experts have employed more complex conceptions of research quality for responsible evaluation, detailed analyses of how impact relates to dimensions such as methodological rigor are lacking. But the increasing number of formal guidelines for biomedical research offer not only the potential to understand the social dynamics of standardization, but also their relations to scientific rewards. By using data from Web of Science and PubMed, this study focuses on systematic reviews from biomedicine and compares this genre with those systematic reviews that applied the PRISMA reporting standard. Besides providing an overview about growth and location, it was found that the latter, more standardized type of systematic review accumulates more citations. It is argued that instead of reinforcing the traditional conception that higher impact represents higher quality, highly prolific authors could be more inclined to develop and apply new standards than more average researchers. In addition, research evaluation would benefit from a more nuanced conception of scientific output which respects the intellectual role of various document types.

Keywords Systematic reviews · Reporting guidelines · PRISMA · Standardization · Document types

JEL Classification D83

Mathematics Subject Classification 62P25

Introduction

In contemporary science policies, much emphasis is put on the evaluation of the producers of biomedical knowledge. The most recent trend in science evaluation, the incorporation of seemingly objective tools to measure publication output or citation impact gave rise to the idea that the former represents productivity, while the latter indicates quality,

✉ Alexander Schniedermann
schniedermann@dzhw.eu

¹ German Centre for Higher Education Research and Science Studies, Berlin, Germany

professionalism or excellence in biomedical research, (de Rijcke et al., 2016; Jappe, 2020; Petersohn et al., 2020). Biomedical researchers and other scientists have reacted to such incentives, not only by dividing their work into more publishable units, but also by incorporating such evaluative categories deeply into their epistemic practices (Müller & de Rijcke, 2017). As a result, a growing number of publications, often with lacking novelty or little to add forced leading experts to proclaim a crisis in biomedical research. Beside the increasing efforts for practitioners and researchers to cope with the growing stock of information, experts criticized the decreasing quality of biomedical research, so that findings are less reliable, less credible or just “false” (Ioannidis, 2005, 696).

Ironically, especially the biomedical research community pursued a century of improving the credibility of its research outputs. Facing the outputs of the “clinical trial industry” due to the 1960s FDA regulation (Meldrum, 2000, 755), experts demanded more systematic assessments of medical knowledge in order to improve the health and life of patients (Chalmers et al., 2002). As a result, medical disciplines increasingly employed meta-analyses and systematic reviews in order to combine multiple studies into an overall and reliable result that can be considered as ‘evidence’ for applicatory contexts and the development of medical treatment guidelines (McKibbon, 1998; Moreira, 2005). In addition, systematic reviews became not only a more common genre in academic periodicals, but also the main focus of multinational organizations that followed the ideal of evidence-based practice, such as the Cochrane Collaboration or the Campbell Collaboration (Simons & Schnieder-mann, 2021).

Method experts constantly improve the recipes for systematic reviews in order to cope with newly discovered varieties of biases and preserve the epistemic credibility of this genre. For example, in a more recent trend, demands for a more transparent reporting of reviews were uttered and manifested in new standards, most notably the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (Moher et al., 2009). Because PRISMA has been implemented in many editorial policies and is highly cited by systematic reviews, its developers ongoingly monitor the actual compliance to further improve the guideline. For example, in their recent evaluation of the PRISMA, Matthew Page and David Moher listed 57 different studies that attempt to analyze to which extent systematic reviews comply with the guideline’s rules (Page & Moher, 2017). They conclude that, although there is room for improvements, the guideline already positively affected the reporting quality of systematic reviews and meta-analyses.

Systematic reviews and their PRISMA-improved counterparts are often settled on top of the hierarchy of biomedical evidence (Goldenberg, 2009). As such, this genre is situated at the intersection of what is perceived as high levels of epistemic quality, and what is actually measured by the wide palette of bibliometric indicators. For example, because review articles generally accumulate higher citation counts than reports of primary research—an observation long known to bibliometricians—they are suspected to be used strategically by some academic journals and individual researchers (Blümel & Schnieder-mann, 2020). Some bibliometric assessments found that systematic reviews and meta-analyses inherit this characteristic and are also used strategically by impact-sensitive actors (Barrios et al., 2013; Colebunders et al., 2014; Ioannidis, 2016; Patsopoulos et al., 2005; Royle et al., 2013). But allegations of strategic behavior in academia are hard to prove solely on the basis of bibliometric measurements. Nevertheless, bibliometric assessments can point towards potential misuses and inform more profound investigations (Krell, 2014; Wyatt et al., 2017).

Several studies elaborated the bibliographic characteristics of systematic reviews, often with a focus on specific domains. Usually, medical researchers and method developers

investigate the “epidemiology of [...] systematic reviews” by focusing on characteristics such language, inclusion criteria or whether a protocol was pre-registered or published (Page et al., 2016). Other observations are the high growth rates and publication counts of systematic reviews and resulting problems (Ioannidis, 2016; Page et al., 2016; Bastian, 2010). Other analyses focus on national affiliations, number of authors, funding sources or citation impact. For example, a study by Alabousi et al. found a positive correlation between the Journal Impact Factor and systematic reviews published in the field of medical imaging (Alabousi et al., 2019).

Studies that take methodological standards such as PRISMA into account often focus on specific sub-fields, or use bibliometric indicators that are inappropriate for the assessment of individual articles. For example, one study found a positive correlation between the quality of reporting of systematic reviews and absolute citation rates in radiology (Pol et al., 2015). A study by Nascimento et al. found a positive correlation between the endorsement of PRISMA and the Journal Impact Factor for systematic reviews about low back pain (Nascimento et al., 2020). Similarly, Mackinnon et al. found a weak correlation between reporting quality and 5-Year Journal Impact Factor in dementia biomarker studies (Mackinnon et al., 2018). Molléri et al. conclude that the rigor of software engineering studies is related to normalized citations, but argue that, at the same time, this may come at the costs of relevance (Molléri et al., 2018).

In order to provide a broader overview about the characteristics and dissemination of systematic reviews in biomedicine, this study focuses especially on the role and impact of (reporting) standards such as PRISMA on systematic reviewing. Although standardization is a common phenomenon in medical research and practice, the willingness and speed in applying new standards differs by medical subfields (cf. Timmermans & Epstein, 2010). For this reason, this study wants to highlight the fields, nations and topics that usually deal with systematic reviews and shed some light on their dynamics in effectively implementing PRISMA. By using different methods to identify systematic reviews and separating them from those that use the PRISMA guideline, this study extends existing research by understanding PRISMA-based systematic reviews as standardized and highly appraised forms of scientific output. Therefore, by analyzing impact measures, it is not only attempted to provide scores irrespective of sub-disciplines and publication years, but also contrast them with standardized and more rigorous forms of systematic reviews.

Methodology

Corpus construction

To compare the bibliometric characteristics of systematic reviews and PRISMA-based systematic reviews, corpora were constructed by a combination of bibliographic data from PubMed and Web of Science. Document types in Web of Science lack in accuracy and necessary detail, especially in the case of review articles (Donner, 2017). In contrast, PubMed contains of several document types related to secondary research and introduced the

“Systematic Review-type in 2019, because of this genres’ importance for medical decision making (NLM, 2018).¹

In a first step, publication types and MeSH keywords were retrieved for 31 million PubMed items via the Entrez-API during March and April 2020 and stored in a PostgreSQL database. In contrast to Web of Science, each of these records is linked to 1.73 different document types on average, up to 10 assignments. The reason for this is a much richer classification system in PubMed consisting of 80 different types which also incorporate methodological information such as ‘Randomized Controlled Trial’ and funding information such as ‘Research Support, Non-U.S. Gov’t’ (Knecht & Marcetich, 2005).

After matching the retrieved records with the inhouse version of Web of Science at the German Kompetenzzentrum Bibliometrie and restricting the publication years from 1991 to 2016 in order to account for changing indexing policies at MEDLINE (NLM, 2017), the resulting set contained 10 million records.

In comparison with the Web of Science classification consisting of “Article”, “Review”, “Editorial”, “Letter”, and “News”, precision and recall were calculated (Baeza-Yates and Ribeiro-Neto, 2011). While sufficiently high for articles (precision=0.94/recall=0.95), they are rather low for reviews (0.85/0.52) which means that while 85% of items labeled “Review” in Web of Science carry the same type in PubMed (true positive), 52% of items are labeled “Review” PubMed, but not in Web of Science (false negative). These results confirm earlier analyses (Donner, 2017), and show the huge differences between document type classification systems in Web of Science and PubMed (see also Harzing, 2013).

To build the guideline-based systematic review set, all items that cite PRISMA or one of its descendants have been separated. PRISMA documents have been identified via DOI and title search and relevant DOI’s have been downloaded from the PRISMA website (www.prisma-statement.org), as well as the EQUATOR network website (www.equator-network.org).

Statistical analysis

For the comparison of systematic reviews and PRISMA-based systematic reviews, several variables have been calculated. For the annual and compound growth rates, basic publication outputs were used. All variations other than document types are based on meta-data from the Web of Science database. As such, field variations are based on extended subject categories. National variations are based on the fractional assignment of items to the authors’ affiliations, taking all affiliations into account. In addition, the country list has been restricted to those that published at least 500 reviews, or have 10% systematic reviews or 2% PRISMA-based systematic reviews over the whole timespan. In addition, countries that belong to the Commonwealth have been separated (cf. Encyclopaedia Britannica, 2021), due to their stronger commitment to evidence-based medicine (Groneberg et al., 2019).

For the impact assessment, all publications in the set have been compared by the major document types. To employ a 3-year citation window, citation analysis is restricted to the publication years between 2009 and 2015. For comparison, the mean citation impact of each group was calculated in terms of absolute citations, mean normalized citation scores

¹ Note that the official term in PubMed is „publication type “ rather than document type. In Web of Science, “publication type” refers to the place of publication, e.g. periodic journal, book etc.

(MNCS) and mean cumulative percentile ranks (CPIN). Both synthetic indicators, MNCS and CPIN are based on normalizations by Web of Science’s subject categories.

In contrast to traditional mean normalized citation scores, percentile measures are more robust against high impact outliers which skew citation distributions of all fields and are a common issue in bibliometric data (Bornmann, 2013). But at the same time, they reduce the data’s level of measurement to an ordinal ranking scale which prohibits any conclusion about the level of differences. However, this does not prohibit conclusions about mean differences of the document types.

Generally, percentile ranks reflect to which citation impact group a publication belongs. Groups can be delineated at wish, for example, a common set is the top 10% percentile, meaning that these papers are cited more often than 90% of rest within this field and publication year (Rousseau, 2012). But citation scores are discrete values which complicate pre-defined rank classes because the percentages are based on the amount of publications while the borders are based on the citation values. For example, if ten publications are cited (0,0,0,0,1,2,2,3,3,3) times, the top ten percent of these would amount one third of the whole set. In order to preserve pre-defined percentile ranking sets, fractional assignments have been introduced which, in turn, give up the binary nature of the ranks (Waltman and Schreiber 2012).

For an optimal percentile measure that can be further aggregated and is also interpretable, Lutz Bornmann and Richard Williams suggested excluding cumulative percentiles, CPEX (2) and including cumulative percentiles CPIN (1) (Bornmann & Williams, 2020). These are calculated for each combination of classification and year of publication. For these sets all possible citation scores are based on a three-year citation window (I) with $j, k \in I$, the amount of papers with j citations (c_j) and the total amount of papers (t).

$$CPIN_j = \sum_{k=0}^j \frac{c_k}{t} * 100 \tag{1}$$

$$CPEX_j = CPIN_j - \frac{c_j}{t} * 100 \tag{2}$$

After CPIN or CPEX have been calculated, each focal paper can be assigned such values based on field classification, publication year, citation window. Although both represent cumulative percentages, their interpretation is different. CPEX represents the amount of other publications that are cited less than the focal one within the field and year of publication. CPEX always has zero as its lowest value because no publications are cited less than zero times. CPIN represents the amount of other publications that are cited equal or less than the focal one. CPIN always has 100 as its highest value because no publication is cited more than the most cited ones.

Results and discussion

The growth and dissemination of systematic reviews

As displayed in Table 1, the analyzed dataset consists of 9.9 million total items. While most of the items are primary articles (8.1 million), substantial parts of scientific

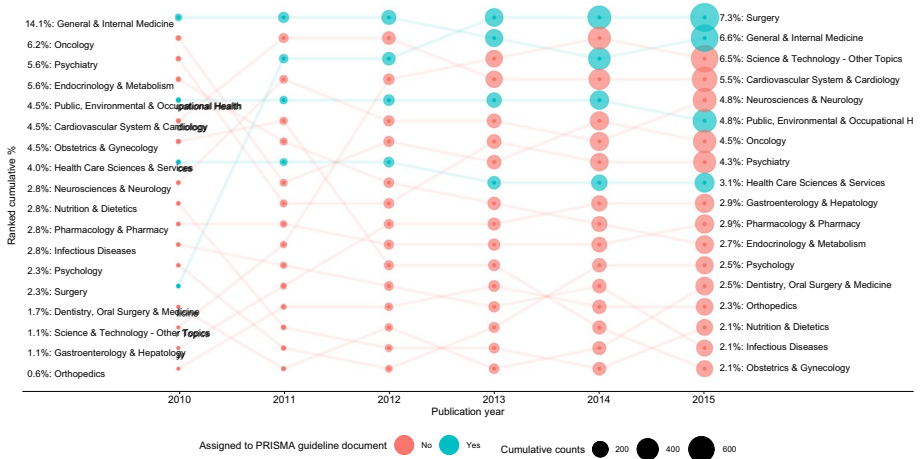


Fig. 2 Ranked cumulative proportions of field assignments and cumulative counts to PRISMA-based systematic reviews from 2010 to 2015. Restricted to fields that have at least a 2% rate of overall assignments

reviews and PRISMA-based systematic reviews have increasing outputs. While systematic reviews increased annually by over a fifth every year before 2009 (22%) and over 14% since then, PRISMA-based systematic reviews almost doubled (82%) every year since the publication of the guideline, leading to an overall growth of 2675% until 2016. In addition, the overall growth of systematic reviews that do not comply with PRISMA halved since its inception (from 300 to 148%).

Systematic reviews and PRISMA-based systematic reviews play important intellectual roles for various medical fields and communities. Figure 1 shows Web of Science’s subject categories that are commonly assigned to systematic reviews or PRISMA-based systematic reviews. “Surgery”, with rather moderate rates of systematic reviews (18%) and PRISMA-based reviews (2.4%) accounts for the highest total number of PRISMA-based reviews (756 assignments).

Whenever “Business & Economics” was assigned to a publication in PubMed, it was a systematic review in 42% of the cases which is the top value in the dataset. High rates are also found in “Information Science & Library Science” (38%) and “Medical Informatics” (37%). On the other site, of all assignments of “Chemistry” only 0.5% go to systematic reviews which is the lowest value. Others are “Biophysics” (1.1%) or “Cell Biology” (1.5%).

While “Evolutionary Biology” has a rather low rate of systematic reviews (7.6%), it even was never assigned to PRISMA-based systematic reviews, representing the lowest value in this category. On the other side, “Surgery”, a field in which the guideline was published, tops this variable with 756 assignments and a rate of 2.4% among all research syntheses, topping even broader categories like “General & Internal Medicine” (688 assignments, 27% rate) or “Science & Technology—Other Topics” (681 assignments, 23% rate). Not surprisingly, fields in which the guidelines have been published offer higher rates in both types of assignments.

How the field assignments to PRISMA-based systematic reviews changed over time is shown in Fig. 2. While fields such as “surgery” raised their share from 2.3% in 2010 to 7.3% in 2015, fields like “Obstetrics & Gynecology” decreased from 4.5% of all assignments

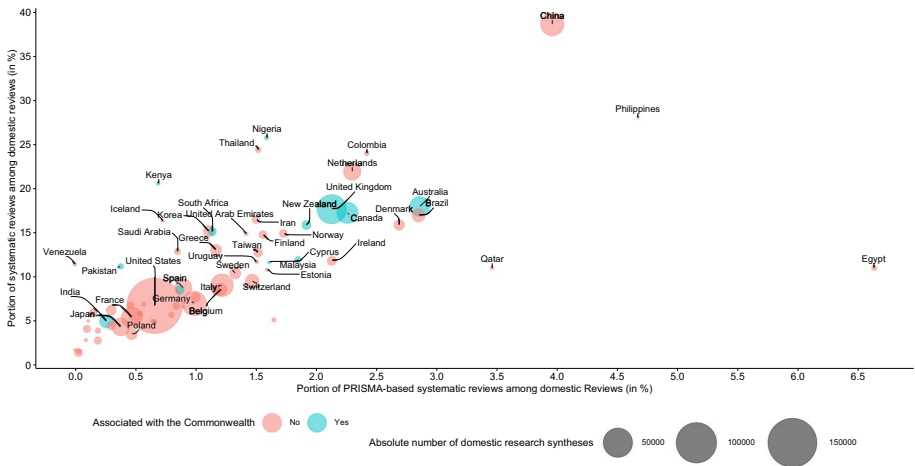


Fig. 3 Overall number of research syntheses, portions of systematic reviews and PRISMA- based systematic reviews by country, and national association with the Commonwealth. Based on fractional assignment of author’s affiliations. Labels are restricted to countries that published at least 500 reviews, have at least 10% systematic reviews or 2% for PRISMA-based reviews accordingly

in 2010 to 2.1% in 2015. Losing its top position in 2010, “General & Internal Medicine” decreased from 14 to 6.6% in 2015 but still marks second rank. Although the ranks change over the years, Fig. 3 shows that the differences in portions and counts between the fields are much smaller compared to what can be observed for systematic reviews in general.

The data visualizes some popular narratives about the dissemination and occurrence of systematic reviews and their standardized counterparts. Generally, the differences show that although the guideline developers attempted to make PRISMA applicable for a broad range of fields and academic cultures, the actual levels of implementation and compliance differ substantially. Accepting a new standard means that established practices have to change. Although standardization and the change in standards is a rather regular phenomenon in biomedicine (Whitley, 2000), such attempts are sometimes accompanied by criticism and reluctance (Solomon, 2015; Timmermans & Berg, 2003; cf. Hunt, 1999). For example, communities with a strong track of qualitative research turned out to be rather critical of systematic reviewing (Cohen & Crabtree, 2008).

One reason for strong objections is the more procedural and prescriptive way of how systematic reviews or meta-analyses create medical evidence (Stegenga, 2011). Beside the objection that this employs a rather narrow perspective on scientific methodology and neglects the plurality in medical research (Berkwits, 1998), a proper execution of systematic review methods requires substantial skills in information retrieval, data cleaning or statistics (Moreira, 2007). Not surprisingly, the data shows that when disciplines such as “Information Science & Library Science” or “Medical Informatics” have been assigned to secondary articles, it was often a systematic review or PRISMA-based systematic review.

Beside objections against systematic reviews in some fields, this genre has grown a substantial footprint in fields that are committed to evidence-based medicine. For example, assignments of “Public, Environmental & Occupational Health” and “Health Care Sciences & Services” regularly go to systematic reviews (30% and 33% respectively) or PRISMA-based systematic reviews (3.5% and 4.2% respectively) with accumulating a total of 499 assignments and 323 assignments respectively, to this genre. These values correspond to

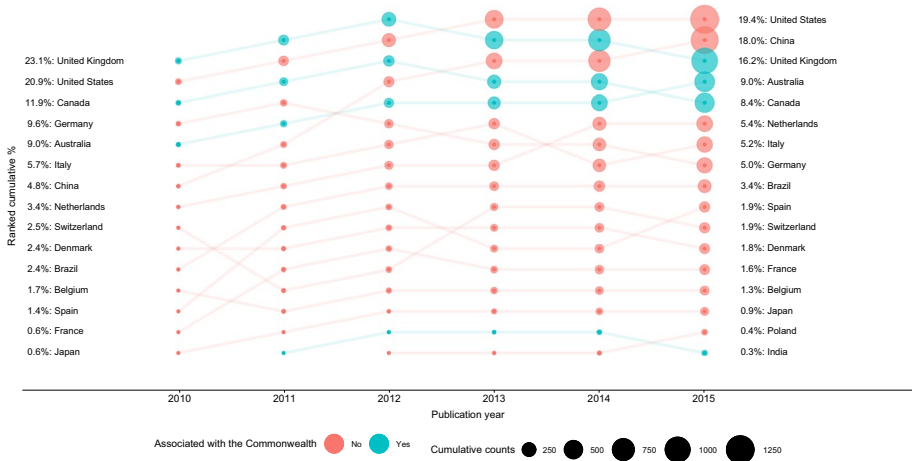


Fig. 4 Ranked cumulative proportions and cumulative counts of country assignments to PRISMA-based systematic reviews from 2010 to 2015

the idea that systematic reviews are important for medical decision making, the development of clinical guidelines or their overall role for public health assessments—contexts in which systematic reviews proved to be fruitful in the past (Moreira, 2005; Hunt, 1999). Similarly, systematic reviews are suggested to reach final conclusions on certain topics by aggregating all available evidence. While this function is crucial for disciplines such as surgery, where a multitude of treatments have to be appraised (Maheshwari & Maheshwari, 2012), scientific disputes are not always solvable by this way (Vrieze, 2018).

Systematic reviews by countries

Although the biomedical sciences are an international community, national variations in the production of research synthesis are observable, as Fig. 3 suggests. While top producers like the United States (195,164 items), the United Kingdom (50,033 items), Germany (33,166 items), and China (29,940 items) publish the majority of research syntheses, they differ in their commitment to systematic reviews and PRISMA-based systematic reviews. While in the U.S. and Germany, less than a tenth of published reviews are systematic reviews (6.8% and 7%), this genre accounts for greater portions of reviews in the U.K. (18%), and especially China (39%). All four countries differ in their commitment to PRISMA-based systematic reviews in a similar manner. With U.S. and Germany having the lowest rates (both <1%), higher rates can be found in the U.K. (2.1%) and China (4.0%). Notably, authors from Egypt published only 587 review articles of which only 11% have been systematic reviews but 6.6% PRISMA-based systematic reviews which is the highest rate in the dataset.

The annual ranks of national assignments to PRISMA-based systematic reviews are displayed in Fig. 4. In the first year after the inception of the guideline, leading producers were the United Kingdom (23%), the United States (21%) as well as Canada (12%). While the U.S. took first position in 2013, accounting for 20% of all PRISMA-based systematic reviews in 2015, especially China showed even higher growth rates. While marking

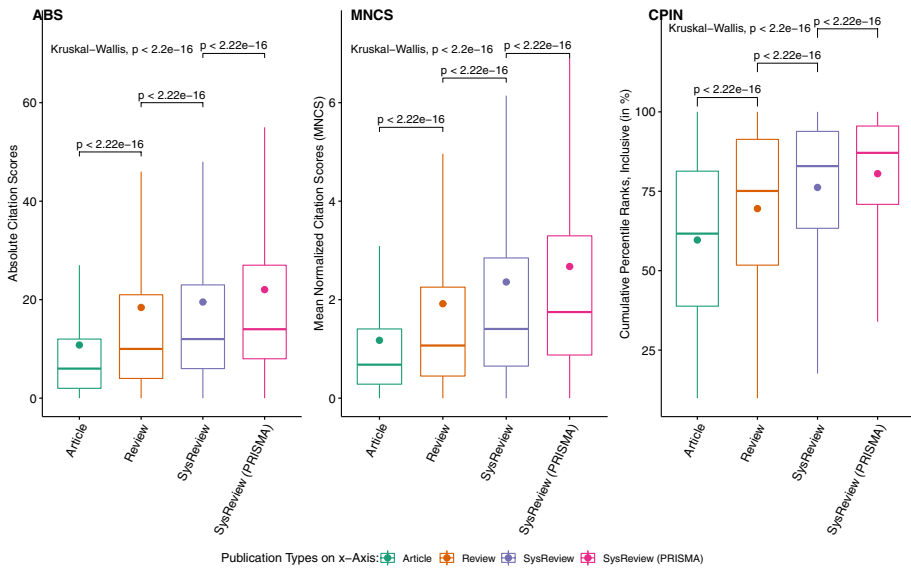


Fig. 5 Box plots with median and mean (dots) values showing citation impact ranks by document type for absolute citations (ABS), mean-normalized citations (MNCS) and cumulative percentile ranks (CPIN), based on publications published from 2009 to 2015 with a 3-year citation window. The y-axes for ABS and MNCS are stripped of extreme outliers for the purpose of better visualization

seventh with 4.8%, its portions grew steadily year over year, taking the second rank in 2015 with over 1000 items and 18% of overall share.

Groneberg et al. have shown that strong commitments towards evidence-based medicine can be associated with the Commonwealth (Groneberg et al., 2019). Similarly, the analysis of document types shows that such countries, especially the U.K., Canada, Australia or New Zealand, produce higher rates of systematic reviews and its PRISMA-based counterpart than other leading science nations (Figs. 3 and 4). Other countries with high rates of these genres, most notably the Netherlands or Denmark, are also among the top ten adopters to evidence-based medicine (ebd.). Although the “evidence-based” movement promotes and comes with a variety of knowledge generation practices, systematic reviewing remains an important cornerstone in EBM as other studies have indicated (Blümel & Schniederermann, 2020; Ojasoo et al., 2001). In that sense, the data provided here corresponds with other studies that mentioned the link between evidence-based paradigms and the emergence of systematic reviews (Timmermans & Berg, 2003; Straßheim and Kettunen 2014).

Besides being fourth in absolute number of reviews, China offers substantially higher rates of systematic reviews (39%) as well as PRISMA-based systematic reviews (4.0%) than all other countries. Although the exact causes may be very complex and unable to identify, two potential causes are proposed. First, being a rising science leader, the majority of Chinese scientists are latecomers to the international community (see Liang et al., 2020). As such, this group may be biased towards a more recent education in the conduct and reporting of biomedical studies. At the same time, Chinese science policy heavily incentivized the production of scientific literature which recently made China the global leader in scientific publishing (Stephen & Stahlschmidt, 2021). Second, due to its recent attempts for healthcare reform, China turned more explicitly towards evidence-based

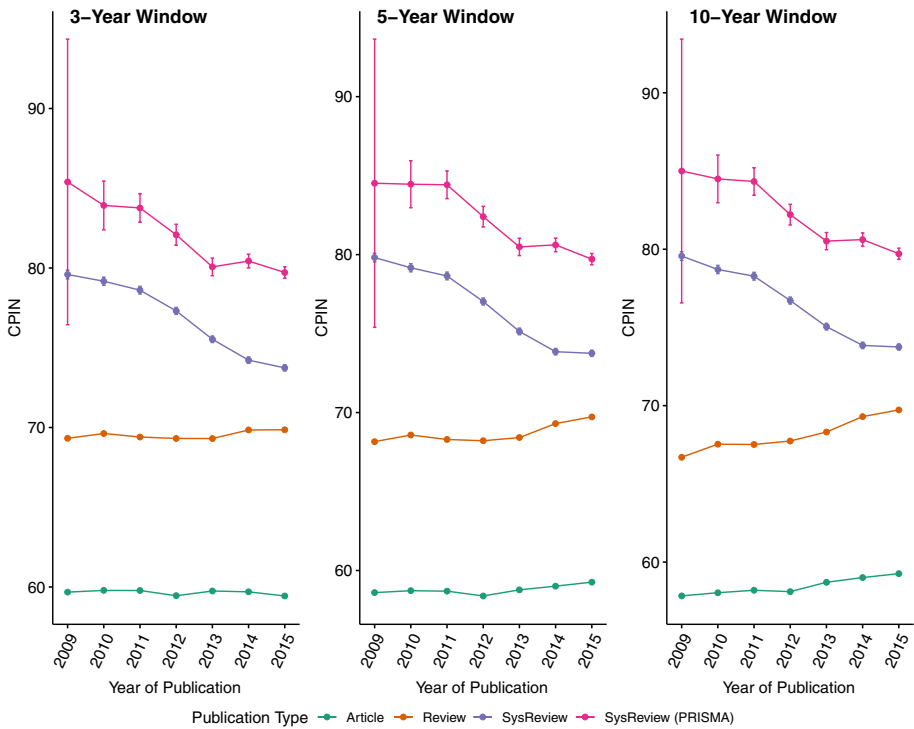


Fig. 6 Average annual cumulative percentile ranks (CPIN) for each document types and corresponding errors, grouped by three different citation windows. As 2018 is the maximum year of citation data, the 5- and 10-year windows are short of data in 2014 and 2009 respectively

medicine and its commitments to systematic clinical research. Together with its growing overall output, China outnumbered even other countries that employ research-based medical systems (Wang & Jin, 2019).

Citation impact of systematic reviews PRISMA

Differences in the citation impact of the analyzed document types are displayed in Fig. 5. All three different citation indicators reveal the same resulting ranks, with the lowest values for articles and highest values for PRISMA-standardized systematic reviews. The latter receive 14 absolute citations (mean = 22), 1.57 normalized citations (mean = 2.67) and the 87% percentile rank (mean = 81%), meaning that 87% of the other document types in the same field receive equal or less citations in three years after publication. In comparison, other systematic reviews receive 12 absolute citations (mean = 19.5), 1.41 normalized citations (mean = 2.36) and reach the 83% percentile rank (mean = 76.2%). Reviews and articles rank lower correspondingly. Mean differences are highly significant with Kruskal–Wallis ($p < 0.001$) for all three sets as well as all pairwise Wilcoxon rank sums ($p < 0.001$).

In Fig. 6, the development of the annual mean CPIN’s shows the dynamics of the impact ranks and corresponding errors for three different citation windows (see Wang, 2013). While 85% of all publications were equally or less cited in a 3-year window

than PRISMA-standardized systematic reviews published in 2009, it decreased to 80% in 2013 and 79.7% in 2015. These developments are similar for the 5-year citation window, coming from 85% in 2009 over 80% in 2013 and 79.9% in 2015. In comparison, systematic reviews came from 80% in 2009 to 74% in 2015 for all three citation windows. Note that since citation data up to 2018 is used, the 10-year window basically represents the whole available citation data. Citation windows seem to play only a minor role. Compared with the 10-year window, the 3-year window bears higher values for articles and reviews, while having slightly lower values for the standardized review formats. Focusing on items published in 2009, the shorter span reveals CPIN differences of +1.84% for articles, +2.62% for reviews, -0.03% for systematic reviews and -0.4% for PRISMA-standardized systematic reviews.

Rankings are the same for all three citation windows. The structurally higher impact of systematic reviews and PRISMA-based systematic reviews is constantly decreasing, since both genres growing faster in publication counts than ordinary articles or reviews. In addition, the results provide greater error values for PRISMA-based systematic reviews since this group has the fewest item counts (6 items in 2009).

The differences in the changes of CPIN values indicates that the document types have different citation patterns, with articles and reviews are cited more consistently in the long run. On a first glimpse, phenomena like delayed recognition (Ke et al., 2015) or the “kiss of death” (Lachance et al., 2014) may influence long-term citation impact of primary research and syntheses differently. However, the different intellectual function of the systematic review, whether standardized or not, in comparison of what scholars now call “narrative review” (Ferrari, 2015) may explain why the former have a shorter citation lifespan. Since systematic reviews are usually designed to achieve consensus on a very particular research question, they draw much more on recent research and become outdated by new findings. Ideally, they are thought to be updated by the same or other authors whenever new trial results or review methods occur (Elliott et al., 2017). This limits their potential citation lifespan, since medical experts are supposed to rely on the most recent findings. In contrast, common reviews are usually of broader scope and may serve a wider array of intellectual functions, such as introductory or educational material, defining the field’s current missions or trends or even render field formation (Grant & Booth, 2009).

The results of the citation analysis provided here show that PRISMA-standardized systematic reviews have a higher citation impact than systematic reviews, reviews or primary articles. In the following, two lines of discussion are proposed.

First, research that complies with reporting standards is credited by higher citation impact afterwards, since it is of higher quality (1). While the conception that ‘citation indicates quality’ still perpetuates in biomedicine and elsewhere (for example, Abramo & D’Angelo, 2011; Durieux & Gevenois, 2010), authors from other domains at least uphold the idea that citations correlate with some specific epistemic values such as rigor or relevance (Molléri et al., 2018). Such a rather traditional conception of impact was common during the early phase of evaluative bibliometrics, but can only found occasionally today. Speaking of a Kuhnian revolution in bibliometrics, Bornmann and Haunschild argue that in contemporary evaluative bibliometrics, citation impact should be considered as only one aspect of a multidimensional concept of research quality, especially there are many different citation behaviors (Bornmann & Haunschild, 2017). For example, authors cite publications if they value the latter’s solidity and plausibility, originality and novelty, scientific value like topical relevance, or societal value, which are complex dimensions that sometimes even conflict each other (Aksnes et al., 2019).

Based on the interpretation that greater quality leads to higher citation impact, there is an important limitation to this study. In the analysis, all systematic reviews that cite the guideline and comply with its first requirement are assigned to the standardized systematic reviews set equally. But meta-studies in biomedicine have shown that the level or quality of guideline compliance varies a lot (Pussegoda et al., 2017). Therefore, to uphold a rather mechanistic relation between quality and citation impact, observed citations in the created set must be normalized by the level of guideline compliance in further research.

Second, beside the interpretation provided above, the results from this study may provide useful input to another interpretation of standards and scientific excellence. So, whatever makes research high impact, it also makes it more open towards standardization and more likely to adopt to new methodological standards (2). Since biomedicine could be understood as what Richard Whitley has called a “professional adhocracy” (Whitley, 2000, 187) standards occur to establish new research practices or tools and further enable formal communication. Authoritative organizations like Cochrane or leading experts in the development of standards form networks in which they negotiate the content and domain of a new standard (Solomon, 2015). By finding consensus, those networks redefine the borders of what counts as a properly reported systematic review which excludes those reviews that do not comply with the standard (for example, Yuan & Hunt, 2009; see also Gieryn, 1999). In this setup, high impact researchers also serve as first movers in adopting the standard. After the leading scientists have defined these standards, other researchers consent afterwards in order to be part of that those that publish ‘the good’ systematic reviews. Similar dynamics have also been observed for academic communities consenting on theories or methods (cf. Luetge, 2004; Zollman, 2007), or the role of reviews in the formation of new disciplines (Blümel, 2020).

In respect to the two different interpretations, the contribution of this analysis to bibliometric research is twofold. Generally, the provided results have revealed some basic characteristics and dynamics of a specific standardization in biomedical research. Together with further research questions about the dissemination and application of PRISMA, for example the characteristics of its developers or adopters, this analysis represents an informative usage of bibliometric data in studying the social aspects of biomedicine or science in general (Gläser & Laudel, 2001; Wyatt et al., 2017).

But identifying standardized research can also contribute to the evaluation of scientific outputs or the development of institutional quality frameworks. As mentioned above, reporting guidelines such as PRISMA have strengthened conceptions about transparency as a fundamental aspect what has to be considered good research. Not surprisingly, reporting guidelines play an important role in the Hong Kong principles for assessing researchers (Moher et al., 2020). Bibliometric data offers the potential to interpret citation in a more diverse form, for example examining the role of method papers, software packages or standards (cf. Li et al., 2017). Similarly, bibliometric analyses of reporting guidelines can enable research evaluations in this respect. In addition, more fine-grained discriminations of document types can ensure that published items are evaluated according to their intellectual contribution.

Conclusion

This study provided a comparative analysis of different document types in biomedical research. In order to understand the relation between methodological quality and citation impact, indexing data from PubMed was used to differentiate between ordinary systematic reviews and those that comply with the PRISMA reporting standard. Besides

providing a general overview about the growth and dissemination of systematic reviews and their standardized counterparts, their citation impact was compared by using different indicators.

The results show that although the growth rates of all biomedical publications decrease, there is still a strong annual growth in systematic reviews and PRISMA-based systematic reviews. Focusing on subject categories, both types of systematic reviewing occur especially in fields that are related to epidemiology and public health. Although the number of assigned subject categories grew, a great portion of PRISMA-based systematic reviews are assigned to fields in which the original guideline was published. While the top producers of scientific literature also dominate the numbers of published systematic reviews, especially countries with a strong focus on evidence-based medicine achieve higher portions of systematic reviews and PRISMA-based reviews. In addition, China is the fourth biggest producer of systematic reviews and also provides substantially higher rates than other leading nations.

Ranking the citation impact of the different document types has revealed that PRISMA-based systematic reviews dominate irrespective of indicator and citation window. It was discussed that this dominance could represent the idea that methodological quality leads to higher citation impact and explain why this conception still perpetuates although it is dismissed in contemporary science studies. In contrast, the results may show that whatever makes authors achieve high citation impact also leads them to willingly apply new methodological standards. Irrespective of which interpretation one favors, bibliometric research could benefit from a more nuanced differentiation of document types in order to evaluate them in respect to their intellectual roles.

Funding Open Access funding enabled and organized by Projekt DEAL. Funded by the German Federal Ministry of Education and Research as part of the program “Quantitative research on the science sector” (01PU17017). Additional support by the German Kompetenzzentrum Bibliometrie (01PQ17001).

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

Data availability All relevant material is provided in the Text.

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

- Abramo, G., & D’Angelo, C. A. (2011). Evaluating research: From informed peer review to bibliometrics. *Scientometrics*, 87(3), 499–514. <https://doi.org/10.1007/s11192-011-0352-7>
- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1), 215824401982957. <https://doi.org/10.1177/2158244019829575>

- Alabousi, M., Alabousi, A., McGrath, T. A., Cobey, K. D., Budhram, B., Frank, R. A., Nguyen, F., Salameh, J. P., Dehmoobad Sharifabadi, A., & McInnes, M. D. F. (2019). Epidemiology of systematic reviews in imaging journals: Evaluation of publication trends and sustainability? *European Radiology*, 29(2), 517–526. <https://doi.org/10.1007/s00330-018-5567-z>
- Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search* (2nd ed.). Addison Wesley.
- Barrios, M., Guilera, G., & Gómez-Benito, J. (2013). Impact and structural features of meta-analytical studies, standard articles and reviews in psychology: Similarities and differences. *Journal of Informetrics*, 7(2), 478–486. <https://doi.org/10.1016/j.joi.2013.01.012>
- Bastian, H., Glasziou, P., & Chalmers, I. (2010). Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? *PLoS Medicine*, 7(9), e1000326. <https://doi.org/10.1371/journal.pmed.1000326>
- Berkwits, M. (1998). From practice to research: The case for criticism in an age of evidence. *Social Science and Medicine*, 47(10), 1539–1545. [https://doi.org/10.1016/S0277-9536\(98\)00232-9](https://doi.org/10.1016/S0277-9536(98)00232-9)
- Blümel, C. (2021). What synthetic biology aims at: Review articles as sites for constructing and narrating an emerging field. In K. Kastenhofer & S. Molyneux-Hodgson (Eds.), *Community and identity in contemporary technosciences*. Springer.
- Blümel, C., & Schniedermann, A. (2020). Studying review articles in scientometrics and beyond: A research agenda. *Scientometrics*, 124(1), 711–728. <https://doi.org/10.1007/s11192-020-03431-7>
- Bornmann, L. (2013). How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top-cited papers. *Journal of the American Society for Information Science and Technology*, 64(3), 587–595. <https://doi.org/10.1002/asi.22792>
- Bornmann, L., & Haunschild, R. (2017). Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact? *Scientometrics*, 110(2), 937–943. <https://doi.org/10.1007/s11192-016-2200-2>
- Bornmann, L., & Williams, R. (2020). An evaluation of percentile measures of citation impact, and a proposal for making them better. *Scientometrics*. <https://doi.org/10.1007/s11192-020-03512-7>
- Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation and the Health Professions*, 25(1), 12–37. <https://doi.org/10.1177/0163278702025001003>
- Cohen, D. J., & Crabtree, B. F. (2008). Evaluative criteria for qualitative research in health care: Controversies and recommendations. *Annals of Family Medicine*, 6(4), 331–339. <https://doi.org/10.1370/afm.818>
- Colebunders, R., Kenyon, C., & Rousseau, R. (2014). Increase in numbers and proportions of review articles in tropical medicine, infectious diseases, and oncology: Increase in numbers and proportions of review articles in tropical medicine, infectious diseases, and oncology. *Journal of the Association for Information Science and Technology*, 65(1), 201–205. <https://doi.org/10.1002/asi.23026>
- Donner, P. (2017). Document type assignment accuracy in the journal citation index data of web of science. *Scientometrics*, 113(1), 219–236. <https://doi.org/10.1007/s11192-017-2483-y>
- Durieux, V., & Gevenois, P. A. (2010). Bibliometric indicators: Quality measurements of scientific publication. *Radiology*, 255(2), 342–351. <https://doi.org/10.1148/radiol.09090626>
- Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., Thomas, J., Agoritsas, T., Hilton, J., Perron, C., Akl, E., Hodder, R., Pestridge, C., & Pearson, L. (2017). Living systematic review: 1. Introduction—The why, what, when, and how. *Journal of Clinical Epidemiology*, 91, 23–30. <https://doi.org/10.1016/j.jclinepi.2017.08.010>
- Encyclopedia Britannica. (2021). *CommonwealthHistory, members, purpose, countries, & facts*. Encyclopedia Britannica. <https://www.britannica.com/topic/Commonwealth-association-of-states>
- Ferrari, R. (2015). Writing narrative style literature reviews. *Medical Writing*, 24(4), 230–235. <https://doi.org/10.1179/2047480615Z.000000000329>
- Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. University of Chicago Press.
- Gläser, J., & Laudel, G. (2001). Integrating scientometric indicators into sociological studies: Methodical and methodological problems. *Scientometrics*, 52(3), 411–434. <https://doi.org/10.1023/A:1014243832084>
- Goldenberg, M. J. (2009). Iconoclast or creed?: Objectivism, pragmatism, and the hierarchy of evidence. *Perspectives in Biology and Medicine*, 52(2), 168–187. <https://doi.org/10.1353/pbm.0.0080>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies: A typology of reviews, Maria J. Grant & Andrew Booth. *Health Information and Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>

- Groneberg, D. A., Rolle, S., Bendels, M. H. K., Klingelhöfer, D., Schöffel, N., Bauer, J., & Brüggmann, D. (2019). A world map of evidence-based medicine: Density equalizing mapping of the Cochrane database of systematic reviews. *PLoS One*, *14*(12), e0226305. <https://doi.org/10.1371/journal.pone.0226305>
- Harzing, A.-W. (2013). Document categories in the ISI web of knowledge: Misunderstanding the social sciences? *Scientometrics*, *94*(1), 23–34. <https://doi.org/10.1007/s11192-012-0738-1>
- Hunt, M. (1999). *How science takes stock: The story of meta-analysis* (1. paperback cover ed). Russell Sage Foundation.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses: Mass production of systematic reviews and meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. <https://doi.org/10.1111/1468-0009.12210>
- Jappe, A. (2020). Professional standards in bibliometric research evaluation? A meta-evaluation of European assessment practice 2005–2019. *PLoS One*, *15*(4), e0231735. <https://doi.org/10.1371/journal.pone.0231735>
- Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences*, *112*(24), 7426–7431. <https://doi.org/10.1073/pnas.1424329112>
- Knecht, L., & Marcetich, J. (2005). New Research Support MeSH® Headings Introduced mid-Year to 2005 MeSH®. *NLM Tech Bull*, *344*(e12). https://www.nlm.nih.gov/pubs/techbull/mj05/mj05_support_heading.html
- Krell, F.-T. (2014). Losing the numbers game: Abundant self-citations put journals at risk for a life without an impact factor. *European Science Editing*, *40*(2), 36–38.
- Lachance, C., Poirier, S., & Larivière, V. (2014). The kiss of death? The effect of being cited in a review on subsequent citations: Journal of the American Society for Information Science and Technology. *Journal of the Association for Information Science and Technology*, *65*(7), 1501–1505. <https://doi.org/10.1002/asi.23166>
- Li, K., Yan, E., & Feng, Y. (2017). How is R cited in research outputs? Structure, impacts, and citation standard. *Journal of Informetrics*, *11*(4), 989–1002. <https://doi.org/10.1016/j.joi.2017.08.003>
- Liang, J., Zhang, Z., Fan, L., Shen, D., Chen, Z., Xu, J., Ge, F., Xin, J., & Lei, J. (2020). A Comparison of the development of medical informatics in China and That in Western Countries from 2008 to 2018: A bibliometric analysis of official journal publications. *Journal of Healthcare Engineering*, *2020*, 1–16. <https://doi.org/10.1155/2020/8822311>
- Luetge, C. (2004). Economics in philosophy of science: A dismal contribution? *Synthese*, *140*(3), 279–305. <https://doi.org/10.1023/B:SYNT.0000031318.21858.dd>
- Mackinnon, S., Drozdowska, B. A., Hamilton, M., Noel-Storr, A. H., McShane, R., & Quinn, T. (2018). Are methodological quality and completeness of reporting associated with citation-based measures of publication impact? A secondary analysis of a systematic review of dementia biomarker studies. *British Medical Journal Open*, *8*(3), e020331. <https://doi.org/10.1136/bmjopen-2017-020331>
- Maheshwari, G., & Maheshwari, N. (2012). Evidence based surgery: How difficult is the implication in routine practice? *Oman Medical Journal*, *27*(1), 72–74. <https://doi.org/10.5001/omj.2012.17>
- McKibbin, K. A. (1998). Evidence-based practice. *Bulletin of the Medical Library Association*, *86*(3), 396–401.
- Meldrum, M. L. (2000). A brief history of the randomized controlled trial. *Hematology/oncology Clinics of North America*, *14*(4), 745–760. [https://doi.org/10.1016/S0889-8588\(05\)70309-9](https://doi.org/10.1016/S0889-8588(05)70309-9)
- Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., Barbour, V., Coriat, A.-M., Foeger, N., & Dornagl, U. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology*, *18*(7), e3000737. <https://doi.org/10.1371/journal.pbio.3000737>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535–b2535. <https://doi.org/10.1136/bmj.b2535>
- Molléri, J. S., Petersen, K., & Mendes, E. (2018). Towards understanding the relation between citations and research quality in software engineering studies. *Scientometrics*, *117*(3), 1453–1478. <https://doi.org/10.1007/s11192-018-2907-3>
- Moreira, T. (2005). Diversity in clinical guidelines: The role of repertoires of evaluation. *Social Science and Medicine*, *60*(9), 1975–1985. <https://doi.org/10.1016/j.socscimed.2004.08.062>
- Moreira, T. (2007). Entangled evidence: Knowledge making in systematic reviews in healthcare. *Sociology of Health and Illness*, *29*(2), 180–197. <https://doi.org/10.1111/j.1467-9566.2007.00531.x>

- Müller, R., & de Rijcke, S. (2017). Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation*, 26(3), 157–168. <https://doi.org/10.1093/reseval/rvx023>
- Nascimento, D. P., Gonzalez, G. Z., Araujo, A. C., & Costa, L. O. P. (2020). Journal impact factor is associated with PRISMA endorsement, but not with the methodological quality of low back pain systematic reviews: A methodological review. *European Spine Journal*, 29(3), 462–479. <https://doi.org/10.1007/s00586-019-06206-8>
- NLM. (2017). *Frequently asked questions about indexing for MEDLINE*. <https://www.nlm.nih.gov/bsd/indexfaq.html>
- NLM. (2018). What's new for 2019 MeSH. *NLM Tech Bull*, 426(e6). https://www.nlm.nih.gov/pubs/techbull/nd18/nd18_whats_new_mesh_2019.html
- Ojasoo, T., Maisonneuve, H., & Dore, J.-C. (2001). Evaluating publication trends in clinical research: How reliable are medical databases? *Scientometrics*, 50(3), 391–404. <https://doi.org/10.1023/A:1010598313062>
- Page, M. J., & Moher, D. (2016). Mass production of systematic reviews and meta-analyses: An exercise in mega-silliness?: Commentary: Mass production of systematic reviews and meta-analyses. *The Milbank Quarterly*, 94(3), 515–519. <https://doi.org/10.1111/1468-0009.12211>
- Page, M. J., & Moher, D. (2017). Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement and extensions: A scoping review. *Systematic Reviews*, 6(1), 263. <https://doi.org/10.1186/s13643-017-0663-8>
- Page, M. J., Shamseer, L., Altman, D. G., Tetzlaff, J., Sampson, M., Tricco, A. C., Catalá-López, F., Li, L., Reid, E. K., Sarkis-Onofre, R., & Moher, D. (2016). Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. *PLOS Medicine*, 13(5), e1002028. <https://doi.org/10.1371/journal.pmed.1002028>
- Patsopoulos, N. A., Analatos, A. A., & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *JAMA*, 293(19), 2362–2366.
- Petersohn, S., Biesenbender, S., & Thiedig, C. (2020). Investigating assessment standards in the Netherlands, Italy, and the United Kingdom: Challenges for responsible research evaluation. In K. Jakobs (Ed.), *Shaping the future through standardization* (pp. 54–94). IGI Global.
- Pussegoda, K., Turner, L., Garritty, C., Mayhew, A., Skidmore, B., Stevens, A., Boutron, I., Sarkis-Onofre, R., Bjerre, L. M., Hróbjartsson, A., Altman, D. G., & Moher, D. (2017). Systematic review adherence to methodological or reporting quality. *Systematic Reviews*, 6(1), 131. <https://doi.org/10.1186/s13643-017-0527-2>
- de Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, 25(2), 161–169. <https://doi.org/10.1093/reseval/rvv038>
- Rousseau, R. (2012). Basic properties of both percentile rank scores and the I3 indicator. *Journal of the American Society for Information Science and Technology*, 63(2), 416–420. <https://doi.org/10.1002/asi.21684>
- Royle, P., Kandala, N.-B., Barnard, K., & Waugh, N. (2013). Bibliometrics of systematic reviews: Analysis of citation rates and journal impact factors. *Systematic Reviews*, 2(1), 74. <https://doi.org/10.1186/2046-4053-2-74>
- Simons, A., & Schniedermann, A. (2021). The neglected politics behind evidence-based policy: Shedding light on instrument constituency dynamics. *Policy and Politics*, 49(3), 1–17. <https://doi.org/10.1332/030557321X16225469993170>
- Solomon, M. (2015). *Making medical knowledge* (1st ed.). Oxford University Press.
- Stegenga, J. (2011). Is meta-analysis the platinum standard of evidence? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 42(4), 497–507. <https://doi.org/10.1016/j.shpsc.2011.07.003>
- Stephen, D., & Stahlschmidt, S. (2021). *Performance and Structures of the German Science System 2021. Studien zum deutschen Innovationssystem* (No. 5–2021; Studien Zum Deutschen Innovationssystem, pp. 1–37). EFI – Expertenkommission Forschung und Innovation. https://www.e-fi.de/fileadmin/Assets/Studien/2021/StuDIS_05_2021.pdf
- Strassheim, H., & Kettunen, P. (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence and Policy: A Journal of Research, Debate and Practice*, 10(2), 259–277. <https://doi.org/10.1332/174426514X13990433991320>
- Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Temple University Press.

- Timmermans, S., & Epstein, S. (2010). A world of standards but not a standard world: Toward a sociology of standards and standardization. *Annual Review of Sociology*, 36(1), 69–89. <https://doi.org/10.1146/annurev.soc.012809.102629>
- van der Pol, C. B., McInnes, M. D. F., Petrich, W., Tunis, A. S., & Hanna, R. (2015). Is quality and completeness of reporting of systematic reviews and meta-analyses published in high impact radiology journals associated with citation rates? *PLoS One*, 10(3), e0119892. <https://doi.org/10.1371/journal.pone.0119892>
- Vrieze, J. (2018). Meta-analyses were supposed to end scientific debates. Often, they only cause more controversy. *Science*. <https://doi.org/10.1126/science.aav4617>
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), 851–872. <https://doi.org/10.1007/s11192-012-0775-9>
- Wang, J.-Y., & Jin, X.-J. (2019). Evidence-based medicine in China. *Chronic Diseases and Translational Medicine*, 5(1), 1–5. <https://doi.org/10.1016/j.cdtm.2019.02.001>
- Whitley, R. (2000). *The intellectual and social organization of the sciences* (2nd ed.). Oxford University Press.
- Wyatt, S., Milojević, S., Park, H. W., & Leydesdorff, L. (2017). Intellectual and practical contribution of scientometrics to STS. In U. Felt, R. Fouche, C. A. Miller, & L. Smith-Doerr (Eds.), *The handbook of science and technology studies* (4th ed.). MIT Press.
- Yuan, Y., & Hunt, R. H. (2009). Systematic reviews: The good, the bad and the ugly. *The American Journal of Gastroenterology*, 104(5), 1086–1092. <https://doi.org/10.1038/ajg.2009.118>
- Zollman, K. J. S. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5), 574–587. <https://doi.org/10.1086/525605>