Standardization in science: effects and issues of guidelines for biomedical reporting

Schniedermann, A.

Chapter 5

The dissemination of PRISMA

The "culture" of reviewing the literature is so fundamental to my own professional life that I too may forget that in comparison with research discoveries one reads about in the press, and for which Nobel Prizes are awarded, reviewing may seem to the uninitiated to be a relatively humdrum topic.

(Garfield, 1987, p. 113)

5.1 <>

In the previous chapters, we have learned about the social and intellectual contexts of the PRISMA guideline and how standards can embody epistemic values or notions of research qualities. In this and the following chapters, I will focus on the dissemination and reception of PRISMA within the biomedical community. In this regard, we have already learned about the dissemination strategies of PRISMA's developers and how the guideline was designed to support its wide adoption. But this and the following chapters will turn towards the arenas where PRISMA is actually used: Systematic reviews in biomedicine. But before I turn to the textual laboratory of conducting and writing systematic reviews (chapter 7), I want to approach the topic from a bird's-eye perspective, i.e., with a scientometric analysis of systematic reviews.

The previous chapters have revealed that the role of systematic reviews in general, as well as the importance of standardized methods and guidelines, differs by contextual factors. Such factors can be biomedical subfields, national policy contexts, or trends in scholarly communication. Another dimension is citation impact analysis, which is always the elephant in the room when we think about scientometric data. Although I use citations to the guideline as identifiers for PRISMA-compliant systematic reviews, I will also compare the citation impact of these reviews to other review types. In the previous chapters, I have argued how PRISMA compliance is a form or dimension of research quality. Thus, by analyzing citation data, I aim to illuminate and discuss the links between *ex ante* standardization, citation impact, and concepts of research quality.

Of course, both – PRISMA usage and citation impact – can only be considered as proxies for the more abstract and latent concept of quality. We will also turn back to this question in 7, when we look at the role of PRISMA for when researchers read or peer review the systematic reviews written by others. However, now we not only look at PRISMA compliance when it investigates levels of standardization. Rather, I will provide a more holistic perspective in which systematic reviews are considered as more standardized forms of review articles. In this sense, this and the next chapter employ an ordinal conception of the level of standardization. This conception considers narrative reviews as least standardized form of secondary literature, followed by systematic reviews as fairly standardized, PRISMA-citing systematic reviews as very standardized, and Cochrane systematic reviews as the most standardized review genre.

Proxies and theoretical constructions permeate bibliometric research. There is a huge variance in how well bibliometric proxies fit the real world. The mindful bibliometrician must account for these variances, either by carefully normalizing the data to make it more comparable and equal, or by being more careful with her conclusions and interpretations of measured differences. In this regard, I interpret the data and results in this chapter by following the principle of modest epistemology which I described in chapter 2.3. This manifests in the usage of median-based statistics and non-linear methods, but also in a more careful and open interpretation of what citation impact means.

Decades of bibliometric research have shown how the publication cultures between the life sciences and other fields differ. There are short journal papers in medicine while it is important to author a monograph in the humanities and social sciences (Aksnes et al., 2019; Hammarfelt & De Rijcke, 2015; Hicks, 2004). This also means that there are longer timespans in the production and reception of scholarly works in those fields, compared to the fast-paced world of medical journals where papers receive their maximum citation impact no later than two years, with a few exceptions (Hornbostel et al., 2008; Ke et al., 2015). For this reason, normalizations on the basis of fields or journals are required not to compare apples with oranges in absolute terms, but with regards to their own reference group (Waltman & van Eck, 2019).

Although the bibliometric analyses of my research are restricted to the medical sciences which mitigates some of these systematic errors, it must be noted that medicine as a field is itself a construct or proxy. In fact, epistemic cultures and publication behaviors in medical subfields can differ as much as they might differ between medicine as a whole and the humanities. For example, it has been argued that nursing science sits at the crossroads between the social and clinical sciences due to its strong tradition of qualitative research, interpretive epistemology, and practice-based knowledge which is often tacit and uncodified (Fedyk, 2023). Because of such differences, we must negotiate the alluring wealth of bibliometric data with careful analysis techniques and modest interpretations.

The epistemic purpose of bibliometric analyses is different from the qualitative analyses. First, they transform or extend the case study, not by making its findings generalizable, but by showing the ubiquity of the case. This emphasizes the translational value of the case study while refraining from the formulation of definitive theories or laws. Second, the quantitative findings can serve as boundary objects that allow us to communicate with other epistemic domains, as I have explained in chapter 2. For example, the data and insights obtained by these analyses can be more easily communicated to research evaluators, metascientists, and medical researchers because quantitative forms of reasoning also dominate these fields. In addition, we have seen in chapter 1.6 how scientometric indicators have long since become an integral part of the work experiences and self-perceptions of biomedical researchers. In this regard, I will use the results from these two chapters to inform interviews and interactions with PRISMA users, which we will see in chapter 7. Third and last, bibliometric analyses can approach some complex dimensions at the macro level despite their shortcomings. For example, the previous two chapters have already highlighted the role that the biomedical subfield or subcommunity plays in the adoption or rejection of the guideline. Although bibliometrics-based field classifications remain proxies that sometimes may hide more than they reveal, they can provide a good starting point for further investigations (see also Held & Velden, 2022).

5.2 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, 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 only little to add, forced leading experts to proclaim a crisis in biomedical research. Besides 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, p. 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” (Meldrum, 2000, p. 755), due to the 1960s FDA regulation, experts demanded more systematic assessments of medical knowledge in order to improve the health and life of patients (Daly, 2005; Solomon, 2015). As a result, medical disciplines increasingly employed meta-analyses and systematic reviews 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 (Daly, 2005).

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., 2009e). 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 what extent systematic reviews comply with the guideline’s rules (Page & Moher, 2017). They conclude that, although there is room for improvement, the guideline has 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 factually 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 & Schniedermann, 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; Patsopoulos et al., 2005;

Royle et al., 2013). But allegations of strategic behavior in academia are hard to prove on the basis of bibliometric measurements alone. Nevertheless, bibliometric assessments can point towards potential misuses and inform more profound investigations (Krell, 2014; S. Wyatt et al., 2017).

Several studies elaborated on 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" (Page et al., 2016, p. 1) by focusing on characteristics such as language, inclusion criteria, or whether a protocol was pre-registered or published. Other observations are the high growth rates and publication counts of systematic reviews and resulting problems, such as perceived information overload (Bastian et al., 2010; Ioannidis, 2016; Page et al., 2016). Other analyses focus on national affiliations, number of authors, funding sources, or citation impact. For example, a study by Alabousi et al. (2019) found a positive correlation between the Journal Impact Factor and systematic reviews published in the field of medical imaging.

Studies that take methodological standards such as PRISMA into account often focus on specific subfields 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 (van der Pol et al., 2015). A study by Nascimento et al. (2020) found a positive correlation between the endorsement of PRISMA and the Journal Impact Factor for systematic reviews about low back pain. Similarly, Mackinnon et al. (2018) found a weak correlation between reporting quality and 5-year Journal Impact Factor in dementia biomarker studies. Using a multidimensional framework to analyze papers from software engineering studies, Molléri et al. (2018) conclude that study rigor but not reporting quality is correlated with citation impact.

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 differ by medical subfield (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.

5.3 Data and corpus construction

To compare the bibliometric characteristics of systematic reviews and PRISMA-based systematic reviews, a unique corpus was constructed by a combination of bibliographic data from PubMed and Web of Science. Document types in Web of Science lack accuracy and necessary detail, especially in the case of review articles (Donner, 2017). In contrast, PubMed consists of several document types related to secondary research and introduced the "Systematic Review-type in 2019 because of this genre's importance for medical decision-making" (NLM, 2018).[52]

In the first step, publication types and MeSH keywords were retrieved for 31 million PubMed items during March and April 2020 and stored in a PostgreSQL database (see Appendix: D.1 PubMed data retrieval tool). 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 in-house version of Web of Science at the German Kompetenzzentrum Bibliometrie and restricting the publication years from 1991 to 2016 to account for changing indexing policies at MEDLINE (NLM, 2017), the resulting set contained 10 million records.

To validate the quality of the different document type systems, precision and recall for the Web of Science classification were calculated, using PubMed as a gold standard (Baeza-Yates & Ribeiro-Neto, 2011). To compare both systems, the rich multiclassification from PubMed was reduced hierarchically to match the Web of Science types "Article," "Review," "Editorial," "Letter," and "News." In this process, these document types were assigned in reverse order so that any item containing "News" was always relabeled as "News," while the "Article" category was assigned last.

[52] 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, or other.

The results can be seen in Table 3. 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” in PubMed but not in Web of Science (false negative). Low overall assignment quality can also be found for editorials (F1-Score = 0.54) or news (F1-Score = 0.72). Although the different type assignments combine into an overall assignment quality of only 0.80, the sheer amount of common articles improves the size-dependent measure by 10% to an average assignment accuracy of 0.90. These results confirm earlier analyses (Donner, 2017) and show the huge differences between classification systems for document types in Web of Science and PubMed (see also Harzing, 2013).

TABLE 3 DOCUMENT TYPE ASSIGNMENT QUALITY IN WEB OF SCIENCE AGAINST PUBMED

Publication Type	Simple Precisions	Simple Recall	F1 Score
Article	0.94	0.95	0.94
Review	0.85	0.52	0.65
Editorial	0.38	0.98	0.54
Letter	0.95	0.95	0.95
News	0.82	0.64	0.72
Micro-average	0.90	0.90	0.90
Macro-average	0.79	0.81	0.80

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 DOIs have been downloaded from the PRISMA website (www.prisma-statement.org), as well as the EQUATOR network website (www.equator-network.org).

5.4 Statistical analysis

For the comparison of systematic reviews and PRISMA-based systematic reviews, different variables have been calculated (see Appendix for more details and formulas). For the annual and compound growth rates, basic publication outputs were used. All variations other than document types are based on metadata 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 due to their stronger commitment to evidence-based medicine (Groneberg et al., 2019).

For the citation analysis, all publications in the set have been compared according to their 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 (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 commonly used 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 the mean differences between the document types.

Generally, percentile ranks reflect to which citation impact group a publication belongs. Groups can be delineated at will; for example, a common set is the top 10% percentile, meaning that these papers are cited more often than 90% of the rest within this field and publication year (Rousseau, 2012). But citation scores are discrete values, which complicate predefined rank classes because the percentages are based on the number 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 to one-third of the whole set, rather than 10 percent. In order to preserve predefined percentile ranking sets, fractional assignments have been introduced, which, in turn, give up the binary nature of the ranks (Waltman et al., 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* (ii), and including cumulative percentiles, *CPIN* (i) (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).

$$i)\ CPIN_j = \sum_{k=0}^{j} \frac{c_j}{t} * 100$$

$$ii)\ CPEX_j = CPIN_j - \frac{c_j}{t} * 100$$

After *CPIN* or *CPEX* have been calculated, each focal paper can be assigned such values based on field classification, publication year, and citation window. Although both represent cumulative percentages, their interpretation is slightly different. *CPEX* represents the number 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 number of other publications that are cited as equal to 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.

5.5 Results and discussion

5.5.1 The growth and dissemination of systematic reviews

As displayed in Table 4, 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 literature are secondary research, or review articles. These consist mostly of non-systematic reviews (950 k), systematic reviews (95 k), as well as PRISMA-based systematic reviews (11 k).

TABLE 4 NUMBER OF PUBLICATIONS AND GROWTH RATES FOR EACH DOCUMENT TYPE

	Number of Publications	Average annual growth rate (in %)		Overall growth (in %)	
		2002-2009	2009-2016	2002-2009	2009-2016
Article	8,122,081	5.94	4.77	54	38
Review	951,215	4.47	4.70	32	38
Systematic Review	96,259	21.92	14.12	300	148
Sys. Review (PRISMA)[53]	11,085	n.a.	81.85	n.a.	2675
Letter	427,579	3.42	3.84	36	27
Editorial	226,085	9.11	4.82	87	39
News	70,157	-0.03	-2.30	0.93	-18
TOTAL	9,904,461	5.75	4.83	52	39

By comparing the timespans before (2002–2009) and after the publication of PRISMA (2009–2016), it becomes clear that the overall publication inflation decreased from 52% to 39%. Besides this demise of genres such as “News” items, which have a shrinking annual output of −2.3%, especially the growth rates of primary research articles slowly decrease from 5.9% before, to 4.8% annually, after the publication of PRISMA.

In contrast, the set shows a growing output of research syntheses. In addition to a slight increase in the annual growth rates of reviews (from 4.5% to 4.7%), both systematic reviews and PRISMA-based systematic reviews have shown 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 has 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 provides a scatterplot of the rates of systematic reviews and PRISMA-standardized systematic reviews among all secondary research. Fields represent Web of Science subject categories that published

[53] For PRISMA-guided reviews, the number of publications from 2009 has been neglected because the guideline was released in 2009, which reduces the timespan for application and citation to less than one year.

at least 500 reviews and 50 systematic reviews from 2009 to 2015. In addition, the coloring reveals which of these categories were assigned to the guideline documents.

The figure shows how *Surgery* has rather moderate rates of systematic reviews (18%) and PRISMA-based reviews (2.4%) but 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 highest rate in the dataset. *Information Science & Library Science* (38%) and *Medical Informatics* (37%) also exhibit high rates. On the other side, of all assignations of *Chemistry*, only 0.5% go to systematic reviews, which is the lowest value. Others are *Biophysics* (1.1%) or *Cell Biology* (1.5%). In contrast, *Evolutionary Biology* has a rather low rate of systematic reviews (7.6%) and was not assigned to any PRISMA-based systematic reviews, representing the lowest value in this category. Not surprisingly, *Health Care Sciences & Services* and *Public, Environmental & Occupational Health,* fields in which the PRISMA guidelines have been published, offer higher rates in both types of assignments.

FIGURE 1 PORTIONS OF SYSTEMATIC REVIEWS BY FIELD

How the field assignments to PRISMA-based systematic reviews changed over time is shown in Figure 2. The figure shows ranked cumulative proportions of field assignments and cumulative counts to PRISMA-based systematic reviews from 2010 to 2015, filtered by a 2% threshold for assignments in 2015.

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 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, Figure 1 showed that the differences in portions and counts between the fields are much smaller compared to what can be observed for systematic reviews in general.

FIGURE 2 PORTIONS OF SYSTEMATIC REVIEWS BY FIELD AND ANNUAL RANK (2010-2015)

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 to 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 are rather common phenomena in biomedicine (Whitley, 1985/2000), such attempts are sometimes accompanied by criticism and reluctance because standards conceal contingencies and nuances, or may not represent the working culture of some subfields altogether (Timmermans & Berg, 2003). For example, communities with a strong track record of qualitative research turned out to be rather critical of systematic reviewing (D. Cohen & Crabtree, 2008).

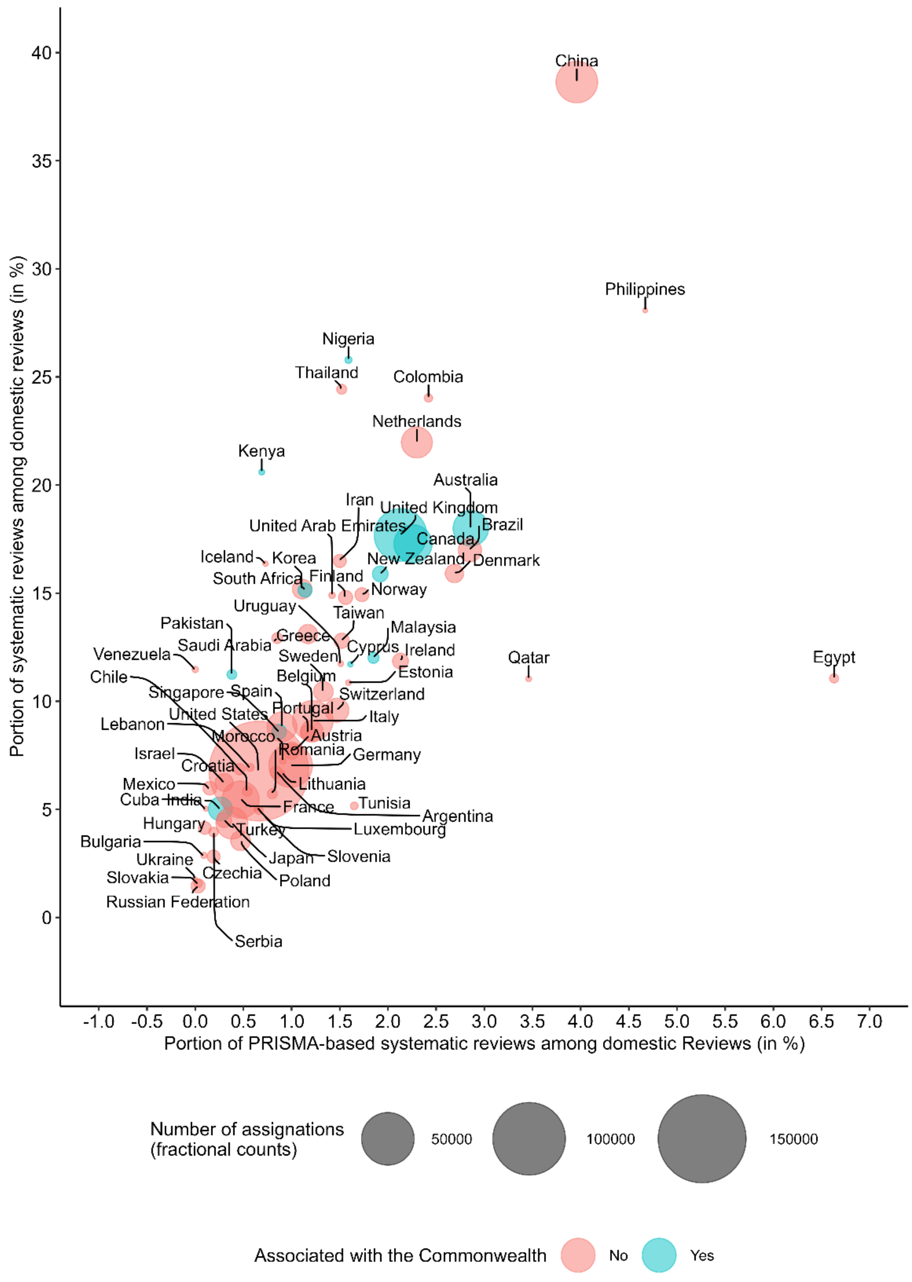
One reason for strong objections is the more procedural and prescriptive way that systematic reviews or meta-analyses create medical evidence. Besides the objection that this employs a rather narrow perspective on scientific methodology and neglects the plurality in medical research (Berkwits, 1998), 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.

Besides 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, assignations 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), accumulating a total of 499 assignations and 323 assignations, respectively, to this genre. These values correspond to 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). 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 and compared (Maheshwari & Maheshwari, 2012), scientific disputes are not always solvable on this way.

5.5.2 Systematic reviews by countries

Although the biomedical sciences are an internationally collaborating community, national research systems vary substantially in their publication output. The national variations in the production of research syntheses and the portions of systematic reviews and PRISMA-based systematic reviews are shown in Figure 3. It shows the overall number of research syntheses, portions of systematic reviews and PRISMA-based systematic reviews by country, and national association with the Commonwealth. The data is based on fractional assignment of the authors' affiliations.

FIGURE 3 PORTIONS OF SYSTEMATIC REVIEWS BY COUNTRY



Although 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 compared to other item types. 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 the 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 value in the dataset.

FIGURE 4 PORTIONS OF SYSTEMATIC REVIEWS BY COUNTRY AND ANNUAL RANK (2010-2015)

The annual ranks of national assignments to PRISMA-based systematic reviews are displayed in Figure 4. It shows cumulative counts and ranks the countries by cumulative portions from 2010 to 2015.

In the first year after the inception of the guideline, leading producers were the United Kingdom (23%), the United States (21%), and Canada (12%). While the U.S. took first position in 2013, accounting for 20% of all PRISMA-based systematic reviews in 2015, China especially showed even higher growth rates. While marking seventh with 4.8%, its portions grew steadily year over year, taking the second rank in 2015 with over 1000 items and 18% of the overall share.

Groneberg et al. (2019) have argued that strong commitments towards evidence-based medicine and a respective dominance in the production of systematic reviews can be associated with the Commonwealth. 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 their PRISMA-based counterparts than other leading science nations (Figure 3 and Figure 4). Other countries with high rates of these genres, most notably the Netherlands or Denmark, are also among the top ten adopters of evidence-based medicine. Although the "evidence-based" movement promotes and comes with a variety of knowledge generation practices, systematic reviewing remains an important cornerstone in evidence-based medicine, as other studies have indicated (Blümel & Schniedermann, 2020; Ojasoo et al., 2001). In that sense, the data provided here generally supports the link between evidence-based paradigms and the emergence of systematic reviews.

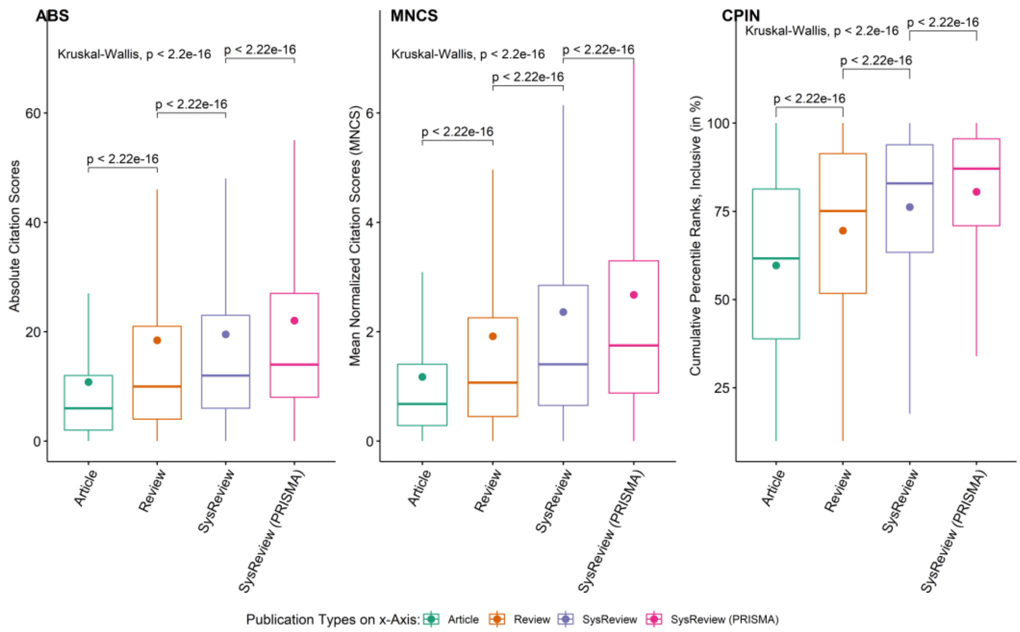
Besides being fourth in absolute number of reviews, China offers substantially higher rates of systematic reviews (39%) as well as PRISMA-based systematic reviews (4.0%) than all other countries. Although the exact causes may be very complex and unable to be identified, two potential causes are proposed. First, being a rising science leader, the majority of Chinese scientists are latecomers to the international science community (Liang et al., 2020). As such, this group may be biased toward 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 at healthcare reform, China turned more explicitly toward evidence-based 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 (J.-Y. Wang & Jin, 2019).

5.5.3 Citation impact of PRISMA-based reviews

Differences in the citation impact of the analyzed document types are displayed in Figure 5. The figure shows the median and mean (dots) citation impact ranks by document type for absolute citations (ABS), mean-normalized citations (MNCS), and cumulative percentile ranks (CPIN). It is based on publications published from 2009–2015 and takes a 3-year citation window into account. The y-axes for ABS and MNCS are stripped of extreme outliers for the purpose of better visualization

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 a median of 14 absolute citations (mean = 22), 1.57 median 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 fewer citations in three years after publication. In comparison, other systematic reviews receive a median of 12 absolute citations (mean = 19.5), 1.41 median 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 ($p < 0.001$).

FIGURE 5 CITATION IMPACT OF DIFFERENT DOCUMENT TYPES (ABSOLUTE, MEAN-NORMALIZED, PERCENTILE RANK)

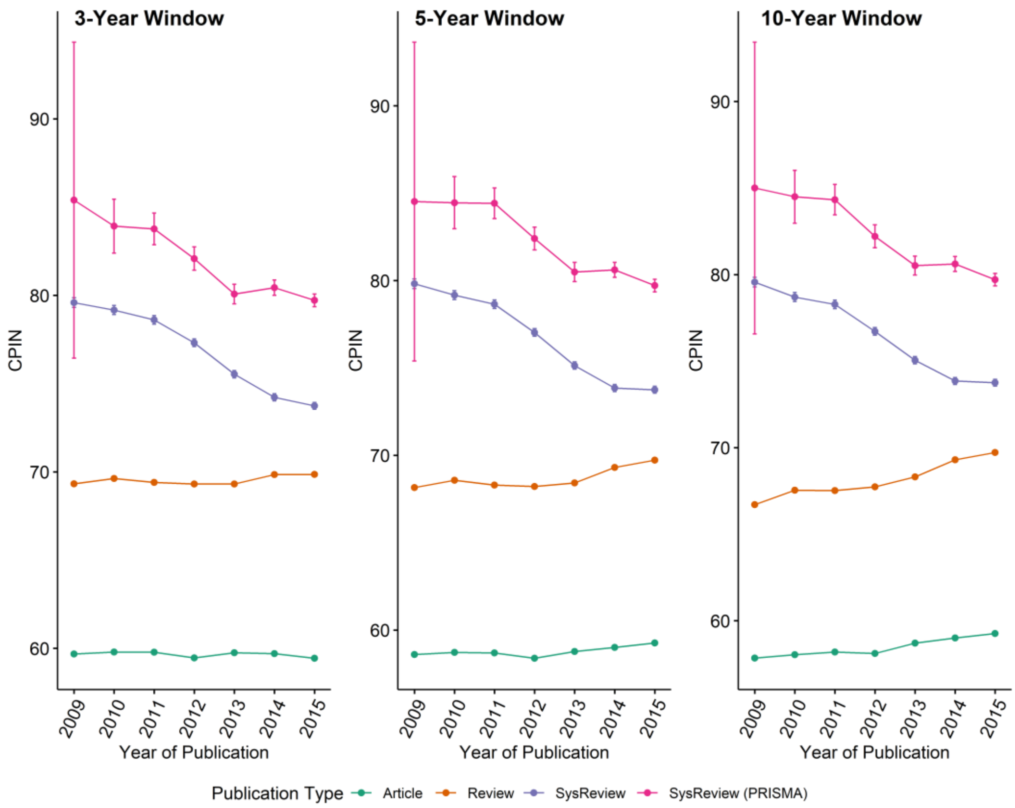


The annual development of citation impact changes over time, which is shown in Figure 6. The figure provides the development of the annual mean cumulative percentile rank

(CPIN) for the different document types. The three panels represent different citation windows to account for possible long-term differences (see also J. Wang, 2013). Since 2018 is the maximum year of citation data, the 5- and 10-year windows are short of data in 2014 and 2009, respectively.

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 to 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. Because we use citation data up to 2018, the 10-year window essentially encompasses all 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.

FIGURE 6 ANNUAL CITATION PERCENTILE RANKS OF DIFFERENT DOCUMENT TYPES BY CITATION WINDOWS



Rankings are the same for all three citation windows. The impact of systematic reviews and PRISMA-based systematic reviews is higher for all years but also constantly decreasing, most likely because both genres experience a stronger growth in publication counts compared to ordinary articles or non-systematic reviews. In addition, the results provide greater error values for PRISMA-based systematic reviews since this group has the fewest item counts (e.g., 6 items in 2009).

The differences in the changes of CPIN values indicate that the document types have different citation patterns, with articles and reviews being cited more consistently in the long run. At a first glance, phenomena like delayed recognition (Ke et al., 2015) or the "kiss of death" (Lachance et al., 2014, p. 1501) may influence the 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 to what scholars now call "narrative review" (Grant & Booth, 2009, p. 98) may explain why the former has 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 quickly become outdated by new primary studies. Ideally, the same or other authors update them whenever new trial results or review methods occur (J. H. Elliott et al., 2017). This limits their potential citation lifespan, since medical experts are supposed to rely on the most recent findings. In contrast, non-systematic 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 rendering 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 possible interpretations are proposed.

First, research that complies with reporting standards is credited by higher citation impact afterwards because it is regarded as of higher quality by its peers who then cite it accordingly (1). While the conception that citation indicates quality still perpetuates in biomedicine and elsewhere (e.g., 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 be found occasionally today. Speaking of a Kuhnian revolution in bibliometrics, Lutz Bornmann and Robin Haunschild (2017) argue that in contemporary evaluative bibliometrics, citation impact should be considered as only one aspect of a multidimensional concept of research quality, especially since there are many different citation behaviors. For example, authors cite publications if they value the latter's solidity and plausibility, originality and novelty, or scientific value, or societal value, which are complex dimensions that sometimes even conflict with 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, the citation impact profiles of the different genres allow for another interpretation of standards and scientific excellence. Whatever causes high-impact research also makes it more likely to adopt new methodological standards such as PRISMA reporting guidelines (2). Since biomedicine could be understood as what Richard Whitley has called a "professional adhocracy" (Whitley, 1985/2000, p. 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 (Yuan & Hunt, 2009; see also Gieryn, 1999). In this setup, high-impact researchers also serve as first movers in adopting new standards. After the leading scientists have defined these standards, other researchers consent afterwards to become part of the group that publishes *the good* systematic reviews. Similar dynamics have also been observed for academic communities consenting on theories or methods (Whitley, 1985/2000), or the role of reviews in the formation of new disciplines (Blümel, 2020). Put in other words, there might be a latent factor that causes both the adoption of PRISMA and also high citation impact.

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; S. 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 of what must 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. 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.

5.6 Conclusion

This chapter provided a comparative analysis of different document types in biomedical research. To understand the relationship between methodological quality and citation impact, index data from PubMed was used to differentiate between ordinary systematic reviews and those that comply with the PRISMA reporting standard. The study provided a general overview of the growth and dissemination of systematic reviews and their standardized counterparts, and it compared their citation impact using different indicators.

The results suggest that systematic reviews and PRISMA-based systematic reviews are a current trend in biomedical publishing. This trend is associated with significantly higher production rates and increased scholarly attention, as evidenced by citation impact. Although it is a trend, systematic reviews have become established and will continue to be a significant part of future biomedical literature and its genre mix. However, we must expect that the rate of production as well as their citation impact will stabilize at a certain level. This chapter left a potential causal relationship between PRISMA usage and citation impact open. But future investigations will reveal if PRISMA-based systematic reviews can hold their citation dominance, or if PRISMA compliance will turn from a minimum standard into a minimum requirement that, in turn, allows for a significant variance in perceived research quality and impact.

The differences among the biomedical subfields that are associated with systematic review publications suggest that this genre has become integral to some disciplines while remaining a side issue in others. However, the strong association with more technically defined subfields such as medical informatics or information science suggests that there might be not only a close fit between systematic reviewing and particular research cultures but also with specific tasks and skills that are required to assemble systematic reviews. In this sense, a closer look into what skills and practices are employed in systematic review projects and how they relate to technical formulas such as the PRISMA reporting checklist would provide a valuable addendum to this line of research.

Lastly, the national patterns in the production of systematic reviews showed the association of the genre with certain political styles. On the one hand, formal evidence syntheses are required for evidence-based paradigms in medical decision-making. On the other hand, the incentive structure of national science policies can draw substantial attention towards systematic reviews because of their citation impact and perceived quality. However, national research cultures may also be affected by current epistemic trends, as these allow them to claim epistemic spheres and intellectual authority. However, further research is required to better understand the social structure of review authors and their organizational settings and contexts.