Guideline Impact Factor – A new indicator to assess journals cited in medical guidelines

Valeria Aman*, Nikita Sorgatz**

* aman@dzhw.eu
orcid.org/0000-0001-8370-780X
German Centre for Higher Education Research and Science Studies GmbH, Berlin, Germany

** sorgatz@dzhw.eu
orcid.org/0000-0003-3878-6267
German Centre for Higher Education Research and Science Studies GmbH, Berlin, Germany

Despite of its many limitations, the Journal Impact Factor (JIF) is widely used to evaluate research institutions and individual researchers. Using references from 41 German medical guidelines we show that clinical relevance as assessed by guideline authors is uncorrelated to the JIF suggesting that a journal’s clinical relevance is independent of its JIF. As a consequence, evaluations solely relying on the JIF end up under-valuing clinically important research. We therefore propose a Guideline Impact Factor (GLIF) quantifying the relevance of journals for medical guideline development as an independent quality criterion for journals.

1. Introduction

This paper presents references in 41 German medical guidelines as a novel source of evidence for the valuation of medical research. Although medical guidelines provide relevant diagnostic, preventative and treatment information and fulfil an important role in helping medical staff to bridge the gap between research and daily practice (Kryl et al., 2012), they remain largely unconsidered in bibliometric performance measures and evaluations (Herrmann-Lingen et al., 2014). There are two main reasons for this, first the dominance of the Journal Impact Factor (JIF) in the valuation of scientific work, especially in medicine. Second, many guidelines are not published in journals and exist only as grey literature not included in bibliometric databases. Other specialised databases for medical guidelines began to emerge recently (Erikksson et al., 2020) but are still in their infancy, which means that the relevance of individual studies or journals for medical guidelines cannot be easily assessed.

While the JIF can serve as an indicator of a journal’s prestige and may help identifying journals with high readership, it is widely recognized that the JIF is not an appropriate indicator of the quality or impact of individual research articles (Bornmann & Daniel, 2008; Seglen, 1997). Moreover, by focussing on the number of citations in academic journals, the JIF underestimates the impact of clinical intervention research compared with basic medical research (Eck et al., 2013), as clinical papers are less cited than basic research papers.

One way of counteracting this bias and getting at the impact of research on clinical practice is through citations in medical guidelines. Medical practitioners such as physicians, surgeons and nurses are expected to keep up with advances in medical knowledge but only have limited time to do so. Ideally, medical guidelines provide practitioners with the information needed enabling them to improve patient outcomes (Grimshaw & Russell, 1993). In addition, guidelines are an easily accessible source of information for patients. Patient groups might also be involved in the creation of guidelines.
In order to gather the best evidence available for the treatment of diseases and to identify or report best practices for treating patients, medical experts review existing evidence and formulate official guidelines for diagnosis and treatment (Woolf et al., 1999). While there is considerable heterogeneity between different guideline programs in terms of institutional processes (like peer-review and consensus mechanisms), available resources and the eventual quality of the produced guidelines, references in medical guidelines are evidence that research findings have been used to inform the day-to-day practice of medical staff (Thelwall & Maflahi, 2016).

Previous work (e.g. by Grant et al., 2000; Lewison & Sullivan, 2008, Andersen, 2013) comparing the literature cited in guidelines to the wider medical literature shows that while guidelines usually reference highly cited papers according to Web of Science (WoS), they are more likely to reference clinical work as well as papers coming from the country the guideline is written for. The cited articles also tend to be more highly cited than comparable articles in terms of the same publication year, issue and journal except for recently published ones, which did not have time to accrue a lot of citations (Thelwall & Maflahi, 2016).

We propose a Guideline Impact Factor (GLIF) to gauge the relevance of scientific journals for medical guideline development. Analogous to the JIF, which is a measure of how often articles from a journal are cited in other journals, the GLIF is a measure of how often articles from a journal are cited in guidelines. There are two reasons in favour for such an approach within the current reward system in medical science. The first one is normative: being cited in a medical guideline corroborates that a study achieved societal impact and guided medical practice (Thelwall & Maflahi, 2016) and should therefore be rewarded. The second one is more technical: as mentioned before, the JIF is biased towards a particular type of research published in Anglo-American journals. Taking citations by medical guidelines into account could mitigate this bias by championing clinical as well as local research.

The remainder of this paper is a proof of concept for the GLIF based on all German medical guidelines from General Medicine and Oncology issued between 2017 and 2022, yielding 41 guidelines in total.

2. Data and Methods

2.1 Data

The focus of the study are German “evidence- and consensus-based” S3 medical guidelines, which are considered to be the most reliable. They are also the most time-consuming type of guideline. In contrast to lower classes of German medical guidelines (i.e. S2e, S2k, S1), these guidelines are characterized by a structured consensus-building process performed by a representative committee using a systematic literature search carried out according to a priori criteria. The guideline process is coordinated by the Association of the Scientific Medical Societies (AWMF), a scientific network of 180 member societies.

We gathered the cited references in S3 guidelines that were published between 2017 and 2022 in General Medicine and Oncology. The 41 guidelines were collected from the society websites in January 2023 and consist of ten General Medicine guidelines by the German College of General Practitioners and Family Physicians (Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin e.V. - DEGAM) and thirty-one Oncology guidelines issued by the German
Guideline Program in Oncology (GGPO, in German: Leitlinienprogramm Onkologie). As the JIF is central to this study, data from Clarivate’s Web of Science (WoS) and Journal Citation Reports is also used.

2.2 Method for reference extraction and matching

Because the guidelines are only available as PDF documents, the references had to be extracted to be turned into structured data. We used a two-step procedure. First, the document parsing and extraction toolchain Parsr was used to extract a total of 29,501 raw reference strings, followed by a manual cleaning stage. In a second step we queried the Crossref database with those strings to attain the relevant bibliographic data (authors, title, journal, publication year, volume, issue, DOI) for each reference in a structured format. References that could not be resolved via Crossref were separated into bibliographic components such as authors, title, journal, and publication year with the help of regular expressions.

This dataset of structured references was used to match the documents cited in the guidelines to documents in the WoS database. About 5,200 references that remained either unparsed or unmatched were revised by hand and a match was found for one third of the revised articles. The matching procedure finally provided a corresponding document in WoS for 25,104 out of 29,501 references (85.1%). In some cases, the names of the journals in the references did not match the name of the journal in WoS due to abbreviations, misspellings, or alternative names. A similar problem was also revealed during our manual revision: articles written in German or ones that were mistakenly provided with a German title were often not matched.

In addition, we did not match cited references published before 1980 because they are not covered in our inhouse-version of WoS and are thus excluded.

2.3 Guideline Impact Factor - GLIF

The GLIF is a journal level metric that puts the number of references to a journal in each reference in relation to all references in a guideline. For each journal, an aggregate measure across all guidelines in which the journal is cited is calculated. A weighting is applied to account for the heterogeneity in terms of reference list length across guidelines and the number of articles across journals. The calculation is described in the formula below.

\[
glif(j) = 1000 \times \sum_{\text{guideline } g} \frac{\#\text{references}(g \text{ to } j) \times \left(1 - \ln \frac{\#\text{references}(g \text{ to } j)}{\#\text{references}(g)}\right)}{\#\text{publications}(j) \times (1 + \ln \#\text{references}(g))}
\]

The formula considers the number of references from a guideline \(g\) to a journal \(j\) in relation to the overall number of references in the journal. Moreover, the denominator takes the total number of citable publications of a journal into account. The publication window of citable items is 10 years and ranges from 2010 to 2019.

---

1 Available at https://github.com/axa-group/Parsr/.
3. Results

3.1 Descriptive results of the cited references in guidelines

A total of 25,104 references from 29,501 guideline references were matched in WoS (85.1%). The number of references in General Medicine guidelines is 3,214, of which 2,328 were matched (72.4%). The share of references in Oncology guidelines matched is higher with 22,776 out of 26,287 references (86.6%). With an average of 321 references per guideline (ranging from 123 to 590) General Medicine guidelines have much shorter reference lists than Oncology guidelines, which have an average of 848 references per guideline (ranging from 268 to 1,688). Reasons for the much longer reference lists in Oncology guidelines may include, among others, more existing literature on the subject and the broader treatment of the disease in the guideline. Figure 1 outlines the distribution of document types among the cited references in WoS. Articles constitute the majority, followed by Reviews, Proceedings and Editorial Material. Others includes document types such as Meeting Abstracts, Letters, Corrections and News Items. Interestingly, reviews are cited more often in General Medicine guidelines than in Oncology guidelines.

![Document types cited in medical guideline](image)

**Figure 1.** Document type of cited references in General Medicine and Oncology.

The analysis of the publication year of the cited references in WoS shows that the differences in the age of the referenced literature are marginal. References in General Medicine guidelines have a median publication year of 2011 (avg. 2009), whereas references in Oncology guidelines have on average a median publication year of 2010 (avg. 2009).

3.2 JIF of cited references

We decided to take the JIF from 2019 as the most up-to-date and reliable JIF as the JIF of some medical journals heavily increased from 2020 on due to publications related to the Covid-19 pandemic (extreme examples include the BMJ, JAMA and The Lancet). With an average of 12.3 the JIF of cited references in Oncology guidelines is slightly higher than in the guidelines on General Medicine (10.6). Figure 2 shows the cumulative distribution of JIF of references cited in General Medicine and Oncology.
Figure 2. Cumulative distribution of JIF of references cited in General Medicine and Oncology.

The median JIF of the cited references in General Medicine is roughly 4.5, whereas the median JIF of the cited references in Oncology is about 5.5.

3.3 GLIF

In order to compare both fields, the GLIF was calculated separately for General Medicine guidelines and Oncology guidelines. Figure 3 is a bubble chart where each bubble represents a journal cited in General Medicine guidelines. The x-axis shows the journal’s JIF in 2019 and the y-axis the GLIF. The bubbles are sized according to the total number of total references to the journal in question. Only journals that were cited at least twice are displayed.

Figure 3. GLIF and JIF (2019) in General Medicine.

The figure shows that the most often cited journal with a total of 142 references is the Canadian Medical Association Journal, an open-access general medicine journal publishing clinical
research, reviews, and practice updates. It is followed by Lancet, JAMA and BMJ. We also see that German journals with a low JIF such as Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, Zeitschrift für Gerontologie und Geriatrie, Deutsches Ärzteblatt International or Pflege turn out to be important for German medical guidelines represented by their high GLIF. Kendall rank correlation coefficient shows that the observations have a dissimilar rank between the two variables GLIF and JIF (Kendall’s τ = -0.027).

Figure 4 illustrates the bubble chart for journals in Oncology guidelines. Note that the legend in the upper-right corner is different from that in Figure 3 and that the x-axis is interrupted due to the outlier journal CA-A Cancer Journal for Clinicians that boasts a JIF of roughly 293 in 2019.

![Guideline Impact Factor (GLIF) and Journal Impact Factor (JIF) in Oncology](image)

**Figure 4.** GLIF and JIF (2019) in General Medicine.

The most often cited journal is the Journal of Clinical Oncology with a total of 1,994 cited references, followed by Cancer (635) and Annals of Oncology (614). Again, we have several journals like Strahlentherapie und Onkologie, Pathologie, and Onkologie that have a low JIF but a high GLIF. Kendall rank correlation coefficient (Kendall’s τ = 0.019) shows that there is almost no correlation between the two variables GLIF and JIF.

Based on Figure 3 and 4, the GLIF indicator suggests that authors of medical guidelines scan through various sources, irrespective of JIF, to arrive at a comprehensive review of the literature.

4. Discussion

Our study suggests that the JIF is not a selection criterion for medical guideline authors and therefore not an appropriate indicator to assess the clinical relevance of medical research. In contrast, the Guideline Impact Factor (GLIF) we propose in this paper provides insight into which research outlets affect patient care and gives a better indication of the clinical impact of journals. The GLIF could be hence used as a complementary indicator in evaluation contexts.

The strength of this analysis lies in the full coverage of German high-quality medical guidelines from six successive years (2017-2022) in two different medical fields. Of course,
this study is not without limitations. One major issue in bibliometric studies is the accuracy of data. While on the one hand, we were able to retrieve a large number of references, the completeness and accuracy of them leaves much to be desired (see also Traylor & Herrmann-Lingen, 2023). As mentioned in Section 2.2 we had difficulties matching German language publications to their WoS counterparts. Combined with the general coverage of the WoS database, this likely biases our results towards English language research. The guidelines also make use of other grey literature published by public bodies, professional institutions or insurers which are not present in WoS. Analysing the references not matched in WoS would provide insight on the role of this body of literature in medical guidelines. This study is based on two fields of medicine in Germany and its results might be dependent on that context. A more comprehensive analysis could provide a better insight into the general reference patterns of medical guidelines and increase the reliability of the GLIF.

We agree with Eriksson et al. (2020) that research on medical guidelines would benefit from well-labelled, structured references which would also greatly ease the calculation of alternative impact measures like the GLIF. Initiatives aiming to create “digital guidelines” – like the one put forward by the GGPO – are still in their early days but show promise in this regard and might be an interesting data source in the future as they would not only simplify the collection of references but also enable researchers to take citation contexts into account.

5. Conclusion

In Germany as well as many other countries, the JIF plays a major role when evaluating medical research and researchers alike. Our results show that a large proportion of cited references in medical guidelines come from specialised journals with a relatively low JIF, suggesting that the JIF does not capture clinical relevance as required by guideline authors or put more strongly, the clinical relevance of a journal is independent of its JIF, at least for the guidelines we studied. This means that the GLIF can be interpreted as an independent quality criterion for journals. Arguably, the GLIF captures the societal impact of medical journals much better than (or at least captures an important dimension neglected by) the JIF by valuing articles that affect clinical practice and thus contribute to better health care. Due to the fact that guidelines reference much more research from the country the guideline is created for, the GLIF also helps to highlight locally relevant research.

It is important to stress that the GLIF is a journal-based measure and cannot give information about the clinical relevance of individual articles or authors. This is of course also the case for the JIF. Bracketing the discussion on the rationality of rewarding individuals based on journal-based metrics, it is not evident why publications in journals with a high GLIF should be valued less than ones with a high JIF. Giving articles published in journals with a GLIF recognition complementary to common journal impact measures like the JIF, would also rectify the lack of acknowledgement of guideline contributions under the common reward system in medicine (Herrmann-Lingen et al., 2014) and increase the valuation of the many journals relevant to clinical practice despite their low JIF. Therefore, we argue that information about a journal’s guideline impact should be included when evaluating medical research using journal-based indicators.
In summary, our study provides important insights into the characteristics of the literature referred to in guidelines as well as showing how solely focussing on the JIF in evaluations will systematically undervalue clinically important research.

**Open science practices**

This study focusses on the JIF and possible alternatives to it. Because evaluations based on JIF are usually using Clarivate’s Journal Citation Reports we decided to use the proprietary WoS database to make the results comparable and to show the feasibility of our approach. Except for the raw reference strings of the studied guidelines (which can be made available upon request) no new software or dataset was produced for this study.

**Acknowledgments**

The authors would like to thank Professor Dr. Herrmann-Lingen, Director of the Department of Psychosomatic Medicine and Psychotherapy at Göttingen University Medical Center, for his initial ideas on the Guideline Impact Factor.

**Competing interests**

There are no competing interests.

**Funding information**

The contribution is funded by the German Federal Ministry for Education and Research (BMBF) with grant number 16PU17025A. Data for WoS in this study were obtained from the German Competence Network for Bibliometrics, funded by the German Federal Ministry for Education and Research (BMBF) with grant number 01PQ17001.

**References**


