

STI 2018 Leiden

*23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"*

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at <https://hdl.handle.net/1887/64521>

ISBN: 978-90-9031204-0

© of the text: the authors

© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Quantifying the reproducibility of scientometric analyses: a case study¹

Manuel J. Cobo*, Tahereh Dehdarirad**, Pablo García-Sánchez* and Jose A. Moral-Munoz***

*** manueljesus.cobo@uca.es; pablo.garciasanchez@uca.es

Department of Computer Science and Engineering, University of Cádiz, Av. Ramon Puyol, s/n, Algeciras, Cádiz, 11202, (Spain)

** tahereh.dehdarirad@chalmers.se

Department of Communication and Learning in Science, Chalmers University of Technology, Hörsalsvägen 2, Göteborg, SE-412 96 (Sweden)

*** joseantonio.moral@uca.es

Department of Nursing and Physiotherapy, University of Cádiz, Avda. Ana de Viya, 52, Cádiz, 11009 (Spain) Institute of Research and Innovation in Biomedical Sciences of the Province of Cádiz (INiBICA), University of Cádiz, Avda. Ana de Viya, 21, Cádiz, 11009 (Spain)

Introduction

A key challenge in science is that information provided in scientific communications must be enough for other researchers in the research field to be able to reproduce the published findings (Stodden, Seiler, & Ma, 2018). That is, a scientific communication (i.e., article), should clearly describe all the steps taken.

Since 1665, when the *Philosophical Transactions of the Royal Society* was established, reproducibility has been used in science to assist the determination of the validity, legitimacy, transparency, importance, and strength of science (Allison, Shiffrin, & Stodden, 2018). This has also helped to improve the credibility of the scientific articles and accelerate new discoveries. Moreover, in recent years, the reproducibility of scientific findings has been a major concern in science, which has turned into a great debate (Allison et al., 2018; Fanelli, 2018; Maniadis & Tufano, 2017; Munafò et al., 2017). Results of some studies suggest that reproducibility of scientific findings is lower than desirable (Begley & Ioannidis, 2015; Glasziou et al., 2014). According to a recent survey published in *Nature*, 52% of those surveyed agree that there is a significant "crisis" of reproducibility (Baker, 2016). Although the term "reproducibility crisis" is not appropriate to describe the current state of science (Baker, 2016), as indicated by Munafò et al. (2017) more effort could be done to improve research practices in order to maximize the efficiency of the research community.

Whilst most replication efforts have focused on biomedicine, health, and psychology (Maniadis & Tufano, 2017), a recent survey of over 1,500 scientists, from various fields, suggests that the problem is widespread (Baker, 2016).

¹ This work was supported by FEDER funds (TIN2016-75850-R)

Reproducibility concerns in one field cannot be directly translated to other fields, as each research field has its own set of reproducibility artifacts (i.e. elements used in the experiment needed to reproduce again the same results). For example, in computational research, digital scholarly artifacts such as datasets and codes are an integral part of it (Stodden et al., 2018), and hence, the research community requires the dissemination of these digital artifacts.

Accordingly, some journals such as Science require authors to make codes and data available. Furthermore, data repositories, such as figshare.com and dryad.org, and software repositories, such as github.com, exist for reproducibility purposes. Although these are good initiatives, they are not well spread across journals and very few of them explicitly require this information to be made available.

In scientometrics studies (Batagelj & Cerinšek, 2013; Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011; Mane & Börner, 2004), datasets are obtained from bibliographic databases (e.g., Web of Science, Scopus, etc.), or they are created from private data sources. Thus, in scientometrics, the procedure used to construct the dataset is an important artifact which must be clearly described. Otherwise it would be impossible for other researchers to reproduce the results. Other reproducibility artifacts in scientometrics are related to data analysis, which may require sharing of data and computer code (Waltman, 2017).

Thus, the main aim of this study is to quantify the reproducibility of a sample of scientometric studies by examining the availability of different artifacts. To do this, an empirical evaluation of a set of 285 articles published in the journal *Scientometrics* in 2017 was carried out. This provides us with a good perspective on the degree of reproducibility in the field of scientometrics. As indicated by Glänzel (1996), the reproducibility of results in scientometrics can only be guaranteed if all sources, procedures and techniques are properly documented in scientific publications. To achieve the aim of the contribution, the following questions are addressed with regards to the studied dataset:

1. Which and to what extent are the reproducibility artifacts available?
2. Which type of database has been used?
3. What databases are used regularly?
4. How many databases are used?
5. Which type of software is used regularly?
6. Are the codes used for data analysis available where ad-hoc software is used?

This contribution is organized as follows. The *Methodology section* describes the methodology followed to measure the reproducibility of articles. The *Dataset section* describes the corpus used in the analysis. The *Results section* describes the main findings. Finally, some conclusions are drawn in the *Conclusions section*.

Methodology

In order to analyse the reproducibility of scientometric studies, a specific set of steps have been taken. In general terms, our methodology is based on three tasks: i) dataset acquisition (collection), ii) definition of the reproducibility measures, and iii) evaluation of the corpus.

Dataset acquisition (collection)

For the purpose of this paper, we used a set of papers that met one of the following criteria: i) focused on science of science analysis, such as, bibliometric, scientometric or science mapping analysis, and ii) papers with a theoretical component but containing a practical example.

According to the above-mentioned criteria we selected a set of articles published in the journal *Scientometrics* in 2017. *Scientometrics* has been selected as it is one of the most important journals in the research field, and it publishes a wide variety of practical, and theoretical studies where a dataset is employed.

We downloaded 375 documents from the *Scientometrics*' journal website. Then, a manual checking was carried out to discard those articles that did not meet the criteria selection described above. The reason is that the journal also publishes pure theoretical articles or articles that could not be reproducible due to their intrinsic characteristic. See for example Rhaïem (2017).

As a result, a total of 285 documents met the criteria and therefore were employed to determine their reproducibility characteristics.

To be able to determine if the experiments or studies carried out were reproducible, the full text of the document were analysed.

Definition of reproducibility criteria

In order to measure the degree of reproducibility of the corpus obtained in the first step, we need to study the key artifacts employed in these documents. Therefore, the corpus was evaluated according to the following key aspects:

- **Workflow.** If the methodology employed (or the steps followed) to analyse the data was clearly defined.
- **Search strategy.** The search made in a bibliographic database or the steps followed to retrieve the data. We checked if the selection strategy of the data was clearly described, and therefore there was no doubt in how the dataset was created and retrieved. This is an important key element in the reproducibility of scientometrics studies. So, a special care must be taken. The search strategy could be defined as a query used in the bibliographical database to retrieve the data (e.g., TS="co-words"), or by a complete textual description. If the search strategy was described vaguely (e.g., "we used some keywords such as, term1 and term2, among others"), or it was not possible to completely reproduce it, we considered the search strategy was not reproducible. In contrast, the search strategy was considered to be reproducible if there was enough information in the paper to be able to obtain the same dataset again. Regarding search strategy, we examined other key elements: time period used, and types of documents (e.g., articles, review, book chapters, etc.).
- **Number of units/cases analysed.** It could be the size of the corpus, the number of authors evaluated, the number of journals, etc. Number of units is important to assure that the same volume of data is evaluated when the research will be reproduced by other researchers.

- **Database.** Usually, the data is obtained from a bibliographic database. The database could be accessible publicly (e.g. arXiv, crossref, open citation, etc) or accessible via a licensed database (e.g., Web of Science, Scopus, patent data, etc.). Furthermore, sometimes the dataset has been made ad-hoc from private sources or databases. Thus, we examined the type of database (e.g., open, licensing or private) used in the analysis.
- **Software.** The tool used to perform the analysis plays a key role in reproducibility. Thus, the type of tool used was examined. We categorized tools into: open (if the software could be used freely, independently from the availability of its source code), licensing, (if the software was available only under license), or ad-hoc, (if the software was specifically developed for the study). In the last case, we also determined the availability of the source code.
- **Dataset.** If the query is well defined and the database is available, other researchers could retrieve the data and perform again the analysis. But sometimes, researchers do not have access to the database, and therefore, although the search strategy were well defined they can not retrieve the data. Although most of the datasets will be based on licensing sources, authors could provide the dataset used in the analysis, without breaking the license terms. For example, authors could provide the DOIs of the documents used, or some initial version of the data employed.

Evaluation of the corpus

The final step in the methodology was the evaluation of the entire dataset to determine the degree of reproducibility according to the criteria described above. Since an automatic way could not be applied, a manual evaluation was done by the authors, by checking the documents.

Results

In this section findings regarding the studied questions are presented. The dataset used in this contribution is available via <http://dx.doi.org/10.6084/m9.figshare.6137501>. It contains the DOIs of the documents evaluated and the results for the different artifacts examined.

Q1. Which and to what extent are the reproducibility artifacts available?

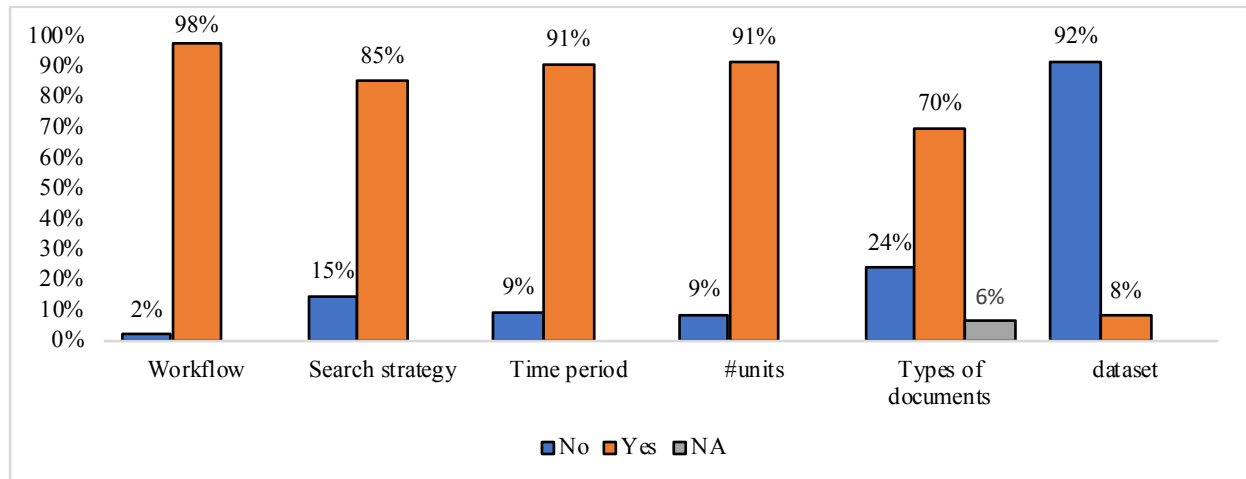
Figure 1 shows the availability of different reproducibility artifacts in the analysed dataset and their corresponding percentage. As can be seen in the figure, 98% of the studies described the workflow and the methodology applied. Although this is a large percentage, we should take into account that we set as reproducibility workflow if in the paper, a methodology is proposed defining the steps needed to perform the analysis. Maybe, if we consider a large corpus, this percentage could decrease, but for Scientoemtrics' papers, the workflow is commonly described.

On the other hand, the search strategy was described in 85% of the articles, not all of them provided a very clear description of the strategy for reproducibility purposes. Also, time period is an important aspect regarding search strategy, being mentioned in the majority of the articles.

Whilst an important key aspect in the reproducibility of scientometric studies is the availability of the dataset, the majority (92%) of the studied articles did not share the dataset

used. Although the search strategy is clearly described in the articles, sometimes, due to the license restriction, it will be impossible for other researchers to access the same data. Sometimes even with a valid licence and the search strategy clearly defined, it could also be impossible to retrieve exactly the same data due to the continuously update of the databases. For the same reason, it is important to publish the dataset used.

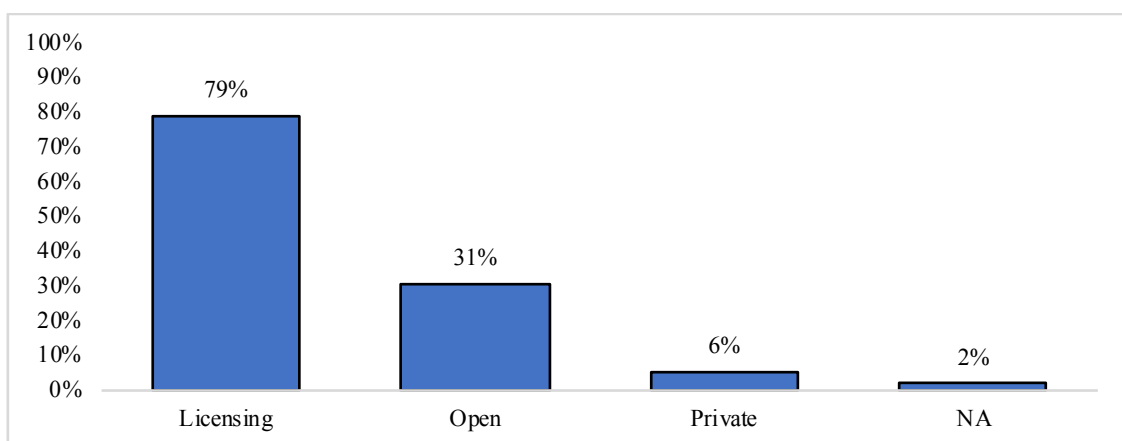
Figure 1. Availability of the reproducibility artifacts and their corresponding percentage for articles published in Scientometrics in 2017.



Q 2. Which type of database has been used?

Figure 2 shows the percentage of database types (open/free, licensing or private) used in the studied set. As can be seen in the figure, while most of the articles used a licensed database such as the WoS, a representative number of articles used an open database, such as DBLP or Google Scholar. Only very few articles used private sources.

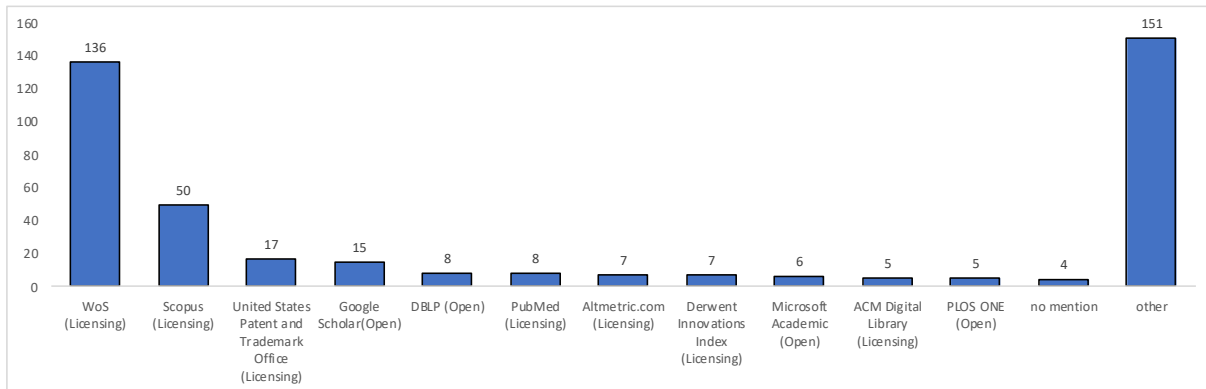
Figure 2. Type of database used in articles published in Scientometrics in 2017 and their corresponding percentage.



Q3. What databases are regularly used?

The most frequently used databases in the studied set are shown in Figure 3. As can be seen in the figure, most of the studies used the WoS and Scopus databases. Also, a large number of studies (*Other*) relied on other data sources such as the national (local) ones (Russian Science Citation Index (RSCI), National Bureau of Statistics and Ministry of Science and Technology of China).

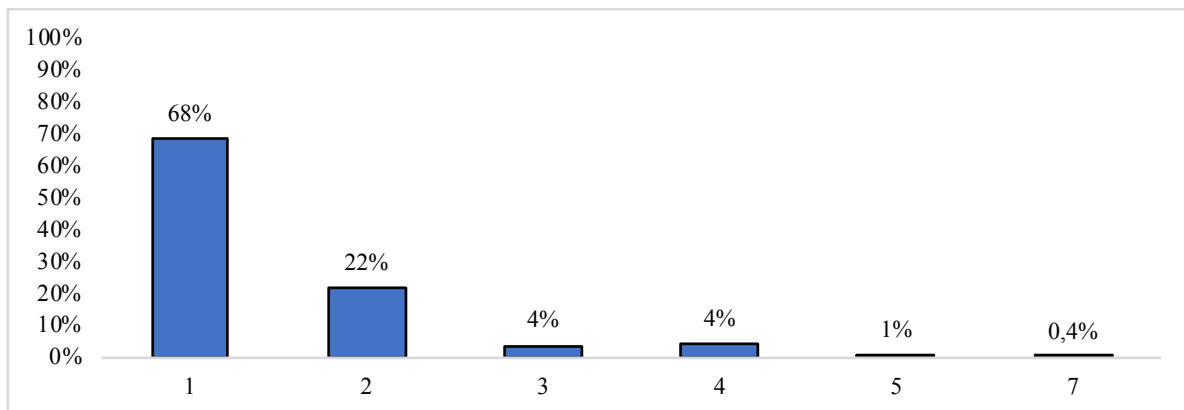
Figure 3. The Most frequently used databases for articles published in Scientometrics in 2017.



Q4. How many databases are used?

Figure 4 shows the number of databases being used in the studied set. Although the majority of articles (68%) used only one database, 31% of the articles used at least two databases. This means that an article could use different types of databases at the same time.

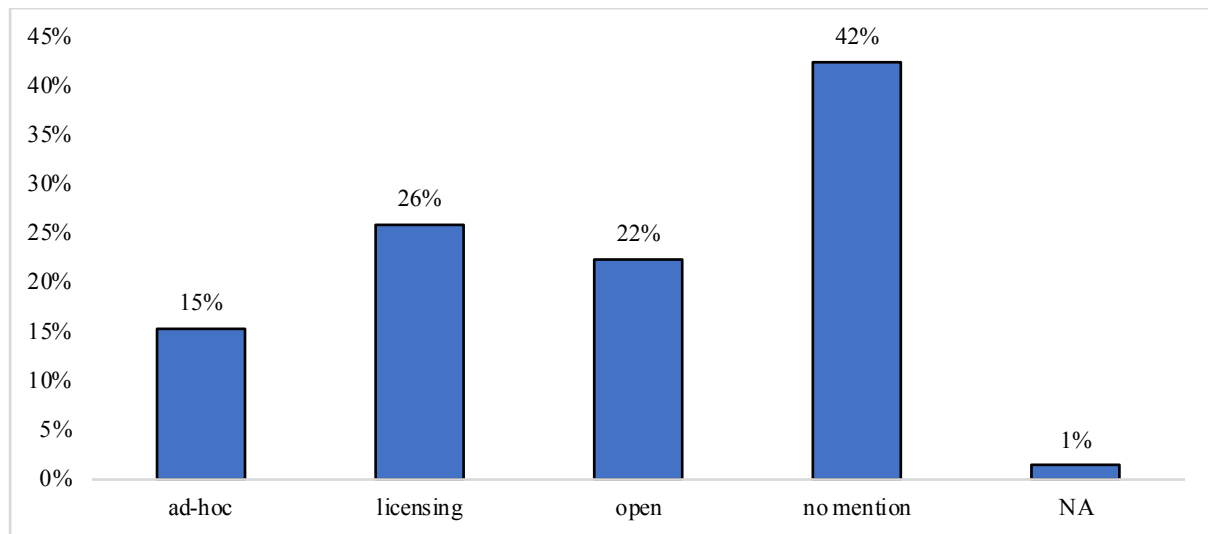
Figure 4. Number of databases used in articles published in Scientometrics in 2017.



Q5. Which type of software is used regularly?

The availability of the tool used is another important aspect with regard to the reproducibility of scientific articles. Figure 5 shows the type of software regularly used in the studied set. As can be seen, 42% of the studies did not mention the software used. Although it is true that some studies could be done using generic software, for reproducibility purpose, it would be nice if the software employed is at least mentioned. As an example of a paper that obviously use some social network tool, but not mentioned it, see

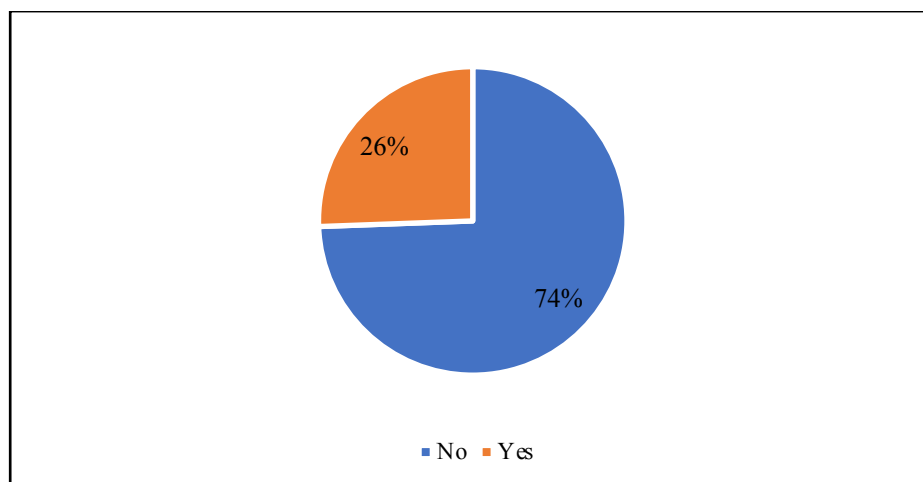
Figure 5. Types of software regularly used in articles published in *Scientometrics* in 2017.



Q6. Are the codes used for data analysis available where ad-hoc software is used?

If a specific or ad-hoc tool was developed to perform the analysis (using R, Python, Matlab or other programming languages), the availability of the source code should be mandatory to reproduce the analysis. Thus, we examined the availability of the source code where ad-hoc software was used. Figure 6 shows the percentage of studies that published the source code when an ad-hoc software was used. As can be seen, most of the studies did not make the source code available, which could make the reproducibility of findings a daunting or an impossible task.

Figure 6. The availability percentage of source code for articles published in *Scientometrics* in 2017.



Conclusions

This contribution examined the reproducibility of a sample of articles published in the journal of *Scientometric* in 2017 by examining the availability of different artifacts.

These artifacts were workflow, search strategy, database, software, the availability of the source code (where applicable) and the availability of the dataset.

Some main finding could be highlighted. Firstly, our data showed that the workflow and the search strategy used were well described in the majority of the examined articles. Nevertheless, the dataset was available only in a few studies. Regarding the most regular used database, our findings showed that whilst 44% of the articles used licensing databases such as Scopus or the WoS, 36% of articles used other databases such as the ones for specific purposes or the national (local) ones. Furthermore, very few articles (26%) made the source code available where ad-hoc software was used.

Considering an aggregation of the reproducibility criterion, we could conclude that only 22 (7,7%) articles could be considered as reproducible, since they define: workflow, search strategy, and dataset. If we also consider the software or source code, the percentage decrease significantly.

Finally, it is important to consider that our study was limited to a small set of scientometrics articles published in the journal of *Scientometrics* in a year. In order to provide a comprehensive perspective on reproducibility in the field of scientometrics a bigger corpus must be examined.

References

- Allison, D. B., Shiffrin, R. M., & Stodden, V. (2018). Reproducibility of research: Issues and proposed remedies. *Proceedings of the National Academy of Sciences*, *115*(11), 2561–2562. <https://doi.org/10.1073/pnas.1802324115>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. <https://doi.org/10.1038/533452a>
- Batagelj, V., & Cerinšek, M. (2013). On bibliographic networks. *Scientometrics*, *96*(3), 845–864. <https://doi.org/10.1007/s11192-012-0940-1>
- Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. *Circulation Research*, *116*(1), 116–126. <https://doi.org/10.1161/CIRCRESAHA.114.303819>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, *62*(7), 1382–1402. <https://doi.org/10.1002/asi.21525>
- Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, *115*(11), 2628–2631. <https://doi.org/10.1073/pnas.1708272114>
- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. <https://doi.org/10.1126/science.aao0185>

- Glänzel, W. (1996). The need for standards in bibliometric research and technology. *Scientometrics*, 35(2), 167-176. doi:10.1007/bf02018475
- Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., ... Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, 383(9913), 267–276. [https://doi.org/10.1016/S0140-6736\(13\)62228-X](https://doi.org/10.1016/S0140-6736(13)62228-X)
- Mane, K. K., & Börner, K. (2004). Mapping topics and topic bursts in PNAS. *Proceedings of the National Academy of Sciences*, 101(Supplement 1), 5287–5290. <https://doi.org/10.1073/pnas.0307626100>
- Maniadis, Z., & Tufano, F. (2017). The Research Reproducibility Crisis and Economics of Science. *The Economic Journal*, 127(605), F200–F208. <https://doi.org/10.1111/ecoj.12526>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 21. <https://doi.org/10.1038/s41562-016-0021>
- Rhaiem, M. (2017). Measurement and determinants of academic research efficiency: a systematic review of the evidence. *Scientometrics*, 110(2), 581–615. <https://doi.org/10.1007/s11192-016-2173-1>
- Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11), 2584–2589. <https://doi.org/10.1073/pnas.1708290115>
- Waltman, L. (2017). Reproducibility concerns in scientometrics differ from other fields. "Reproducible Scientometrics Research: Open Data, Code, and Education" (ISSI2017), Wuhan, China.
- Zhang, J., & Guan, J. (2017). Scientific relatedness and intellectual base: a citation analysis of un-cited and highly-cited papers in the solar energy field. *Scientometrics*, 110(1), 141–162. <https://doi.org/10.1007/s11192-016-2155-3>