

# Big Data and Complex Data Analytics: Breaking Peer Review?

F. Schwendicke<sup>1</sup> , M.L. Marazita<sup>2</sup>, N.S. Jakubovics<sup>3</sup>, and J. Krois<sup>1</sup> 

Journal of Dental Research  
2022, Vol. 101(4) 369–370  
© International Association for Dental Research and American Association for Dental, Oral, and Craniofacial Research 2022



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/00220345211070983  
journals.sagepub.com/home/jdr

**Keywords:** artificial intelligence, big data, bioinformatics, biostatistics, deep learning/machine learning, dental informatics/bioinformatics

With the advent of the data era, dental, oral, and craniofacial research is driven more and more by data (Schwendicke and Krois 2022). For instance, the *Journal of Dental Research* receives a continuously growing number of submissions dealing with large data sets, including routinely collected text records, images, or claims, or prospectively collected large biological data sets. In addition, many submissions focus on the complex analysis of these data sets—for example, involving shallow and deep machine learning. Considering that employing big data and analyzing them via machine learning has been relevant in medical research for several years, this is encouraging; dental, oral and craniofacial research is joining the vanguard in these innovative research strategies!

However, both the underlying data sets and the involved analytic methods bear a number of problems when considering how research is traditionally appraised—that is, via peer review. First, the employed data sets are in many cases not available for reanalysis. Data protection or intellectual property considerations limit the option to reevaluate the integrity, fairness, or generalizability of data. This, however, comes with significant and nondetectable risks of bias, for example. Second, the analytic strategies employed for leveraging these large data sets are usually complex, involving a combination of open access libraries and individual code. While the capacity of our field to adopt contemporary programming languages and use them for advanced analytic purposes is to be commended, it again poses a range of problems: the employed packages are not always stringently validated, nor is the code necessary to combine them and crunch the data regularly accessible, oftentimes again due to reasons of intellectual property. Hence, as a result, meaningful replication is often impossible; thus, faults and biases remain undetected. Third, even assuming that data and code were accessible, appropriate peer review is extremely difficult to arrange: reviewers with sufficient data science skills to appraise such studies remain a scarce resource in most journals' editorial boards. Health data science is a young discipline, but also the methods are advancing at a fast pace, reducing the number of individuals available for peer review to a vanishingly small number of those who professionally keep up with advances in data fields and are familiar with dentistry. In addition, systematic and comprehensive appraisal of data and code is a task of days or even weeks, often requiring replication of the analytic environment before

being able to implement and scrutinize the analyses. It is unlikely in most cases that reviewers will regularly spend that much time reviewing articles—currently at very limited incentives and/or at the expense of the institution funding them.

A number of possible options to address these problems are conceivable:

- *Enforce open data and open code.* This is attractive as it is immediately applicable. It is not expected that open data and code necessarily improve peer review but strengthen the option of postacceptance review and replication by peers. This could well mean a period of more corrections or even retractions in the first years—and journals should consider improved mechanisms for collating postpublication updates—but should lead to more rigorous and transparent analyses given the expected scrutiny after publication. In addition, the potential to reuse large data sets or integrate them into meta-analyses provides added value to data sets that are often difficult or costly to acquire. For maximum benefit, the data must be well formatted and well described with appropriate metadata. The International Committee of Medical Journal Editors now requires data-sharing statements to be included in manuscripts in member journals that report the results of clinical trials (Taichman et al. 2017).
- *Realign incentives for peer review, allowing for more time and systematic reviewing efforts.* 1) Paying reviewers would at least solve the asymmetry of private publication companies benefiting from review and the

<sup>1</sup>Department of Oral Diagnostics, Digital Health, Health Services Research, Charité–Universitätsmedizin Berlin, Berlin, Germany

<sup>2</sup>Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, School of Dental Medicine, and Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

<sup>3</sup>School of Dental Sciences, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

#### Corresponding Author:

F. Schwendicke, Department of Oral Diagnostics, Digital Health, Health Services Research, Charité–Universitätsmedizin Berlin, Aßmannshäuser Str. 4-6, 14197, Berlin, Germany.  
Email: falk.schwendicke@charite.de

public paying for it. However, there are many issues with cash rewards for peer review: the difficulty of evaluating peer review and linking remuneration to the depth and quality of the review, the problems associated with organizing contracts for peer reviewers, the potential for payments to incentivize acceptance, and the increased costs of the publication process (Brainard 2021). 2) Incentivize peer review activities institutionally—for example, membership in an editorial board or the number of reviewed articles are reflected by institutions when providing intramural funding but also during career development. Here, systems to document peer review, such as Clarivate's Publons platform, can help to keep track of peer review activity.

- *Establish standards and norms for data and advanced data analyses in dentistry onto which to build.* These would be analogous to the well-established standards available in areas such as materials testing (Schmalz et al. 2021). These would also provide guidance on conducting and presenting data-driven research, assist authors and reviewers alike, and are likely to stepwise raise the bar. Some early examples are available, such as the CONSORT-AI extension for conducting and reporting clinical trials involving artificial intelligence (Schwendicke and Krois 2021).
- *Push for authors to make data and code testable, even if not accessible.* This could occur, for example, by implementing machine learning models via modern application programming interfaces, dashboarding approaches, or other minimum viable software products. These could be tested by reviewers, even subscribers, or could be made freely available after publication.
- *Engage additional reviewers.* In cases where manuscripts are highly interdisciplinary, it may be necessary to engage more reviewers to provide expertise in specific aspects of the article (Ghosh et al. 2012). However, even if reviewers are asked to focus on specific areas, the increased number of peer review requests could quickly lead to reviewer fatigue.

Clearly, there are many challenges to the peer review of articles based on large and complex data sets or sophisticated analytic tools. Nevertheless, it is essential that we evaluate data-driven manuscripts effectively to realize the full potential

of this field. The forthcoming special issue of the *Journal of Dental Research* entitled “Data-Driven Analytics for Dental, Oral, and Craniofacial Health Care,” will give us an opportunity to test our current peer review processes and learn for the future. We look forward to the presentation of outstanding science in this exciting and rapidly growing area of research and resources.

### Author Contributions

F. Schwendicke, contributed to data conception and design, drafted and critically revised the manuscript; M.L. Marazita, N.S. Jakubovics, J. Krois, contributed to data acquisition, analysis, or interpretation of data, critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.



### Declaration of Conflicting Interests

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

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

F. Schwendicke  <https://orcid.org/0000-0003-1223-1669>  
J. Krois  <https://orcid.org/0000-0002-6010-8940>

### References

- Brainard J. 2021. The \$450 question: should journals pay peer reviewers? ScienceInsider [accessed 2021 Nov 30]. <https://www.science.org/content/article/450-question-should-journals-pay-peer-reviewers>
- Ghosh SS, Klein A, Avants B, Millman KJ. 2012. Learning from open source software projects to improve scientific review. *Front Comput Neurosci.* 6:18.
- Schmalz G, Jakubovics N, Schwendicke F. 2021. Normative approaches for oral health: standards, specifications, and guidelines. *J Dent Res* [pub ahead of print 25 Oct 2021] in press. doi:10.1177/00220345211049695
- Schwendicke F, Krois J. 2021. Better reporting of studies on artificial intelligence: CONSORT-AI and beyond. *J Dent Res.* 100(7):677–680.
- Schwendicke F, Krois J. 2022. Data dentistry: how data are changing clinical care and research. *J Dent Res.* 101(1):21–29. doi:10.1177/00220345211020265
- Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, Hong ST, Haileamlak A, Gollogly L, Godlee F, et al. 2017. Data sharing statements for clinical trials: a requirement of the international committee of medical journal editors. *Ann Intern Med.* 167(1):63–65.