Invited Editorial

Data-Driven Dental, Oral, and Craniofacial Analytics: Here to Stay

Journal of Dental Research
2022, Vol. 101(11) 1255–1257

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00220345221120564
journals.sagepub.com/home/jdr

F. Schwendicke[1] and M.L. Marazita[2]

Keywords: artificial intelligence, big data, bioinformatics, biostatistics, deep learning/machine learning, data science

Dental, oral, and craniofacial (DOC) health care and research are generating an increasing amount of data that are expected to be used to foster a deeper understanding of our patients' health and disease, thus allowing more effective, efficient, and safer care. Some even postulate that data and data analytics may shape something coined "data dentistry," with data-driven approaches and applications disseminating widely and deeply into DOC research and practice (Schwendicke and Krois 2021). Data are generated not only prospectively (e.g., in scientific studies)—with such prospectively collected data being increasingly large (e.g., omics or imagery data sets)—but also during clinical routine, involving history data, clinical records, images, specific tests, imaging, prescriptions, and so forth.

Big Data and data analytics in all its forms, including artificial intelligence (AI), and machine learning, have been successfully used in medicine and dentistry in recent years. The *Journal of Dental Research* has received an ever-increasing number of submissions involving such analytical strategies both for research and for clinical applications. This special issue of the *Journal of Dental Research* focuses on such aspects of DOC Big Data and advanced data analytics, aiming to display not only the breadth of data and analytical strategies currently used in DOC, their translational efforts, and their promised impact but also the challenges and current difficulties in the field as applied to DOC.

The explosion of truly Big Data in biology and health care began with the Human Genome Project (https://www.genome.gov/human-genome-project), an international initiative spearheaded by public-private research partnerships, with the goal of mapping the entire human genome (Watson 1990; Gibbs 2020). The project was envisioned in the 1980s, took off in the 1990s (Watson 1990), was considered finished in the early 2000s despite many gaps remaining (Lander et al. 2001), with a final complete assembly published in 2022 (Nurk et al. 2022). The data and tools developed from the Human Genome Project are publicly available, notably through the National Center for Biotechnology Information (https://www.ncbi.nih.gov/genome/guide/human/). The Human Genome Project remains the largest collaborative biological project in human history; it has spawned countless offshoots and has made possible today's initiatives in precision health care and genomic research, including in DOC (Marazita 2012).

To fulfill national and international goals in precision health care, there have been concomitant data collection efforts over many years to collect and make available data from population- or cohort-based electronic health records, coupled with genomic, metabolomic, and many more "omics" to allow extremely large-scale projects investigating the biological and epidemiological underpinnings of human health and disease conditions. A relatively early example of such resources is the UK Biobank established in the early 2000s (www.ukbiobank.ac.uk), which contains accessible in-depth genetic and health information—including dental (Galloway 2011)—from half a million UK participants. More recently, the US has launched the All of Us project (allofus.nih.gov), with the goal of enrolling 1 million participants with electronic health records, imaging, and whole-genome sequencing. To date, All of Us has enrolled about 330,000 participants. Many similar projects are ongoing around the world, and 1 article in this special issue (Divaris et al. 2022) includes a description of a consortium of such projects focused on DOC conditions. Similarly, other authors of articles in this special issue summarize existing and available large biological data resources, for example, the online data repository FaceBase (facebase.org; Schuler et al. 2022) and the developing Human Oral and Craniofacial Cell Atlas (Caetano et al. 2022), and a new tool to visualize the gene and miRNA expression patterns across developmental stages developed from CleftGenDB data as well as regulatory networks (Yan et al. 2022).

Using these large, structured or unstructured data sets of different modalities, ideally also recorded repeatedly over time, is challenging, but undoubtedly worthwhile. Traditional analytic tools may not fully leverage the wealth of information in these data sets or may not even be able to work with such Big Data at all. To use such data fully and appropriately, there has been a concomitant explosion in data analytics, notably

[1]Department of Oral Diagnostics, Digital Health, Health Services Research, Charité – Universitätsmedizin, Berlin, Germany
[2]Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, School of Dental Medicine, and Department of Human Genetics, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

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

including the application of AI. This special issue highlights the application of AI for developing clinically relevant tools, for example, evaluating factors contributing to the long-term survival of root canal treatments using data from the National Dental Practice Based Research Network (Thyvalikakath et al. 2022), the diagnosis of oral squamous cell carcinoma (Yang et al. 2022), automated 3-dimensional (3D) cephalometric landmarking (Dot et al. 2022), the prediction of 3D post-orthodontia facial changes in adults (Park et al. 2022), the predictive classification of tooth removal procedures (de Graaf et al. 2022), and the association of components of computer-aided design/computer-aided manufacturing resins with clinical outcomes (Li et al. 2022).

In addition to the development of clinical tools and assessments, other articles in this special issue address the application of analytical strategies to Big Data for research purposes in DOC science, for example, evaluating the impact of early childhood oral health on later academic performance (Wehby 2022); B/plasma cell infiltration in periodontitis, which may lead to a molecular diagnosis (Huang et al. 2022); novel AI network analyses to identify multimorbidity clusters in people with periodontitis (Larvin et al. 2022); and the value of information and economics around AI for caries detection (Schwendicke, Cejudo Grano de Oro, et al. 2022).

Harnessing Big Data presents challenges beyond analytic pathways and methods: data may not be accessible at all or, they are if accessible, may not conform to current FAIR (Findable, Accessible, Interoperable, and Reusable) principles, often being siloed due to data protection reasons or not usable given limited interoperability. Peer-reviewing research using such data, analyzed using the described complex (and often not fully explainable) methods, has been found to be challenging and may require new strategies for researchers and publishers alike (Schwendicke, Marazita, et al. 2022). Efforts toward accessing DOC data or using them, even if they are not directly accessible, are required, in concert with harmonizing data formats and analytical streams. The authors in this special issue address a variety of such issues, for example, difficulties in appropriately benchmarking different AI applications (Schneider et al. 2022), the lack of interoperability of available DOC data (Rajkumar et al. 2022), DOC research data availability and quality according to the FAIR principles (Uribe et al. 2022), harmonizing caries disease phenotypes for large-scale genomics studies (Divaris et al. 2022), federated learning in dentistry (Rischke et al. 2022), conceptualization of social epigenomics for oral health (Gomaa 2022), and trustworthy AI in DOC research (Ma et al. 2022).

DOC is uniquely positioned to embrace advanced data analytics, with a wealth of routinely collected, longitudinal data being complemented by a wide range of large prospectively collected data sets to allow deeper insights into individuals' health. This special issue displays the breadth of research in our domain along with its chances, achievements, and challenges. The *Journal of Dental Research* remains focused on fostering innovative, data-driven science in the field and to help unlock the potential of Big Data and advanced analytics for driving better DOC health care for all.

Author Contributions

F. Schwendicke, M.L. Marazita, contributed to conception, design, data acquisition, analysis, and interpretation, drafted and 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 author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

F. Schwendicke https://orcid.org/0000-0003-1223-1669

References

Caetano A, Kretzschmar K, Shazib M, Sharpe P, Kimple A, Opasawatchai A, Warner B, Krivanek J, McKay L, Freire M, et al. 2022. A roadmap for the human oral and craniofacial cell atlas. J Dent Res. 101(11):1274–1288.

de Graaf W, Riet T, Lange J, Kober J. 2022. A multiclass classification model for tooth removal procedures. J Dent Res. 101(11):1357–1362.

Divaris K, Haworth S, Shaffer J, Anttonen V, Beck J, Furuichi Y, Holtfreter B, Jönsson D, Kocher T, Levy S, et al. 2022. Phenotype harmonization in the GLIDE2 Oral Health Genomics Consortium. J Dent Res. 101(11):1408–1416.

Dot G, Schouman T, Chang S, Rafflenbeul F, Kerbrat A, Rouch P, Gajny L. 2022. Automatic 3-dimensional cephalometric landmarking via deep learning. J Dent Res. 101(11):1380–1387.

Galloway J. 2011. Putting the teeth into the UK Biobank. Prim Dent Care. 18(1):6–12.

Gibbs RA. 2020. The human genome project changed everything. Nat Rev Genet. 21(10):575–576.

Gomaa N. 2022. Social epigenomics: conceptualizations and considerations for oral health. J Dent Res. 101(11):1299–1306.

Huang D, Liu L, Chen Y, Wang L, Yang F, Li X, Luo S, Yang L, Wang T, Song D. 2022. Dissecting B/plasma cells in periodontitis at single-cell/bulk resolution. J Dent Res. 101(11):1388–1397.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409(6822):860–921.

Larvin H, Kang J, Aggarwal VR, Pavitt S, Wu J. 2022. Systemic multimorbidity clusters in people with periodontitis. J Dent Res. 101(11):1335–1342.

Li H, Sakai T, Tanaka A, Ogura M, Lee C, Yamaguchi S, Imazato S. 2022. Interpretable AI explores effective components of CAD/CAM resin composites. J Dent Res. 101(11):1363–1371.

Ma J, Schneider L, Lapuschkin S, Achtibat R, Duchrau M, Krois J, Schwendicke F, Samek W. 2022. Towards trustworthy ai in dentistry. J Dent Res. 101(11):1263–1268.

Marazita ML. 2012. The evolution of human genetic studies of cleft lip and cleft palate. Annu Rev Genomics Hum Genet. 13:263–283.

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. 2022. The complete sequence of a human genome. Science. 376(6588):44–53.

Park YS, Choi JH, Kim Y, Choi SH, Lee JH, Kim KH, Chung CJ. 2022. Deep learning-based prediction of the 3D postorthodontic facial changes. J Dent Res. 101(11):1372–1379.

Rajkumar NMR, Muzoora MR, Thun S. 2022. Dentistry and interoperability. J Dent Res. 101(11):1258–1262.

Rischke R, Schneider L, Müller K, Samek W, Schwendicke F, Krois J. 2022. Federated learning in dentistry: chances and challenges. J Dent Res. 101(11):1269–1373.

Schneider L, Arsiwala-Scheppach L, Krois J, Meyer-Lueckel H, Bressem KK, Niehues SM, Schwendicke F. 2022. Benchmarking deep learning models for tooth structure segmentation. J Dent Res. 101(11):1343–1349.

Schuler RE, Bugacov A, Hacia JG, Ho TV, Iwata J, Pearlman L, Samuels BD, Williams C, Zhao Z, Kesselman C, et al. 2022. FaceBase: a community-driven hub for data-intensive research. J Dent Res. 101(11):1289–1298.

Schwendicke F, Cejudo Grano de Oro J, Garcia Cantu A, Meyer-Lueckel H, Chaurasia A, Krois J. 2022. Artificial intelligence for caries detection: value of data and information. J Dent Res. 101(11):1350–1356.

Schwendicke F, Krois J. 2021. Data dentistry: how data are changing clinical care and research. J Dent Res. 101(1):21–29.

Schwendicke F, Marazita ML, Jakubovics NS, Krois J. 2022. Big data and complex data analytics: breaking peer review? J Dent Res. 101(4):369–370.

Thyvalikakath T, LaPradd M, Siddiqui Z, Duncan WD, Eckert G, Medam JK, Rindal DB, Jurkovich M, Gilbert GH. 2022. Root canal treatment survival analysis in national dental PBRN practices. J Dent Res. 101(11):1328–1334.

Uribe SE, Sofi-Mahmudi A, Raittio E, Maldupa I, Vilne B. 2022. Dental research data availability and quality according to the fair principles. J Dent Res. 101(11):1307–1313.

Watson JD. 1990. The human genome project: past, present, and future. Science. 248(4951):44–49.

Wehby GL. 2022. Oral health and academic achievement of children in low-income families. J Dent Res. 101(11):1314–1320.

Yan F, Simon LM, Suzuki A, Iwaya C, Jia P, Iwata J, Zhao Z. 2022. Spatiotemporal microRNA-gene expression network related to orofacial clefts. J Dent Res. 101(11):1398–1407.

Yang SY, Li SH, Liu JL, Sun XQ, Cen YY, Ren RY, Ying SC, Chen Y, Zhao ZH, Liao W. 2022. Histopathology-based diagnosis of oral squamous cell carcinoma using deep learning. J Dent Res. 101(11):1321–1327.