Data Dentistry: How Data Are Changing Clinical Care and Research

F. Schwendicke[1] 🆔 and J. Krois[1] 🆔

Abstract

Data are a key resource for modern societies and expected to improve quality, accessibility, affordability, safety, and equity of health care. Dental care and research are currently transforming into what we term *data dentistry*, with 3 main applications: 1) medical data analysis uses deep learning, allowing one to master unprecedented amounts of data (language, speech, imagery) and put them to productive use. 2) Data-enriched clinical care integrates data from individual (e.g., demographic, social, clinical and omics data, consumer data), setting (e.g., geospatial, environmental, provider-related data), and systems level (payer or regulatory data to characterize input, throughput, output, and outcomes of health care) to provide a comprehensive and continuous real-time assessment of biologic perturbations, individual behaviors, and context. Such care may contribute to a deeper understanding of health and disease and a more precise, personalized, predictive, and preventive care. 3) Data for research include open research data and data sharing, allowing one to appraise, benchmark, pool, replicate, and reuse data. Concerns and confidence into data-driven applications, stakeholders' and system's capabilities, and lack of data standardization and harmonization currently limit the development and implementation of data dentistry. Aspects of bias and data-user interaction require attention. Action items for the dental community circle around increasing data availability, refinement, and usage; demonstrating safety, value, and usefulness of applications; educating the dental workforce and consumers; providing performant and standardized infrastructure and processes; and incentivizing and adopting open data and data sharing.

Keywords: artificial intelligence, clinical studies/trials, computer vision, decision making, deep learning, personalized medicine

Introduction

In the 17th century, the experimental and theoretical sciences emerged; these are now considered the 2 basic research paradigms for understanding nature. In recent decades, computer simulations have become a third paradigm. Today, at the dawn of the Big Data Era, a fourth paradigm has emerged, focusing on data-intensive sciences (Bell et al. 2009). More and more industries embrace the data-driven paradigm and make better use of data to create value. Data are considered a key resource for striving modern societies and fulfill a set of characteristics to prompt a technological revolution (Perez 2002). Among others, these are the following: the costs of using data are low and decreasing over time, data are an inexhaustible resource as they can be used as many times as needed and by as many agents as technologically feasible at the same time, data are fundamental for a wide range of applications across industries and markets, and the smart usage of data increases the efficiency of processes and procedures (Klingenberg et al. 2019). While large technology companies identified the value of data already in the 1990s and 2000s, coining "data as the new oil" (Economist 2017), in health care, data have only recently been acknowledged as possibly facilitating a better, safer, more reliable, affordable, and accessible care. A number of data-driven technologies have entered health care in the past decade, for example, artificial intelligence (AI); sensors including wearables, ingestibles, and implantables; social media; and clinical data and electronic health records (eHR), to name a few. Nearly all technologies expected to reshape health care until 2030 are related to data (Fig. 1A).

Economic, demographic, and epidemiologic shifts put immense pressure on health care systems, with a growing and aging global population consuming an increasingly costly care (Peres et al. 2019). Concomitant to an expected increase in global health care expenditures to over $10 trillion by 2030 (Chang et al. 2019), significant shifts in expenditure toward diagnostics, prevention, precision medicine, and digital health are likely (Fig. 1B). Recent acquisitions of big technology companies in health care point into the same direction (Fig. 1C).

Data are seen as central to facilitate quality, accessibility, and equity of oral and dental care in the coming decades (World Health Organization 2020). The guiding principles of making better use of data to create value may, if rigorously applied, promote a shift in dentistry toward data-driven decision making and the dissemination of data-driven applications—a metamorphosis to a new state that we refer to as "data dentistry." This review focuses on 3 main areas of such data dentistry: 1) medical data analysis using deep learning, 2) data-enriched clinical care, and 3) data from and for dental research. These sections are followed by a brief summary of challenges to the adoption and advancement of data dentistry. We then conclude

[1]Department of Oral Diagnostics, Digital Health and Health Services Research, Charité–Universitätsmedizin Berlin, Berlin, Germany

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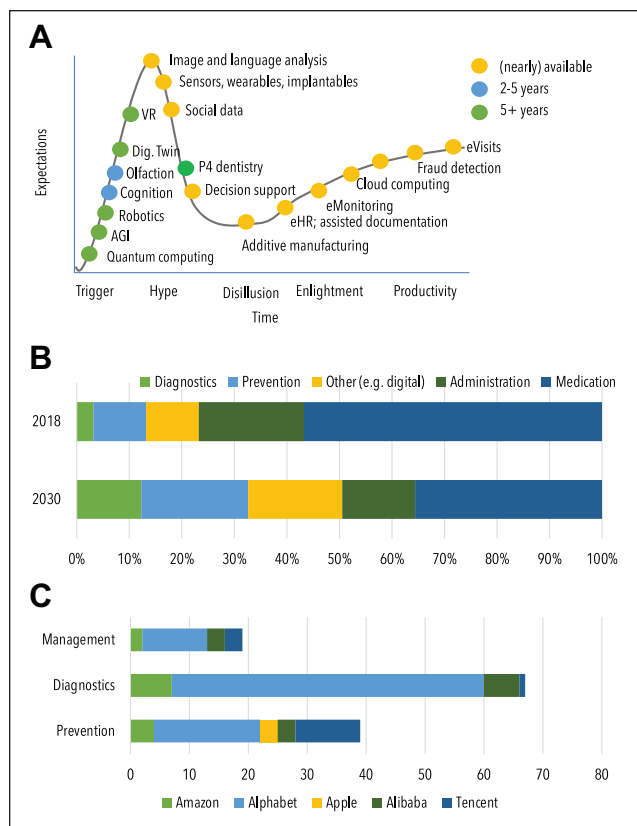Email: falk.schwendicke@charite.de

Figure 1. Data-driven or related technologies in dentistry come with significant changes in health care budgets and industry investment. (**A**) The Gartner Hype Cycle provides a representation of the "maturity of technologies and applications," which are classified according to their availability and located along the evolutionary stages from initial trigger and inflated expectations (hype) over possible disillusionment to increasing useful adoption and productivity. A range of data-driven or related technologies had been predicted in 2010 to be available by 2020, and most of them—located on the right side—are in productive use in dentistry by now. Others, predicted in 2020 to be available in the future, are rather concepts or in early application stages right now and will need to evolve both technically but also regarding their useful integration into the workflow. AGI, advanced generalized intelligence; eHR, electronic health record; P4, personalized, precise, preventive, and participatory; VR, virtual reality. (**B**) The composition of per-capita health care expenditures is expected to experience significant shifts. (**C**) International technology industry is increasingly investing in health care (annual investment was increased 12-fold since 2015); diagnostics and prevention are the main focus. Panels (B) and (C) are generated based on data from Solbach et al. (2019).

this review with action items for the dental and oral medicine community.

Medical Data Analysis Using Deep Learning

In 2013, the global health data volume was estimated to be 153 exabytes (1 exabyte = 10^{18} bytes). Seven years later, in 2020, the estimated volume was 2,314 exabytes; for comparison, the total global data volume in 2000 was 3 exabytes (Statista

2018). Machine learning (ML) is central when it comes to analyzing such large amounts of data and putting them to productive use. In ML, inherent statistical patterns in a data set are learnt by machines. Most ML applications employ supervised learning, where data points and data information (labels) are provided and used for training and iteratively improving the mapping of a datapoint and its label, eventually allowing to "self-label" new, unseen data. Deep learning (DL), a subfield of ML, has seen a dramatic surge over the past decade, driven by the increased availability of large data sets, powerful computational resources, and open-source software frameworks (LeCun et al. 2015). DL leverages artificial neural networks (ANNs) that have proven to be particularly useful for the processing of imagery (computer vision), (written) language, and speech (natural language processing [NLP]), with data representations being stepwise handed over between layers, allowing for increasing complexity and abstraction (Esteva et al. 2019). In the following, we briefly sum up the main applications in this area: computer vision and NLP.

Computer vision allows machines to understand and work with images and videos; related tasks are image classification, object detection, and segmentation. Most common in this field is the application of convolutional neural networks (CNNs). These networks use convolutions to extract features from images, such as colors, textures, edges, geometric forms, and macroscopic structures, and pass a numerical representation of them to an ANN as its input vector. Depending on the aimed task, the ANN then maps this input to an output, such as classifying the image, detecting objects, or segmenting pixels. In dentistry, CNNs have been used for the automated detection and classification of anatomic landmarks (including teeth), mainly on cephalometric, panoramic, or 3-dimensional radiographs (e.g., cone beam computed tomography [CBCT]) but also photography (Schwendicke et al. 2019). The accuracy of these applications has been found to be similar to that of clinicians. Notably, there are only a few comparative studies allowing strong conclusions as to this relative performance against humans. In contrast to humans, DL can assess a cephalometric, panoramic, or CBCT data set within seconds, saving a significant amount of time and freeing up resources for other tasks. CNNs have also been used to detect apical lesions and periodontal bone loss on periapical and panoramic radiographs, as well as caries lesions on bitewings and periapicals, usually with accuracy similar and some cases superior to humans. On radiographs, the detection of restorations and of bony pathologies, such as osteoporosis or cysts, as well as prediction of growth patterns or skeletal relations has been attempted using CNNs. On photographic imagery, the detection and classification of caries lesions, mucosal and skin lesions, and facial profiles have been performed (Schwendicke et al. 2019). Only a few studies have assessed the cost-effectiveness (Schwendicke, Rossi, et al. 2020) or other impacts of computer vision technologies for dental care or dental public health; their generalizability and robustness in dentistry remain uncertain (Schwendicke, Samek, et al. 2020). A range of dental computer

vision applications are by now market-ready, and some have acquired regulatory approval, too.

Notably, computer vision in dentistry has so far focused on single tasks, that is, more comprehensive multitask detection or classification has not yet been performed. The embedding of context (e.g., clinical, demographic, or historic data) has also not yet been performed. Data protection (see below) and the efforts in labor required to provide human-derived labels for each image (supervised learning) currently limit the available amount of data for training, reducing the models' accuracy and generalizability. Strategies involving heavy data augmentation (Ronneberger et al. 2015), federated (distributed) learning on multiple (independent) data sets (Bonawitz et al. 2019), and unsupervised (e.g., using generative adversarial networks) (Goodfellow et al. 2016) or active learning (semisupervised learning and "human-in-the-loop approach," with humans controlling all or only the uncertain machine-derived labels) may help to overcome these limitations.

NLP allows machines to infer meaning from text and speech but also to translate or generate them. Common DL architectures in this field are recurrent neural networks (RNNs) or, more specifically, long short-term memory networks, which are particularly useful for sequential data inputs (Sutskever et al. 2014). However, these kinds of networks do not perform well on long and complex sequences. More recently, transformer networks such as bidirectional encoder representations from transformers (BERTs) and generative pretrained transformers are setting new standards and consistently outperform other approaches. These types of network architectures combine CNNs together with attention models and allow for parallel sequence computation and keeping track of long-distant dependencies within sequences (e.g., for text, valuable information that is important to infer meaning is often far apart).

The 2 most obvious sources for such language data in health care are eHR and voice recordings. NLP can leverage existing large (but oftentimes unstructured) amounts of language data via data mining, replacing manual omission or redaction, and developing prediction models. Dental eHRs have been mined for such purpose, with high classification accuracy for text-to-content matching (Chen et al. 2021). Similarly, NLP has been used to extract pain features from eHR, allowing to develop shallow ML models for predicting temporomandibular disorders (Nam et al. 2018). Combining NLP with computer vision allows automated labeling of imagery via eHR entries. Speech recognition and knowledge extraction facilitate comprehensive transcriptions of clinical visits (patient-provider conversations), automating today's labor-intensive manual reporting (Shickel et al. 2018). NLP applications have a good chance of reducing the administrative costs, for example, making billing more efficient by extracting relevant information from unstructured medical reports and assigning medical codes (e.g., *International Classification of Diseases* codes) automatically, and they provide effective clinical decision support by, for example, the identification and prevention of medication prescribing errors (Rozenblum et al. 2020).

Data-Enriched Clinical Care

The integration of the multitude of available data from the individual level (e.g., demographic, social, and clinical data obtained via records mining, clinical assessment, omics analyses, and real-time consumer data from wearables and tracking device), setting level (e.g., geospatial, environmental, or provider-related data), and systems level (e.g., health insurance, regulatory, and legislative data) has been shown to enrich and affect nearly all steps of clinical care (Fig. 2).

Besides (DL-based) medical analytics, omics technologies are an emerging part of the clinical workflow in dental care and will affect diagnosing and characterizing conditions but also predicting their course and thereby guiding therapy. For example, the use of omics profiling has been suggested to assist the tailored decision between different bone regeneration protocols and materials (Calciolari and Donos 2018). More prominent is the proteomic, transcriptomic, metabolomic, and microbiomic analysis of saliva. So far, however, accurate, affordable, and scalable tools for this purpose, usable in primary dental care, are unavailable.

Generally, there is a strive for continuous real-time monitoring rather than "on-off" episodic assessments. The availability of consumer products supporting such real-time health data collection disrupts the demand side of health care. The use of wearables for health monitoring and improvement has been researched in a wide range of areas; we here present examples where systematic reviews and meta-analyses are available; these are spread along patients' clinical pathway (Fig. 2): monitoring and training to improve balance and gait in Parkinson disease, stroke, neuropathy, or frail patients, with positive but inconsistent effects (Gordt et al. 2018); improving lifestyle and related outcomes like mobility and weight, with positive and consistent effects (Ringeval et al. 2020); activity interventions for cardiovascular diseases, showing significant small to medium improvements in activity levels (Hodkinson et al. 2019); monitoring multiple sclerosis severity, with high correlation between predicted and observed severity (Vienne-Jumeau et al. 2020); computerized cognitive behavioral therapy for depression, attention-deficit/hyperactivity disorder, autism, anxiety, psychosis, and eating disorders, with mixed results (Hollis et al. 2017); monitoring heart rate variability, with small but acceptable errors compared with clinically available measurement options (Dobbs et al. 2019); improving maternal health during pregnancy, including weight management, gestational diabetes mellitus, and asthma, with moderate to large positive effects (Chan and Chen 2019); and chronic disease management using goal setting, virtual social support, e-health programs, feedback, and diaries, with demonstrated benefits on weight, hemoglobin A1c, and exercise levels (Kamei et al. 2020).

Fewer data are available on wearables for oral and dental health. The most obvious use-case for assessing oral health–related behavior and outcomes is the toothbrush. Experimental toothbrushes employing accelerometers, magnetic sensors, and 3-dimensional visualizations have been comprehensively
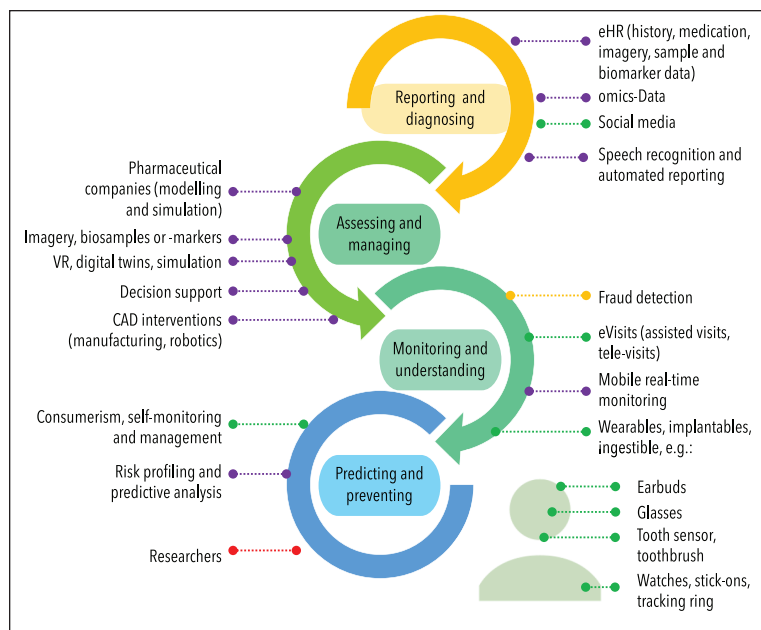
Figure 2. The data-driven clinical workflow. Data provided or used by different stakeholders (purple: provider; green: patient; yellow: payer; red: researcher) are permeating clinical care. CAD, computer-assisted design; eHR, electronic health record. This figure is available in color online.

tested and shown to support patient education and improve oral hygiene outcomes (Lee et al. 2007, 2012). Removable mouth-guards measuring glucose or uric acid concentrations in saliva have been used for research purposes but have not entered the consumer stage (Kim et al. 2015; Arakawa et al. 2016). The real-time analysis of saliva, while being promising, using nano-sensors (e.g., attached to teeth) has been conceptualized but not reached market readiness yet.

Besides individual-level data, the use of geospatial, environmental, or provider-related data has been suggested. Geospatial analyses allow one to map and predict the spread of infectious diseases, to assess the accessibility of health care institutions, to support environmental exposure analysis, or to improve logistic planning and strategic sampling in clinical research. In dentistry, they have been employed to evaluate water fluoride coverage (Curiel et al. 2020), geographic incentives affecting providers' decision making (Ghoneim et al. 2020), or the accessibility of services (Eke et al. 2019), for example. Provider-level data have been suggested to support the tailored application of diagnostic tools for caries detection, reflecting the individual disease spectrum and provider-related test conduct and application thresholds (Schwendicke et al. 2018).

On systems level, decision makers may use data of statutory or private insurers, employer-based insuring organizations, or pension schemes to better relate input (disease prevalence and spectrum, resources, workforce), throughput (processes, incentives), output (services provided, provider behavior, system's transformation), and outcomes (clinical outcomes, costs,

equity) data to guide and improve health care organization (Schrappe and Pfaff 2017).

The growing number of data sources will contribute to a deeper understanding of health and disease and a precise, personalized, predictive, and preventive approach in diagnostics and management, as well as in dentistry (Hamburg and Collins 2010; Flores et al. 2013; Schwendicke, Samek et al. 2020). Notably and as mentioned, harvesting and integrating such data will be only possible if these data are systematically collected and made available, which is not the case in many countries worldwide at present.

Data from and for Dental Research

A major source of research waste lies within data management (Glasziou et al. 2014). Oftentimes, the yielded data are not transparently and comprehensively reported or provided, limiting replication and reuse (Naudet et al. 2018) and contributing to the "reproducibility crisis" (Stupple et al. 2019). Open research data and data sharing are strategies to tackle these limitations and reduce waste, supported or enforced by major funding agencies and journals. Data sharing enables research information to be appraised, pooled, or employed for replication. Open benchmark data allow validating prediction models and demonstrating their transportability and comparative performance.

In dentistry, open data are so far uncommon, and while dental researchers acknowledge the value of sharing data, they remain critical toward data security and reuse (Vidal-Infer et al. 2018; Spallek et al. 2019; Cenci et al. 2020). There are, however, pilot projects demonstrating the feasibility of large-scale data sharing (Walji et al. 2014), and a number of studies demonstrated the value of cross-center data sharing. Notably, replication studies or the wide pooling of trial data (e.g., for individual participants meta-analysis) remain uncommon.

Considering the chronic nature of most dental conditions and their long-term sequels (on health but also on future treatment needs and costs), data reuse has been employed to populate research simulation models (Qu et al. 2019; Schwendicke 2019). Pulling and pooling data from a wide range of data sources allow extrapolation and modeling of chains of events, whose observation is usually beyond clinical studies (Fig. 3).

Generally, data should be at the center of future trials' conception, conduct, and reporting (Fig. 4). Data-centered study designs increase the quality and yield of traditional (clinical) trials and open up the option to complement them, mainly when it comes to data attainment but also usage. For example, the usage of eHealth applications, often involving patients' own data generators (e.g., wearables, implantables, ingestibles: the so-called "bring your own device" approach) or other data sources (e.g., social media), allows one to leverage a wealth of
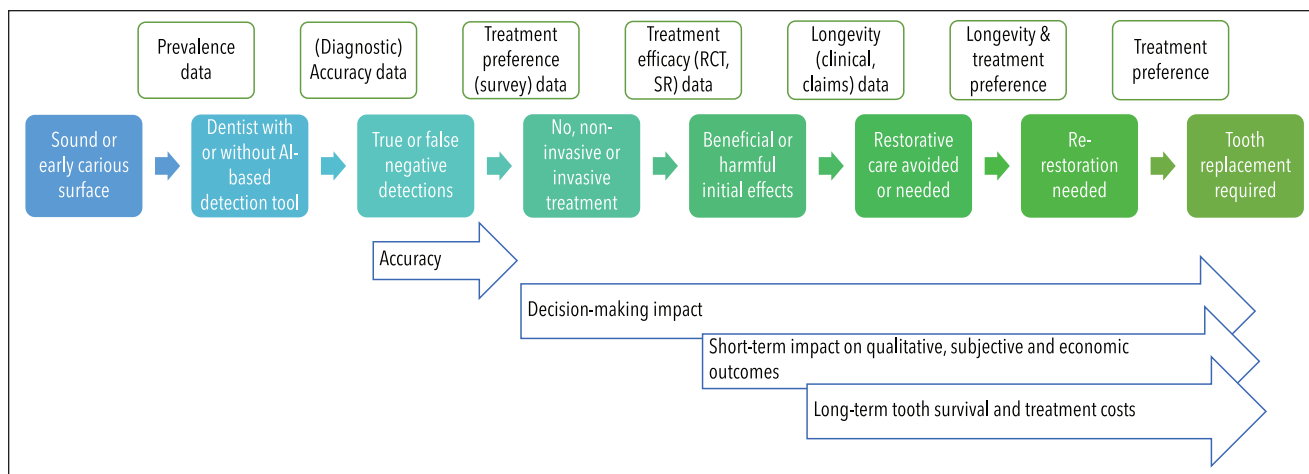
Figure 3. Modeling a chain of events (filled colored boxes in the middle row) based on various data sources and study types (green boxes in the upper row) allows one to reflect the long-term impact of decisions on health and further outcomes (blue arrow boxes). RCT, randomized controlled trial; SR, systematic review. This figure is available in color online.
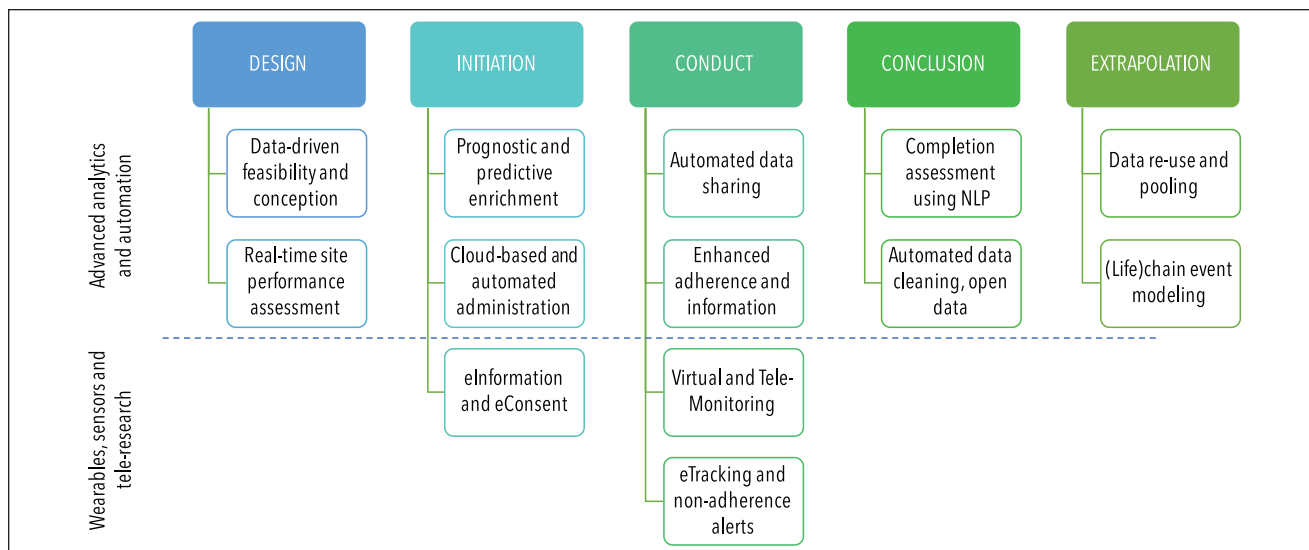
Figure 4. Clinical research in the data era. Advanced analytics and automation as well as the wealth of remote, real-time data affect clinical research, including trial design, initiation, conduct, conclusion, and beyond (e.g., extrapolation; also see Fig. 3). Trials will be tested for their conception and feasibility based on existing data and simulations, and possible trial sites can be assessed prior to initiation in real time for expected recruitment and so on. Inclusion criteria can be individualized based on biomarker, social, or historical data, allowing prognostic and predictive enrichment and leading to more targeted samples with higher statistical power. Trial administration and monitoring as well as data attainment and controlling may be fully remote and cloud supported. Recruitment, including patient information and consent, will be automated and electronic; social media channels may be used to individually approach fitting participants. Adherence to follow-up and therapy will be monitored real time; site visits by patients and monitors may be performed and supported electronically, reducing efforts, costs, and attrition. Comprehensive and broad (real-time) data will be collected (compare Fig. 2) and analyzed using deep learning. Data cleaning and/or imputation will be performed automatically, supporting a transparent and reliable data workflow. Open data policies facilitate replication, extrapolation, and pooling. NLP, natural language processing.

real-world (routine) data at low cost, even from populations that are usually hard to reach.

The availability of open-source frameworks for developing research apps supporting the usage of eHealth data, like ResearchKit (https://developer.apple.com/researchkit/, for iOS) and ResearchStack (http://researchstack.org, for Android), will support this movement. There are by now also standards and

recommendations toward employing patient-generated electronic outcomes in clinical trials (https://c-path.org/programs/eproc/). While patient-collected, routinely generated data offer a wide range of advantages over purposively (prospectively) collected (scientist-derived) data and promise additional insights, they are prone to biases generally affecting routine data (Table).

Table. Traditional versus Data-Focused, Crowd-Sourced, Distributed Clinical Research.

Item	Traditional Research	Crowd-Source, Distributed Research
Recruiting and consent	Pull mode, manual	Push mode, automated
Setting	Clinical care or related settings	Real-world settings
Populations	Specified, controlled, homogeneous	Less specified, less controlled, heterogeneous
Data	Narrow, episodic, clinically attained	Big, real-time and continuous, clinical, social, geographical, patient attained
Costs per data point	High	Low
Bias and challenges	Selection, attrition, reporting	Selection, memory bias, limited validation of many instruments; evaluation complex and costly

Modified from Inomata et al. (2020).

Main Challenges

A number of challenges for implementing and sustaining data-driven applications have been identified and apply to dentistry, too.

1. Concerns and confidence: The balance between the public interest in attaining data and individual data protection demands has been handled differently across the globe. Questions around broad consent, data donation, and cybersecurity but also around bias, fairness, and responsibility of AI and other data-driven applications have been raised. Confidence in abstract and complex data products that support the dental workforce (e.g., diagnostic support tools) and allow nondental professionals to perform certain tasks (e.g., dental screening via handheld devices in nursing homes by nurses) may grow over time as the public health benefits will become apparent. However, transparency, trustworthiness, and explainability are fundamental for the uptake of data-driven applications.
2. Pitfalls, bias, and failures: For data-driven application in health care, there is reasonable concern about handing over critical medical decisions to computers. The stakes are high, as any treatment decision will affect patients' well-being. Failures of data-driven applications are most often rooted in biases that are not always apparent and therefore difficult to compensate. Sample selection bias (i.e., training on nonrepresentative, small, and siloed data sets) leads to limitations in generalization, for example, on different devices or patient populations (Krois et al. 2021). Further bias originates from distributional shift in the target population; in such cases, data-driven systems may confidently make erroneous predictions based on "out-of-sample" inputs (Challen et al. 2019). Humans, including clinicians, are not perfect either and tend to give significance to evidence that supports their presumptions (confirmation bias). Automation bias and automation complacency, respectively, refer to phenomena where clinicians are in favor of accepting the guidance of an automated system, especially when challenged by multiple concurrent tasks (Parasuraman and Manzey 2010).
3. Capabilities: Stakeholders' (e.g., users' and consumers') capabilities toward adapting, employing, and appraising data applications are currently limited. Educating the future health care workforce in data literacy seems crucial.

Professionals need to be enabled to access, interpret, appraise, manage, and ethically handle data (Calzada Prado and Marzal 2013); there is a call for the "data-driven physician" (Stanford Medicine 2020). In a survey of 523 US physicians and 210 medical students and residents, nearly three-quarters of medical students and nearly half of all physicians are planning to pursue additional education around data (e.g., advanced statistics or data science), providing evidence that young professionals are aware of the future challenges for their profession and that training in this field is currently insufficient (Stanford Medicine 2020). Medical training and education need to keep pace with technological and data-driven developments. Notably, a good basic science background is already a prerequisite for entering dental and medical school, but little emphasis is placed on data analysis and applied mathematics. Data literacy should be a core competence in dental under- and postgraduate curricula. Introducing such new technologies into dental education is possible. One example for this is Computer-Assisted Design and Computer-Assisted Manufacturing (CAD-CAM): by now, many dentists routinely employ this technology and leverage their strengths while knowing how to cope with their weaknesses. Postgraduate courses on CAD-CAM are widely available and allow graduated dentists to learn and take up this technology, too. Once data-driven applications are available on the market and have proven their additional value, postgraduate courses and training offers will become available and will educate an "informed user"—someone who is not necessarily an expert but can actively navigate the field. Furthermore, democratizing data science via automation will allow medical professionals and researchers to perform or replicate data science exercises on their own (e.g., on open data), increase trust, and bridge the gap between the technical and medical domains. The data era and data dentistry will be a chance and not only a threat: it may help to push the profession toward a more critical and literate stance toward scientific data, as indicated by the exploding wealth of educational activities and their adoption by medical and dental professionals.

Besides individuals' capabilities, the technical capabilities need to be provided; storing and analyzing data at such scale require continuous upgrading of computer-processing power. Three technological pillars are of

importance; storage, transmission, and compute. All 3 have shown considerable growth rates over the past decades. For instance, the amount of data actually stored in data centers has increased 8-fold since 2015 (Statista 2021). The exponential growth of bandwidth and compute over the past decades is famously described by empirical laws such Edholm's law (after Phil Edholm), Moore's law (after Gordon Moore), or Koomey's law (after Jonathan Koomey). Computing power has increased constantly over the decades; consumer computing currently reaches Giga scale (10^9 floating point operations per second [FLOPS]), modern graphic cards (GPUs) perform in the Terra scale (10^{12} FLOPS) range, and industry computing is in the Peta scale (10^{15} FLOPS) range. Distributed computing recently even broke the Exa-FLOPS barrier (10^{18}) (Foldingathome 2020).

Notably, irrespective of the system's capabilities, making data and compute available for the society's greater good remains uncertain; computational resources are expensive, and access to those will not be distributed equally. Furthermore, in many domains, data are kept private and not made accessible for financial gains, something that should not be expected to be different in health care.

4. Standardization: Data exchange and usage require harmonization and standardization. Efforts toward systematizing medical terminology (e.g., SNOMED: The Systematized Nomenclature of Medicine Clinical Terms, containing 300,000 uniquely identified, logically defined, and hierarchically arranged terms, or MedDRA: Medical Dictionary for Regulatory Activities, a standardized medical terminology facilitating medical product regulatory information exchange) support the semantic interoperability of data, allowing cross-sectional data exchange. Public agencies, trusted institutions, or community-based initiatives may take a prominent role in establishing standards and benchmarking frameworks to ensure quality, generalizability, and transparency of data-driven applications (International Telecommunication Union 2018).

Action Items for the Dental and Oral Medicine Community

From the above said, we derive 3 areas that require action for the community.
1. Data availability, refinement, and usage: To fulfill the promises of data dentistry, dental data silos need to be broken up and made accessible for secure integration and use in research and clinical care. On one hand, real-time patient-derived (routine) data will allow one to better capture the wider socioeconomic-behavioral or environmental determinants of oral health. On the other hand, researchers should not only rely on "big" retrospectively and routinely obtained data but aim to attain similarly large, prospectively and purposively collected data, for example, via data sharing and pooling. Employing such

structured and counterbiased data sets (i.e., they are also biased but in a different direction than routine data) will allow one to validate and enhance prediction models or simulations. Besides data generation, dental researchers should contribute to the development of data-driven applications; they, in contrast to developers and engineers, have the domain expertise and are aware of deficits and needs. For AI in dentistry, knowledge-based annotation (labeling) strategies for data and the abstraction of insights from other medical disciplines will be appreciated by the engineering community (Schwendicke, Samek, et al. 2020).
2. Demonstrate value and usefulness: Data-driven health care is slowly permeating dentistry. Technological hurdles like high costs and the perceived lack of relevance have limited its adoption in general dental practice. Increasing the scientific underpinning of the usefulness, cost-effectiveness, generalizability, fairness, and robustness of dental data-driven applications and demonstrating their impact for increasing overall health are required.
3. Dental workforce and community: Educating the dental workforce and prioritizing data literacy in future dental curricula as well as supporting a closer cooperation between dental and data science professionals to bridge interprofessional gaps will address a range of the described implementation barriers. Standardized infrastructure and processes for cross-discipline data exchange and use should be enhanced. Open data, allowing pooling, replicating, benchmarking, or further reuse, should be incentivized. Dental domain knowledge should be reflected when developing or advancing regulatory processes (e.g., involving living AI).

Conclusions

Data are expected to improve the quality, accessibility, affordability, safety, and equity of health care. Dental care and research are currently transforming into what we term *data dentistry* with modern technologies, allowing one to master unprecedent amounts of data for analysis and simulation, as well as fostering biologic understanding, the development of new diagnostics and therapeutics, and—overall—a more precise, personalized, predictive, and preventive care. Open research data and data sharing allow one to appraise, benchmark, pool, replicate, and reuse data. To make full use of the potential of data dentistry, concerns and confidence into data-driven applications need to be addressed; the availability of data and labels increased; stakeholders' and systems' capabilities improved; standardized data formats, infrastructure, and processes employed; and open research data and data sharing incentivized.

Author Contributions

Declaration of Conflicting Interests

Funding

ORCID iDs

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

J. Krois (iD) https://orcid.org/0000-0002-6010-8940

References

Arakawa T, Kuroki Y, Nitta H, Chouhan P, Toma K, Sawada S, Takeuchi S, Sekita T, Akiyoshi K, Minakuchi S, et al. 2016. Mouthguard biosensor with telemetry system for monitoring of saliva glucose: a novel cavitas sensor. Biosens Bioelectron. 84:106–111.

Bell G, Hey T, Szalay A. 2009. Beyond the data deluge. Science. 323(5919):1297–1298.

Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, Kiddon C, Konečný J, Mazzocchi S, McMahan HB. 2019. Towards federated learning at scale: system design. arXiv preprint [accessed XXXX]. https://arxiv.org/abs/1902.01046.

Calciolari E, Donos N. 2018. The use of omics profiling to improve outcomes of bone regeneration and osseointegration: how far are we from personalized medicine in dentistry? J Proteomics. 188:85–96.

Calzada Prado J, Marzal MÁ. 2013. Incorporating data literacy into information literacy programs: core competencies and contents. Libri. 63(2):123–134.

Cenci MS, Franco MC, Raggio DP, Moher D, Pereira-Cenci T. 2020. Transparency in clinical trials: adding value to paediatric dental research. Int J Paed Dent. 31(Suppl 1):4–13.

Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. 2019. Artificial intelligence, bias and clinical safety. BMJ Qual Safety. 28(3):231–237.

Chan KL, Chen M. 2019. Effects of social media and mobile health apps on pregnancy care: meta-analysis. JMIR Mhealth Uhealth. 7(1):e11836.

Chang AY, Cowling K, Micah AE, Chapin A, Chen CS, Ikilezi G, Sadat N, Tsakalos G, Wu J, Younker T, et al. 2019. Past, present, and future of global health financing: a review of development assistance, government, out-of-pocket, and other private spending on health for 195 countries, 1995–2050. Lancet. 393(10187):2233–2260.

Chen Q, Zhou X, Wu J, Zhou Y. 2021. Structuring electronic dental records through deep learning for a clinical decision support system. Health Inf J. 27(1):1460458220980036.

Curiel JA, Sanders AE, Slade GD. 2020. Emulation of community water fluoridation coverage across us counties. JDR Clin Trans Res. 5(4):376–384.

Dobbs WC, Fedewa MV, MacDonald HV, Holmes CJ, Cicone ZS, Plews DJ, Esco MR. 2019. The accuracy of acquiring heart rate variability from portable devices: a systematic review and meta-analysis. Sports Med. 49(3):417–435.

Economist. 2017. The world's most valuable resource is no longer oil, but data [accessed 2021 June 5]. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Eke PI, Lu H, Zhang X, Thornton-Evans G, Borgnakke WS, Holt JB, Croft JB. 2019. Geospatial distribution of periodontists and us adults with severe periodontitis. J Am Dent Assoc. 150(2):103–110.

Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. 2019. A guide to deep learning in healthcare. Nat Med. 25(1):24–29.

Flores M, Glusman G, Brogaard K, Price ND, Hood L. 2013. P4 medicine: how systems medicine will transform the healthcare sector and society. Personal Med. 10(6):565–576.

Foldingathome. 25 Mar 2020. Twitter post [accessed 6 May 2020]. https://mobile.twitter.com/foldingathome/status/1242918035788365830

Ghoneim A, Yu B, Lawrence HP, Glogauer M, Shankardass K, Quiñonez C. 2020. Does competition affect the clinical decision-making of dentists? A geospatial analysis. Comm Dent Oral Epidemiol. 48(2):152–162.

Glasziou P, Altman DG, Bossuyt P. 2014. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 383(9913):267–276.

Goodfellow I, Bengio Y, Courville A. 2016. Deep learning. Cambridge (MA): MIT Press.

Gordt K, Gerhardy T, Najafi B, Schwenk M. 2018. Effects of wearable sensor-based balance and gait training on balance, gait, and functional performance in healthy and patient populations: a systematic review and meta-analysis of randomized controlled trials. Gerontology. 64(1):74–89.

Hamburg MA, Collins FS. 2010. The path to personalized medicine. N Engl J Med. 363(4):301–304.

Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillan B, Bower P, Panagioti M. 2019. Accelerometer- and pedometer-based physical activity interventions among adults with cardiometabolic conditions: a systematic review and meta-analysis. JAMA Netw Open. 2(10):e1912895.

Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, Davies EB. 2017. Annual research review: digital health interventions for children and young people with mental health problems: a systematic and meta-review. J Child Psychol Psychiatry. 58(4):474–503.

Inomata T, Sung J, Nakamura M, Fujisawa K, Muto K, Ebihara N, Iwagami M, Nakamura M, Fujio K, Okumura Y, et al. 2020. New medical big data for p4 medicine on allergic conjunctivitis. Allergol Int. 69(4):510–518.

International Telecommunication Union. 2018. Focus group on "artificial intelligence for health" [accessed 2021 Jan 3]. https://www.itu.int/en/itu-t/focusgroups/ai4h/pages/default.aspx

Kamei T, Kanamori T, Yamamoto Y, Edirippulige S. 2020. The use of wearable devices in chronic disease management to enhance adherence and improve telehealth outcomes: a systematic review and meta-analysis. J Telemed Telecare [epub ahead of print 20 Aug 2020]. doi:10.1177/1357633X20937573

Kim J, Imani S, de Araujo WR, Warchall J, Valdés-Ramírez G, Paixão TR, Mercier PP, Wang J. 2015. Wearable salivary uric acid mouthguard biosensor with integrated wireless electronics. Biosens Bioelectron. 74:1061–1068.

Klingenberg CO, Borges MAV, Antunes JJAV. 2021. Industry 4.0 as a data-driven paradigm: a systematic literature review on technologies. J Manufactur Techn Management. 32(3):570–592.

Krois J, Garcia Cantu A, Chaurasia A, Patil R, Kumar Chaudhari P, Gaudin R, Gehrung S, Schwendicke F. 2021. Generalizability of deep learning models for dental image analysis. Sci Rep. 11(1):6102.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature. 521(7553):436–444.

Lee KH, Lee JW, Kim KS, Kim DJ, Kim K, Yang HK, Jeong K, Lee B. 2007. Tooth brushing pattern classification using three-axis accelerometer and magnetic sensor for smart toothbrush. Annu Int Conf IEEE Eng Med Biol Soc. 2007:4211–4214.

Lee YJ, Lee PJ, Kim KS, Park W, Kim KD, Hwang D, Lee JW. 2012. Toothbrushing region detection using three-axis accelerometer and magnetic sensor. IEEE Trans Biomed Eng. 59(3):872–881.

Nam Y, Kim HG, Kho HS. 2018. Differential diagnosis of jaw pain using informatics technology. J Oral Rehab. 45(8):581–588.

Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, Ioannidis JPA. 2018. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in the *BMJ* and *PLoS Medicine*. BMJ. 360:k400.

Parasuraman R, Manzey DH. 2010. Complacency and bias in human use of automation: an attentional integration. Hum Factors. 52(3):381–410.

Peres MA, Macpherson LMD, Weyant RJ, Daly B, Venturelli R, Mathur MR, Listl S, Celeste RK, Guarnizo-Herreno CC, Kearns C, et al. 2019. Oral diseases: a global public health challenge. Lancet. 394(10194):249–260.

Perez C. 2002. Technological revolutions and financial capital: the dynamics of bubbles and golden ages. Cheltenham (UK): Edward Elgar.

Qu Z, Zhang S, Krauth C, Liu X. 2019. A systematic review of decision analytic modeling techniques for the economic evaluation of dental caries interventions. PLoS One. 14(5):e0216921.

Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. 2020. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Int Res. 22(10):e23954.

Ronneberger O, Fischer P, Brox T. 2015. Dental x-ray image segmentation using a u-shaped deep convolutional network. International Symposium on Biomedical Imaging (ISBI). 1–13.

Rozenblum R, Rodriguez-Monguio R, Volk LA, Forsythe KJ, Myers S, McGurrin M, Williams DH, Bates DW, Schiff G, Seoane-Vazquez E. 2020. Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation. Jt Comm J Qual Patient Saf. 46(1):3–10.

Schrappe M, Pfaff HE. 2017. Einführung in die versorgungsforschung. In: Pfaff H, Neugebauer E, Glaeske G, Schrappe M, editors. Lehrbuch versorgungsforschung, 2. Auflage. Stuttgart (Germany): Schattauer. p. 1–63.

Schwendicke F. 2019. Less is more? The long-term health and cost consequences resulting from minimal invasive caries management. Dent Clin North Am. 63(4):737–749.

Schwendicke F, Elhennawy K, El Shahawy O, Maher R, Gimenez T, Mendes FM, Willis BH. 2018. Visual and radiographic caries detection: a tailored meta-analysis for two different settings, Egypt and Germany. BMC Oral Health. 18(1):105.

Schwendicke F, Golla T, Dreher M, Krois J. 2019. Convolutional neural networks for dental image diagnostics: a scoping review. J Dent. 91:103226.

Schwendicke F, Rossi JG, Göstemeyer G, Elhennawy K, Cantu AG, Gaudin R, Chaurasia A, Gehrung S, Krois J. 2020. Cost-effectiveness of artificial intelligence for proximal caries detection. J Dent Res. 100(4):369–376.

Schwendicke F, Samek W, Krois J. 2020. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 99(7):769–774.

Shickel B, Tighe PJ, Bihorac A, Rashidi P. 2018. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform. 22(5):1589–1604.

Solbach T, Kremer M, Grünewald P, Ickerott D. 2019. Driving the future of health. PWC Strategy [accessed 2020 Dec 2]. https://www.strategyand .pwc.com/de/de/studien/2019/die-zukunft-der-gesundheit-vorantreiben/ Driving-the-future-of-health.pdf.

Spallek H, Weinberg SM, Manz M, Nanayakkara S, Zhou X, Johnson L. 2019. Perceptions and attitudes toward data sharing among dental researchers. JDR Clin Trans Res. 4(1):68–75.

Stanford Medicine. 2020. Stanford medicine 2020 health trends report: the rise of the data-driven physician [accessed 27 Dec 2020]. https://med .stanford.edu/content/dam/sm/school/documents/Health-Trends-Report/ Stanford%20Medicine%20Health%20Trends%20Report%202020.pdf

Statista. 2018. Total amount of global healthcare data generated in 2013 and a projection for 2020 [accessed 2021 Mar 4]. https://www.statista.com/statis tics/1037970/global-healthcare-data-volume/

Statista. 2021. Data center storage capacity worldwide from 2016 to 2021, by segment [accessed 2021 Mar 4]. https://www.statista.com/statistics/638593/ worldwide-data-center-storage-capacity-cloud-vs-traditional/

Stupple A, Singerman D, Celi LA. 2019. The reproducibility crisis in the age of digital medicine. NPJ Digit Med. 2:2.

Sutskever I, Vinyals O, Le QV. 2014. Sequence to sequence learning with neural networks. Paper presented at: Proceedings of the 27th International Conference on Neural Information Processing Systems—Volume 2. Montreal, Canada: MIT Press. https://dl.acm.org/doi/10.5555/2969033.2969173

Vidal-Infer A, Tarazona B, Alonso-Arroyo A, Aleixandre-Benavent R. 2018. Public availability of research data in dentistry journals indexed in journal citation reports. Clin Oral Investig. 22(1):275–280.

Vienne-Jumeau A, Quijoux F, Vidal PP, Ricard D. 2020. Wearable inertial sensors provide reliable biomarkers of disease severity in multiple sclerosis: a systematic review and meta-analysis. Ann Phys Rehabil Med. 63(2):138–147.

Walji MF, Kalenderian E, Stark PC, White JM, Kookal KK, Phan D, Tran D, Bernstam EV, Ramoni R. 2014. Bigmouth: a multi-institutional dental data repository. J Am Med Info Assoc. 21(6):1136–1140.

World Health Organization. 2020. Oral health: achieving better oral health as part of the universal health coverage and noncommunicable disease agendas towards 2030 [accessed 2021 June 5]. https://apps.who.int/gb/ebwha/ pdf_files/EB148/B148_8-en.pdf