


Using big data to promote precision oral health in the context of a learning healthcare system

Joseph Finkelstein, MD, PhD¹ ; Frederick Zhang, BA²; Seth A. Levitin, BS²; David Cappelli, DMD, MPH, PhD³

1 Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA

2 Center for Bioinformatics and Data Analytics in Oral Health, College of Dental Medicine, Columbia University, New York, NY, USA

3 Department of Biomedical Sciences, School of Dental Medicine, University of Nevada, Las Vegas, NV, USA

Keywords

precision medicine; big data; learning health system; public health dentistry.

Correspondence

Joseph Finkelstein, Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Tel.: 212-659-9596

Fax: 212-423-2998

e-mail: joseph.finkelstein@mssm.edu.

Received: 2/14/2019; accepted 12/2/2019.

doi: 10.1111/jphd.12354

Journal of Public Health Dentistry **80** (2020) S43–S58

Summary

There has been a call for evidence-based oral healthcare guidelines, to improve precision dentistry and oral healthcare delivery. The main challenges to this goal are the current lack of up-to-date evidence, the limited integrative analytical data sets, and the slow translations to routine care delivery. Overcoming these issues requires knowledge discovery pipelines based on big data and health analytics, intelligent integrative informatics approaches, and learning health systems. This article examines how this can be accomplished by utilizing big data. These data can be gathered from four major streams: patients, clinical data, biological data, and normative data sets. All these must then be uniformly combined for analysis and modelling and the meaningful findings can be implemented clinically. By executing data capture cycles and integrating the subsequent findings, practitioners are able to improve public oral health and care delivery.

Introduction

Precision medicine has become a key component in the modern delivery of healthcare. The aim of precision medicine is to tailor healthcare delivery to individual patients, thereby improving care delivery and outcomes while decreasing costs.¹ In determining the best care for an individual, precision medicine takes various factors into account in addition to disease history, such as an individual's environment, genome, and socioeconomic status. By combining these elements, a profile can be generated for each individual that better predicts their health outcomes and addresses risk factors. Combined information provides an opportunity to tailor treatment to the needs of the individual patient based on their risk factors. To successfully integrate precision medicine into oral healthcare, three major challenges must be addressed: development of up-to-date evidence-based guidelines,^{2,3} integration of large analytical data sets,^{4,5} and translation of new knowledge into routine clinical care delivery.^{6,7} In the fields of dentistry and oral health, implementation of these factors represents a major obstacle to the delivery of precision oral medicine.⁸

In recent years, oral health sciences called for a greater emphasis on evidence-based practice, especially as advances in

the field of healthcare and medicine have accelerated.^{9–11} Without evidence, clinical practice is empirical, anecdotal, and experiential. In the field of dentistry, certain therapies or procedures are used in clinical practice without substantive evidence for their use. For example, a systematic review of scaling and root planning therapy and its adjuncts found that of the nine adjuncts investigated in the paper, only four showed evidence of benefit over scaling and root planning treatment alone.¹² Another review could not locate long-term studies that determined the duration that positive effects from scaling could be maintained. Accordingly, the review could not determine the optimal recall frequency.¹³ Scaling and root planning therapy is a routine procedure that is considered “first-line therapy” for chronic periodontal disease, so the lack of literature that outlines how long therapy should be maintained is concerning. The lack of well-established, evidence-based guidelines is partially a result of the limited availability of data sets in dental research. A call for evidence-based dentistry as the foundation for modern dental practice has been recently reiterated by an editorial published by Dr. Robert J. Weyant.²

The second challenge facing precision oral health implementation is the limited amount of integrative analytical data

sets, which precludes the possibility of making evidence-based decisions. The root causes of this problem are manifold. One significant problem is that insurance claims in the United States, which are a major source of data for dental research, do not have diagnostic codes, but rather treatment codes, when used for dental insurance.¹⁴ The data elements used to track information in a dental clinic can vary widely from those used for research purposes. A study compared the data elements used by the National Institute of Dental and Craniofacial Research's Cancer Data Standards Registry and the Dental Information Model, a taxonomy of patient records based on data charting by practicing general dentists. This study found that only 46% of the data elements based on a clinical setting overlapped with the data elements used in research.¹⁵ These obstacles hinder the development of large, analytical data sets that can be used for research purposes. Improvements in data curation and aggregation in dentistry and oral medicine is a necessary step for the integration of precision medicine practices into oral healthcare. An urgent need for intelligent and harmonized aggregation of heterogeneous oral health data sets has been well articulated in several recent publications.^{4,8}

The third challenge is the slow translation from research to routine care delivery. A survey of clinical dental school faculty found that only 47 percent of faculty integrate evidence-based dentistry into their teaching.¹⁶ One potential cause for this is that the search for high-quality data is overwhelming and cumbersome, and that clear clinical practice guidelines are not available.¹⁷ An easily accessible resource, based on up-to-date evidence could help address this need and improve the implementation of guidelines into clinical practice. Optimal approaches for facilitating translation of research findings into practice have been the subject of broad discussion across all domains of care delivery as reflected by multiple publications in the field of dissemination and implementation science.^{18,19} In order to address the challenge of implementing personalized healthcare into routine practice, the Learning

Healthcare System (LHS) paradigm has been introduced to capitalize on the great potential of high-scale healthcare data combined with patient perspectives into integrated models of continuous care improvement.²⁰

A LHS was defined by the Institute of Medicine²¹ as a system in which “science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience.” A scalable infrastructure for big data aggregation and analytics provides necessary framework for personalized health implementation in the context of LHS.²² Thus, one potential solution to the three challenges outlined above could be the implementation of big data and data science practices into both the dental research infrastructure and routine clinical oral healthcare.

The National Institute of Dental and Craniofacial Research has emphasized the importance of evidence-based treatment and developed ways to implement data science systematically through funding, education, and fellowships.⁹ The translation of new knowledge discovered by big data analytics into evidence-based dental practice has potential to improve public oral health outcomes.⁴ The term “precision public health” has arisen as a descriptor for this approach. Precision public health has been defined as a method to “[improve] the ability to prevent disease, promote health, and reduce health disparities in populations by applying emerging methods and technologies for measuring disease, pathogens, exposures, behaviors, and susceptibility in populations; and developing policies and targeted implementation programs to improve health.”²³ This approach requires the usage of big data in order to predict risk profiles, improve our understanding of the pathogenesis of disease, and develop targeted treatment strategies, among other applications.²⁴

In this article, we will be discussing the potential role of big data in promoting the oral health of the public. The implementation of big data into oral health research and practice is still a developing process; however, successful implementation of big data practices can overcome the challenges preventing the integration of precision medicine and oral health. By addressing these challenges properly, the field of oral health can become integrated into the age of precision medicine while also addressing dental public health needs, a concept we have illustrated in Table 1. Through a scoping review, we will examine the sources of data available for oral health research, its applications for public oral health care delivery, and current barriers of implementation.

Methods

Scoping review has been used in this article as a method to map out the available literature regarding a specific topic

Table 1 Challenges and Solutions Facing the Integration of Precision Medicine into Oral Public Health

Challenge	Solution	Precision medicine	Dental public health
Lack of up-to-date evidence-based guidelines	Knowledge discovery pipelines based on big data and health analytics	X	X
Limited integrative analytical data sets	Intelligent integrative informatics approaches	X	X
Slow translation to routine care delivery	Learning health systems	X	X

of research in order to characterize the current landscape and identify future directions in this field. The scoping review methodology is based on the following five steps as previously described²⁵:

1. Identify a research question
2. Identify relevant studies
3. Evaluate and select studies to include
4. Chart the data
5. Collect, summarize, and report the results

Research question

The primary goal of this scoping review was to identify current trends and future applications in the use of big data in oral healthcare. To that end, the research question we developed was as follows: “What is the current role of big data in oral health care and what new applications are being developed or proposed at this time?”

Search strategy

We conducted a literature search of the PubMed database in order to identify papers to be included in this study. Our search only included articles that were either published in English or had an English translation available. We used the following Boolean search term: (“big data” OR “data mining” OR “data science” OR “data repository” OR “precision dentistry”) AND (“oral health” OR “dental health” OR “dental care” OR “oral hygiene” OR “dental medicine” OR “dental public health”). This search term yielded 78 potential papers to be included in our study.

Study selection

After our initial search, we established inclusion and exclusion criteria for papers to be included in our final analysis. The inclusion criteria were as follows:

1. The paper had to investigate the applications of big data, data mining, or data aggregation for oral health research.
2. The paper had to be available in English.
3. We included papers where big data may not have been the primary focus, but data mining or other data analysis techniques were still used on large data sets in the methodology of the study.

We then established the following exclusion criteria:

1. Perspective or opinion pieces were not included in this review.
2. Other review articles were not included in this review.

Results

Data charting

After applying article inclusion and exclusion criteria, we found 21 total papers to be included in the final study. The scoping review identified three major categories of literature regarding big data and oral health to be discussed in this review. They are as follows:

1. *Aggregation of big data for precision oral health*: The implementation of big data in oral health practice requires the establishment of data streams to aggregate and warehouse data. In our search, we identified four papers that describe the establishment of a data repository for oral health research.
2. *Use of big data for predictive analytics and knowledge discovery in oral health*: Currently, the most established implementation of big data in oral health research is secondary outcomes analysis. This process consists of using data mining and other data analysis techniques as established in the field of biomedical informatics to identify new patterns and trends from existing data sets.
3. *Development of future applications in oral health using big data*: Beyond secondary outcome analysis studies, we identified studies that represent forays into novel areas of research. These studies did not fit either of the above categories but represented new applications of big data into the field of oral health, such as the role of social media in oral health education.

Data extraction

The general characteristics of each of the studies included in this review were extracted and summarized in Table 2. In Table 3, we synthesized the different types of primary data sources used for the 14 secondary outcomes analysis papers included in this review based on the country where the study was conducted.

Data repositories and sources: aggregation of big data for precision oral health

Of the four papers devoted to developing data repositories we identified in this review, three were based in the United States: BigMouth, the Consortium for Oral Health-Related Informatics (COHRI), and the National Dental Practice-Based Research Network (DPBRN). The BigMouth data repository was initially developed as a repository linking the electronic health records (EHRs) of four dental schools in the United States. This repository has since expanded by including eight dental schools to house over 1.2 million oral health records. COHRI is a consortium of over

Table 2 General Characteristics of the Studies Included

Author	Country of origin	Study description
Data repository development von Bültzingslöwen et al., 2019	Sweden	The Swedish Quality Registry for Caries and Periodontal Diseases is a database of electronic patient dental records collected from affiliated dental care organizations ²⁶
Walji et al., 2014	USA	The BigMouth data repository is a collection of EHRs which was initially collected from four dental schools in the United States ²⁷
Gilbert et al., 2013	USA	The National Dental Practice-Based Research Network is a network of practicing and academic dentists and researchers who collaborate for data collection and research purposes ²⁸
Stark et al., 2010	USA	The Consortium for Oral Health-Related Informatics is a consortium of over 20 dental schools designed to share best practices and develop standardized data collection tools including BigMouth ²⁹
Predictive analytics Rao et al., 2019	Canada	EHRs from the Canadian Hospitals Injury Reporting and Prevention Program database were mined to identify the incidence of toothbrush-related injuries ³⁰
Suni et al., 2013	Finland	Municipal dental records in Finland were mined to develop Kaplan–Meier survival curves for caries-free permanent teeth and restoration survival distribution ³¹
Käkilehto et al., 2009	Finland	EHRs from four public dental health centers in Finland were mined to develop Kaplan–Meier curves for restorations of different restorative materials ³²
Raedel et al., 2017	Germany	Claims data from a large German health insurance company were mined to develop a Kaplan–Meier survival curve for posterior tooth restorations ³³
Lee et al., 2018	Korea	The Korea National Health and Nutrition Examination Surveys from 2010 to 2015 were mined to develop a decision tree model for predicting risk of periodontal disease ³⁴
Chan et al., 2016	Taiwan	EHRs from the National Health Insurance research database in Taiwan were mined to identify differences in outcomes between patients who receive conventional periodontal therapy and patients who receive comprehensive periodontal therapy ³⁵
Su et al., 2019	Taiwan	Data from the Taiwanese Nationwide Oral Cancer Screening Program were mined to determine the relationship between anatomic site of oral cancer and its staging and mortality ³⁶
Nalilah et al., 2013	USA	The Nationwide Emergency Department Sample database was mined to discover the relationship between mental illness and dental disease ³⁷
Thyvalikakath et al., 2015	USA	EHRs from the Indiana University School of Dentistry were mined in order to develop a model for predicting risk of periodontal disease ³⁸
Rai et al., 2019	USA	EHRs from the University of Colorado School of Dental Medicine were mined in order to identify factors associated with partial edentulism ³⁹
Filker et al., 2013	USA	EHRs from the Nova Southeastern University College of Dental Medicine were mined to find characteristics associated with caries risk level, including geographic median income level ⁴⁰
Boland et al., 2013	USA	EHRs from the Columbia University College of Dental Medicine were linked to medical records of the same patients at a nearby hospital and analyzed in order to identify associations between medical and dental diseases ⁴¹
Kalenderian et al., 2016	USA	Data from the BigMouth data repository were queried for patients diagnosed with chronic moderate periodontitis and analyzed for the percentage that received treatment that followed current evidence-based guidelines ⁴²
Tiwari et al., 2019	USA	Data claims for Medicaid-enrolled children from 13 states were mined to find the association between number of routine pediatric physician visits and preventive dental visits in children ⁴³
Future applications Huber et al., 2019	USA	Text-based social media posts responding to the 2016 ADA sealants guideline across a variety of different platforms were analyzed for their alignment with the ADA guideline ⁴⁴
Helmi et al., 2018	USA	The Media Cloud searchable big data platform was queried for published digital media related to community water fluoridation. These media were then analyzed for their stance on community water fluoridation ⁴⁵
Liu et al., 2013	USA	The data elements from the Cancer Data Standard Registry and Repository and the Dental Information Model were compared to each other in order to characterize the overlap in data elements used for dental research purposes as opposed to general clinical dental records ⁴⁶

Table 3 Primary Data Sources Used from Each Country

Country	Primary data sources used
United States	Dental school EHRs Hospital EHRs Academic data repositories Insurance claim databases
Canada	Hospital EHRs
Finland	Public health center EHRs
Germany	Insurance claim databases
Korea	Public health screening
Taiwan	Insurance claims database Public health screening

20 dental schools that share the common goal to further the field of oral health informatics and improve standardized data collection. The third US-based data repository is the DPBRN, which is sponsored by the NIDCR. This research network consists of practicing dentists and academics in the field of oral health who have agreed to share data and conduct research in a collaborative fashion. The fourth paper describing the development of a data repository was based in Sweden and presented the Swedish Quality Registry for Caries and Periodontal Diseases (SKaPa). In contrast to the data repositories being developed in the United States, which were primarily based on collecting data from academic institutions, SKaPa was developed as a registry for both private and public dental clinics in Sweden.

Predictive analytics for knowledge discovery

Of the data sources used for predictive analytics and knowledge discovery in the studies that were based in the United States, the most common was EHR data from dental schools, which was used in four of the seven studies included in this review. Other data sources used were the Nationwide Emergency Department Sample, which is a national database of emergency department visits, the BigMouth data repository described above, and Medicaid claims data. Studies based in countries outside of the United States used data from a variety of sources. A Canadian study queried the Canadian Hospitals Injury Reporting and Prevention Program, which is a hospital-based database of injury and poisoning events. Two studies based in Finland both used EHR data from community health centers. A study in Taiwan used data from the National Health Insurance research database, which is a database of claims data from Taiwan's mandatory single-payer insurance program. The other study from Taiwan used data from the Taiwanese Nationwide Oral Screening Program, a public health initiative to screen adults in Taiwan who have risk factors for oral cancer. These patients were then linked to data from Taiwan's National Death Registry. A study in Germany also used claims data

from a large national health insurance company. Finally, the study from Korea used data from the Korean National Health and Nutrition Examination Surveys. This includes a survey of health and nutrition behaviors as well as a physical health examination, including an oral health examination.

Large EHR data sets were used successfully to generate significant research output on a variety of topics ranging from risk factors for periodontal disease, incidence of rare adverse events, tooth survival curves, predictive data mining for periodontal disease, and outcomes of comprehensive periodontal treatment. Other areas where EHR data were used for secondary analysis included edentulism risk factors, obesity impact on oral cancer outcomes, posterior restoration outcomes, dental school strategic planning, quantitative model generation, and outcomes of dental preventive visits. A broad range of methodologies was employed from simple descriptive statistics to logistic regression, Kaplan–Meier survival analysis, decision tree models and various machine-learning approaches.

Future directions

Despite successful use of existing data sets for big data analytics, many articles emphasized limitations of their data sets and need for more comprehensive integrated framework for big data allowing simultaneous inclusion of multiple basic science, clinical and social science domains related to oral health. Existing data repositories such as SKaPa, National Dental Practice-Based Research Network, BigMouth were instrumental in building useful predictive models but lacked generalizability due to the limitations of the patient populations (dental school data, Medicaid data, insurance data) included in these databases. However, these repositories provided useful and valuable initial information for implementing precision dental public health strategies. A number of articles utilized big data generated by social media to characterize oral health in diverse patient population and to generate tailored messaging promoting oral health guidelines. Overall, a consensus framework for future directions of big data for precision oral health evolved from these articles that comprised a vision of synergistic and harmonized aggregation of multiple heterogeneous data sets pertinent to oral health and dental care delivery from sequencing data, proteomics, metabolomics, to EHR data, exposome, and social media and environmental data. Future directions using Big Data were identified in articles describing generation of supporting materials for evidence-based dentistry, tailored healthcare guideline sharing via social media, development of common data elements for sharing clinical and research data at point of care, and innovative approaches for identifying and targeting population subgroups for preventive care.

Discussion

The scoping review identified three major categories of articles reflecting the current state and potential future use of big data in precision dental public health. A number of articles described approaches of aggregating multiple heterogeneous data streams to create harmonized systematic representation of oral disease prevention and care delivery. Despite multiple limitations, these data sets are rapidly increasing in volume and complexity requiring innovative approaches for their visualizations and analysis.⁴⁷ The second group of articles deals with various applications of big data analytics for knowledge discovery and predictive modeling in the field of precision oral health.⁴⁸ This area of research has been growing in geometric progression demonstrating significant potential in identifying previously unrecognized subpopulations and approaches for oral health prevention and management.⁴⁹ The third cohort of articles represent various innovative solutions based on precision data analytics that address challenging problems in dental public health.⁷

Where to gather data

Big data has been defined as large data sets that adhere to the four V's: volume (size of the data set), velocity (speed at which data are generated), variety (different types of data being incorporated), and veracity (accuracy of data reported).⁵⁰

Some authors specify two additional V's⁵¹: value (relevance of the data), and variability (evolution and seasonality of diseases). Overall, the wide spectrum of available data sets can be represented by four major data streams.⁵² The first stream is patient-generated data from self-report, wearable devices, ambient data capture, social media, and devices maintained throughout patient homes. The second stream comprises clinical data such as EHRs pathology reports, billing generated by routine care delivery and provider characteristics. The third stream includes biologic data such as genetics, proteomics, microbiome and multi-omics⁵³ The fourth is represented by normative data sets, which are data carefully collected in clinical trials, nationwide observations, and population surveys.^{54,55} The combination of these data streams can be used to fuel knowledge discovery in research in order to deliver personalized care, a concept summarized in Figure 1.

Patient-generated data can come from new devices or applications that capture objective measurements that are regularly being introduced to the market. In the era of "smart" devices, these data have become even more accessible for patients, providers, and researchers alike. A 2018 study used toothbrushes that were connected to a smartphone application in order to track brushing habits such as frequency, duration, and surface coverage.⁵⁶ More advanced systems such as 3-D motion tracking devices and "selfie" tooth brushing video interventions have also been tested in pilot studies in

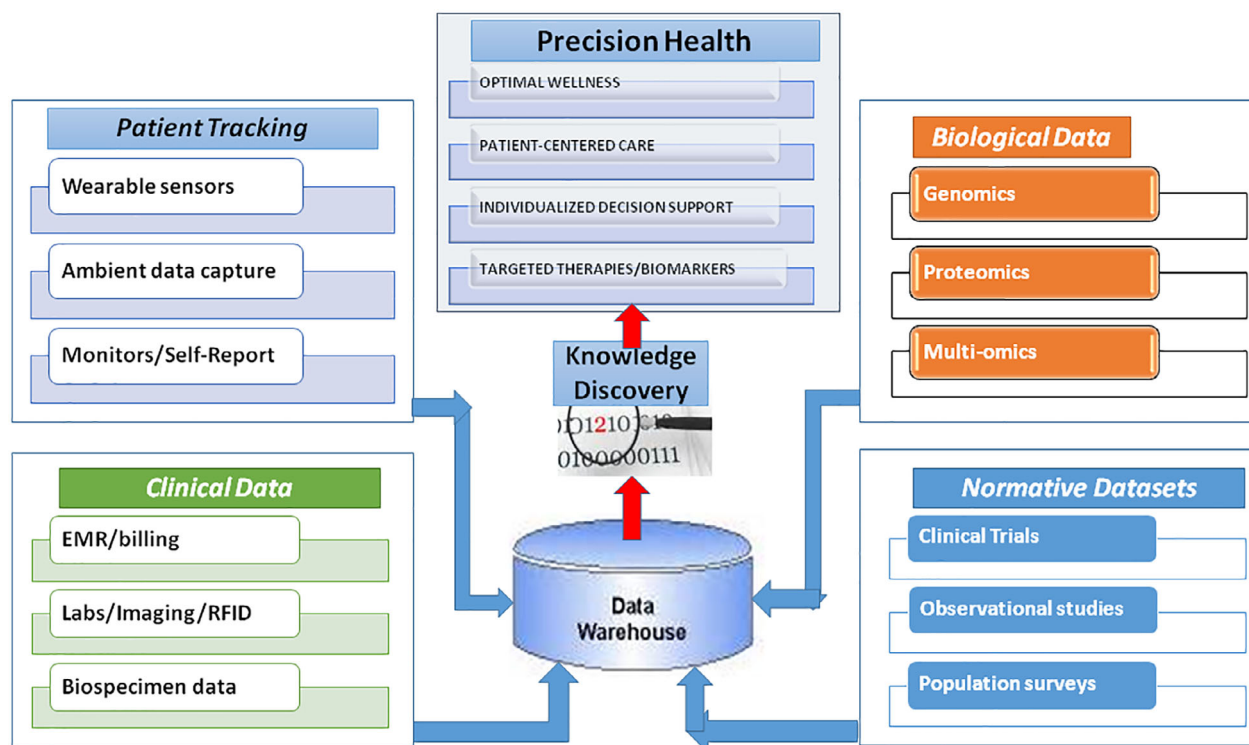


Figure 1 Summary of knowledge discovery process using big data.

order to provide an even more detailed analysis of tooth brushing technique.^{57,58} Currently, the implementation of this data collection method is still in its relative infancy. As these types of technology develop and gain widespread adoption, these data can be analyzed to provide immediate feedback to patients. In addition, potential exists for these data to be documented and uploaded into a central data repository to provide new avenues of research.

Our scoping review identified the development of major data repositories aimed at collecting comprehensive dental data as another important trend in oral health research. As we demonstrated in Results section, in the United States three major initiatives have emerged in recent years as forerunners for the big data trend in oral health. The largest academic initiative is the Consortium for Oral Health-Related Informatics (COHRI), a consortium of members from over 20 dental schools aiming at standardizing oral health data collection and improving informatics utilization in dental education, health care, and research.²⁹ The second data repository, BigMouth, was developed out of this consortium. The BigMouth data repository established the technical foundation and developed a data governance framework for secondary analysis of electronic dental records including patient demographics, diagnoses, medical history, dental history, procedures, odontogram, periodontal chart and treatment provider information.²⁷ The third US-based data repository is the Dental Practice-Based Research Network (DPBRN), which is sponsored by the NIDCR. This research network consists of practicing dentists and academics in the field of oral health who have agreed to share data and conduct research in a collaborative fashion.²⁸ Outside of these data repositories that have been specially developed for academic purposes, clinical data can also come from a wide range of other sources. Data from Medicaid claims, private insurance claims, emergency department records, and many other primary data sources have been used in oral health data analysis studies across the globe.^{33,37,43} Recent studies identified potential limitations of using EHR data generated solely by dental schools as they may represent a biased patient sample.⁵⁹ Similarly, administrative and claims records such as Medicaid data represent a very valuable resource use of which requires understanding of its strengths and limitations.⁶⁰ Future research should capitalize on growing availability of integrated comprehensive data sets representing all facets of oral health.⁴

With the rise of the disciplines of the OMICS (genomics, proteomics, metabolomics, etc.), biological data from in vitro and in vivo sampling has also entered the age of big data. Genomic sequencing of the oral microbiome has already helped identify novel strains of cariogenic bacteria to improve our understanding of the pathogenesis of caries.^{61,62} The accelerating rate of production of next generation sequencing

(NGS) data has led to the development of data repositories for oral microbiome genomics. Repositories such as the Human Oral Microbiome database and the Human Microbiome Consortium have emerged as data aggregate to provide standardized and easily accessible sources of oral microbiome genomic and taxonomic data to researchers and clinicians alike.^{63,64} Genome-wide association studies have also been performed on traditionally at-risk populations such as US Hispanics and children to further identify characteristics that can be used to stratify these patients.^{65,66} However, the relationship between genetics and caries risk is still unclear. Currently, these studies have only found modest associations between gene variants and caries risk. A large number of databases describing genes, proteins, and other biological factors currently exist in order to facilitate research regarding the biological mechanisms of disease.⁶⁷ Conducting more NGS experiments to find links between biological factors and oral disease can improve integration of oral health pathways into these databases and allow bioinformatics research in this field to become more accessible.⁶⁸

Finally, the generation of data from normative data sets such as large clinical trials and national public health surveys represent another important source of data for oral health research at the population level. The randomized clinical trial has long been recognized as the gold standard for clinical investigational research. The data collected from clinical trials have been pivotal in clarifying the efficacy of many therapeutic agents used for caries and periodontal disease control. In addition, data from clinical trials can be analyzed to find a variety of secondary outcomes. For example, the X-ACT clinical trial for xylitol lozenge therapy in adults found that xylitol supplementation did not produce a statistically significant effect on caries reduction.⁶⁹ Beyond this primary finding, baseline data collected from this clinical trial were analyzed to determine risk factors for root caries.⁷⁰ However, in recent years clinical trials in the field of oral health have come under some scrutiny for their low adherence to best reporting practices⁷¹. Prospective registration of clinical trials is necessary to promote study fidelity and data sharing. Unfortunately, some studies of dental clinical trials have found that only around 24–25 percent of clinical trials in the fields of dentistry and orthodontics are prospectively registered.^{72,73} Because of this, the volume of normative data sets available for public oral health research is lacking in some regards. Improved adherence to clinical trial registration and reporting guidelines and promotion of wide data sharing can facilitate the data analysis pipeline for oral health research.⁷⁴

Social media platforms are another source of useful data for oral care delivery. In Japan, researchers set out to determine if social networks among older adults impacted the determinants of oral health. They found that the extent of

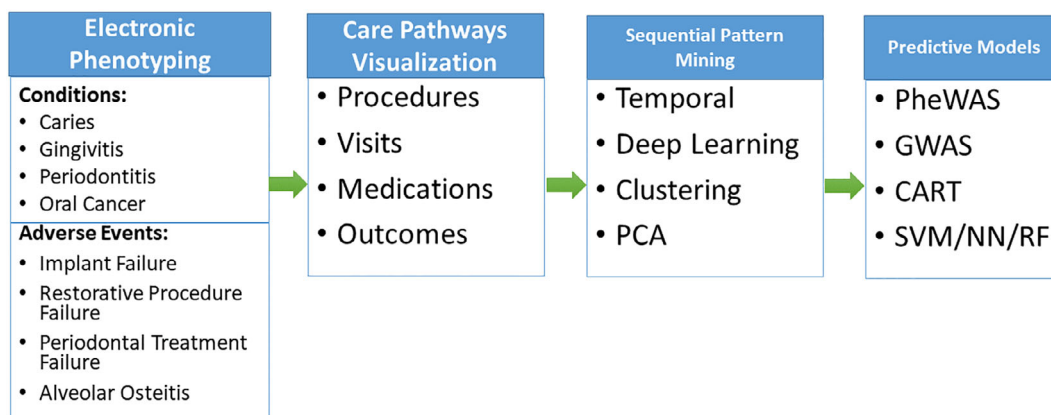


Figure 2 Illustration of knowledge discovery pipeline using electronic data. PCA: principal component analysis; PheWAS: phenome-wide association study; GWAS: genome-wide association study; CART: classification and regression trees; SVM: support vector machine; NN: neural network; RF: random forest.

edentulism in older adults is negatively correlated with the number of social networks.⁷⁵ Another study looked at bullying on Twitter related to dentofacial features and orthodontic treatment. They identified cases of bullying, qualified them, and looked for coping mechanisms the victims and their families had for the mistreatment.⁷⁶ These studies demonstrate the value of data gleaned from social media. Data gathered from social media demonstrates that researchers can measure the positive or negative impact of psychosocial variables on oral health. The data present vast opportunities for tailored interventions both by clinicians and public healthcare professionals.

How to use data

Using the four major data streams discussed above, big data provide four analytic domains that can be combined to deliver optimal precision oral health. In the *clinical domain*, multiple research opportunities exist to study precision dental care delivery tailored to specific patient profiles. Development of real-time decision support tools for individualized diagnosis and treatment planning based on a multitude of relevant factors provided before, during and after the dental encounter will significantly improve the quality of dental care and patient satisfaction. In the *socio-behavioral domain*, identifying oral health risk factors specific to particular population subgroups and delivering targeted preventative interventions using digital media will greatly facilitate individualized oral health on a population level. In the *translational science domain*, research on how the wired digital operatories access and utilize data from outside streams, including a patient's genetic traits and microbiome to facilitate personalized care delivery will be supported by Common Data Models and cross-linked biomedical ontologies such as the Observational Health Data

Science and Informatics (OHDSI) framework.⁷⁷ In the *educational domain*, research on utilizing multiple data streams to better monitor student performance and identify areas for personalized improvement will promote personalized education and individualized student support tailored to individualized performance profile. In Figure 2 below, we illustrate an example of how electronic data can be used to facilitate knowledge discovery and eventually predictive modeling. Electronic phenotyping allows precise identification of specific oral health conditions and syndromes in the presence of data gaps and ambiguity of large heterogeneous oral health data. Care pathways visualization and analytics provides temporal representation of oral healthcare delivery process. Sequential pattern mining identifies characteristic trajectories of dental conditions and allows automated identification of patient subgroups not readily discernible from a manual chart review. Electronic phenotyping results can be correlated with treatments or medications used in the delivery of care in order to provide a holistic picture of each patient. Data mining or other analysis techniques can then be applied to these data to discover new patterns or associations and to identify optimal personalized treatment pathways. Finally, predictive models of outcomes and ideal treatments can be established based on individual patient characteristics. Similar analytical workflow can be applied to identify optimal precision health pathways both for individual patients as well as for unique patient subgroups such as elderly.

The development of clinical decision support systems (CDS) could greatly improve integration of evidence-based dentistry into clinical practice. A recent survey of dental clinicians found that most providers were amenable to the idea of implementing CDS into their daily practice and saw potential for improving quality of care, patient oral health, and other similar benefits.⁷⁸ These tools can give providers summaries

of current evidence for every step of their decision-making process based on analysis of digitally entered patient information or integration with patient EHR.⁷⁹ Development and integration of these tools into routine clinical care is still in progress. A CDS developed in New York provides dental hygienists with assessments and recommendations for screening of chronic, systemic conditions such as hypertension and diabetes.⁸⁰ As vital signs such as blood pressure are measured and recorded, the CDS provides assessments of personalized disease risk and prompts for referral to a specialist in appropriate situations as established by the current hypertension management guidelines.⁸¹ Another CDS was developed by Machado et al. for use in dental trauma management and was found to improve adherence to evidence-based guidelines in both dental students as well as experienced pediatric dentists.⁸² The use of CDS for tobacco abuse screening and interventions has been found to improve both provider adherence to current evidence and patient outcomes.^{83,84}

Data mining techniques have high potential for improving our understanding of particular socio-behavioral risk factors for oral disease. These techniques can be applied to populations that are already considered at-risk in order to identify individual risk factors that can be addressed using targeted public health interventions. This strategy is particularly useful in the field of oral health because the development of caries and periodontal disease can be chronic and insidious in onset, leading to diagnosis in late stages of disease. Yoon et al. used big data and deep learning algorithms on a large sample of Latino patients to identify demographic, behavioral, and psychological factors associated with tooth mobility⁸⁵ and other indicators of oral health status in older adults.⁸⁶ Other studies utilized big data in order to develop risk prediction profiles for development of periodontal disease,^{34,38} implant failure,⁸⁷ peri-implantitis,⁸⁸ and alveolar osteitis.⁸⁹ Using big EHR data, Boehm A et al. uncovered patient determinants of care utilization compliance in a student dental clinic.⁹⁰ In addition to risk prediction of oral disease, research in this field can be used to find novel associations between oral health behavior and systemic health. For example, a study conducted in 2018 found that regular dental visits were independently associated with lower stroke risk.⁹¹ Similar studies can be used to further integrate the fields of medicine and dentistry in order to improve the overall health of the patient. Once individual risk factors have been identified, preventive interventions can be developed in order to increase patient awareness of these risk factors, change health behavior, and improve patient outcomes. An example of this is the My Smile Buddy iPad application, which engages families of at-risk populations to identify health behavior that puts their children at risk of caries development and encourages them to improve their oral health habits.⁹²

Basic science data can be integrated into the clinician workspace so that biomarkers predicting disease risk and therapy outcomes can be identified for optimal treatment planning. The integration of oral microbiome data and genetic testing into clinical decision making can improve diagnostic precision and the risk stratification process. The principle of “deep phenotyping” can also be applied to improve our understanding of disease staging and outcomes. Deep phenotyping refers to “the precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and described” in order to facilitate a more comprehensive understanding of the pathologic basis of disease.⁹³ A study on oral microbial profiles used principal components analysis to determine the microbial profile of healthy patients when compared to chronic and aggressive periodontitis patients.⁹⁴ Deep phenotyping of periodontitis patients can provide new insights into pathologic phenotypic characteristics that are predictive of tooth loss.⁹⁵ Similar principles can also be applied to implant patients in order to develop more comprehensive risk profiles of patients likely to develop implant failure or peri-implantitis.⁹⁶ As we briefly described earlier, there is currently lack of systematic evidence to support the use of genotyping in clinical dental practice.⁹⁷ However, as patient genotyping becomes more widespread and our understanding of the role of genomics in oral health increases, there may be a role in the future for genetics to become another dimension for clinicians to stratify patients.⁶⁸

Finally, data analysis can be implemented in the educational domain in order to improve clinician training for future performance. The application of big data analytics in the area of dental education spans from improving student training at dental schools to continuous professional improvement and to ongoing real-time support via EHR at the point of care. Recent studies demonstrated potential of big data in predicting academic outcomes and professional performance in graduate students.^{98,99} This approach was shown to have implications for optimizing personalized learning and improving student assessments based on individualized feedback provided in a timely fashion.¹⁰⁰ Additionally, use of real life examples drawn from dental EHR in the process of dental education demonstrated promising results in predoctoral³⁹ and postdoctoral training.^{101,102} As evidence-based dentist training is being considered as an essential means for improvements in the quality of patients care,¹⁰³ data analytics workflows for ongoing reporting of dental care quality metrics and providing instructive feedback to dental professionals will be increasingly used in the context of LHS. Implementation of point of care CDS based on individual patient profile and driven by big data analytics has been shown to be an additional vehicle for delivering best clinical practices and supporting ongoing clinician education.¹⁰⁴

While social media can provide valuable data, social marketing is a powerful tool for the promotion of health messages, ones that can target and reduce oral health disparities to subpopulations by changing attitudes, increasing knowledge, and impacting behavior.¹⁰⁵ The Mighty Mouth oral health program, which positions oral health as essential to overall health, is a great example of a program that embraces social marketing. It emphasizes the immediate rewards of good oral health, such as fresher breath, and frames oral health as easy, important, and cost-effective. It is presented as fun and informative, rather than demanding more scientific and informational approach. With big data, oral healthcare providers can develop effective social marketing programs to tailor the message and improve care delivery. Social media can serve as a powerful source of patient-generated data for big data analytics as well as effective media for targeted messaging for oral health promotion.¹⁰⁶

As of the July 2017, Hispanics in the United States accounted for 18.1 percent of the total population.¹⁰⁷ They face major general health and, more specifically, oral health inequities. More than half of Hispanics over the age of 64 will suffer from tooth decay and they are more likely to experience delays in accessing dental care. There has been a widespread call by the greater dental community to address the oral health needs of this population.¹⁰⁸ Hispanic Dental Association convened a workshop of health care providers and other experts to examine the current state of Hispanic oral health research and identify gaps in existing data and research methods. Research and development priorities were outlined by this workshop to better meet the oral healthcare needs of Hispanic patients and to implement standardized, validated instruments using a comprehensive data collection infrastructure.¹⁰⁹ Another disadvantaged population whose oral healthcare needs require urgent attention are older adults.¹¹⁰ In a recent article using cross-sectional analyses of the British Regional Heart Study (BRHS) and the Health, Aging, and Body Composition (HABC) Study,¹¹¹ markers of poor oral health were associated with disability and poor physical function in older populations. The authors proposed further prospective investigations of these associations and underlying pathways. These research initiatives require an increase in population-based studies, social and behavioral sciences, health promotion and communications, gene-environment interactions, and research training and workforce development. Broad inclusion of minority populations and other disadvantaged populations in big data initiatives was identified as a crucial component of addressing oral health inequities.¹¹² By utilizing big data and precision oral health, public health professionals can promote oral health in minority communities and attempt to reduce these disparities.

Three major challenges

A lack of up-to-date evidence-based guidelines has been previously noted in the literature.¹¹³ This paucity of evidence-based practices can be corrected by employing knowledge discovery pipelines revealed through big data analytics and predictive modeling. Knowledge discovered in multiple interconnected databases allows extract useful, nontrivial, and valid patterns from large heterogeneous data sets,¹¹⁴ avoiding the “garbage in garbage out” risk presented by the large quantities of data available. Broadly, this pipeline may entail the following steps: electronic phenotyping→care pathways visualization→sequential pattern mining and predictive models. Electronic phenotyping is essential for big data research using EHRs. This concept is based on the notion that every disease and syndrome has a unique digital signature that can be used to automatically identified cases in EHR even in the presence of erroneous or insufficient coding. This is well illustrated by a dry socket signature in EHR. When a patient is diagnosed with a dry socket, they must have had an extraction first, then returned within a set number of days with pain, open socket, or missing blood clot,^{59,89} with or without other possible symptoms. Even without the proper code for a dry socket, we still have an electronic signature of this event in the record and possible accompanying side effects as well. Electronic phenotyping can be learned by utilizing data science and classifying the exclusive signatures of conditions, adverse events, and procedures. An analytical system can read past what is stated in the note, and with high sensitivity and specificity, recognize the classes of the encounter, creating care pathway visualizations. The next step is employing sequential pattern mining, to ultimately build precise predictive models. Overall, this approach can continue knowledge discovery and a quality improvement cycle.

Implementing these concepts requires verified, integrative, and harmonized data sets, designing intelligent integrative information architecture to manage these data. The systems are integrative because they utilize multiple heterogeneous data sets from basic science, electronic medical records, patients, environments, and so on. It is insufficient to simply merge silos into one data set. It is required that there be an understanding of what the data are and how they are interrelated; in that sense, the data sets must be intelligent. Data sets can only know about selves if the system includes meta-data that explains what is in the data set and what are relationships between internal and external data elements. The underlying systems will understand these relationships using cross-linked biomedical ontologies that will facilitate comprehensive knowledge discovery in a harmonized systematic way.

This process of learning and intervening has already begun, utilizing social media. An innovative example of

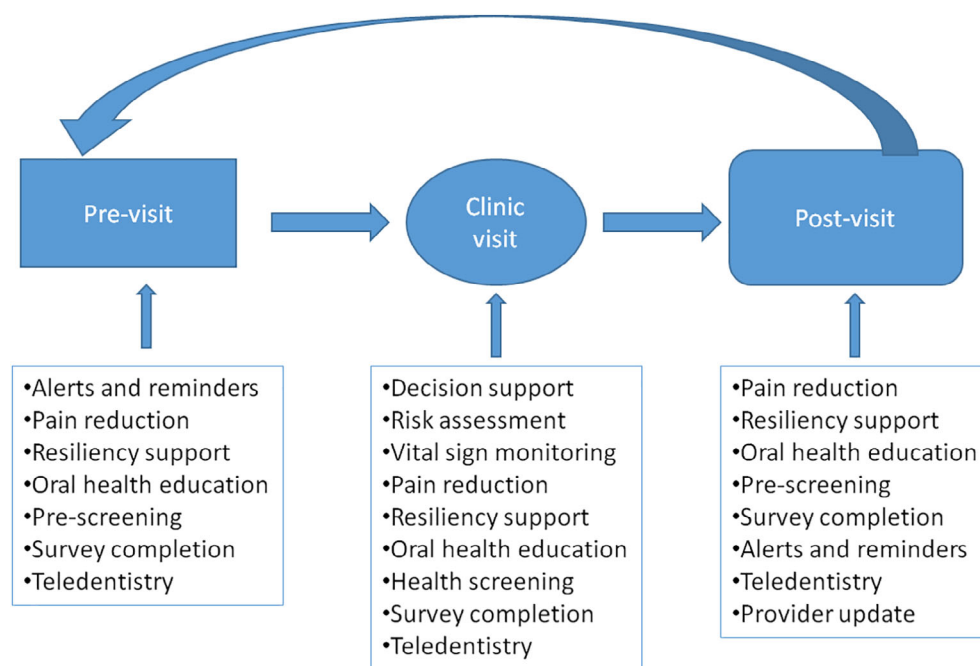


Figure 3 Example workflow of precision medicine integration into clinical practice.

using big data analytics for targeted public health interventions was recently demonstrated by a team at Columbia University that developed a system that identifies foodborne illnesses in NYC restaurants by analyzing Yelp reviews.¹¹⁵ The system utilizes logistic regression trained with bias-adjusted augmented data, which has identified 10 outbreaks and 8,253 complaints of foodborne illness associated with NYC restaurants since 2012. Perhaps the best description of necessary pieces of a learning health care system was given by Dr. Friedman, who chairs the Department of Learning Health Sciences at the University of Michigan School of Medicine in Ann Harbor. He outlines five major points that have been frequently repeated¹¹⁶:

1. One can learn from every patient's characteristics and experiences.
2. Best practice knowledge is immediately available to support decisions.
3. Improvement is continuous through ongoing study.
4. An infrastructure enables this to happen routinely and with economy of scale.
5. All of this is part of the culture.

There is a slow progress in implementing these five major tenets of LHS into dental care.¹¹³ The major prerequisite of successful introduction of sustainable LHS ecosystem in dental care is establishment of comprehensive integrated intelligent big data infrastructure combining multiple heterogeneous data streams related to oral health. This is both a challenge to and impetus for the implementation of precision oral healthcare. Precision oral health is

a cyclical process and once it reaches its action potential, it will be a self-perpetuating system when it is implemented in the context of LHS. It is a cycle of: capturing data actively and passively—organizing data into actionable information—analyzing outcomes—learning new protocols, and cycling back to the first step. From the patient's perspective, the process has begun before the patient has even set foot into the office. In Figure 3, we illustrate the self-perpetuating cycle of precision medicine principles into oral healthcare.

We can potentially use the data to identify a patient's preferences and recognize moments when s/he can be delivered messaging tailored to their individual oral health needs. During the patient's clinic visit, there is decision support, risk assessment, vital sign monitoring, pain reduction, resiliency support, oral health education, health screening, and survey completion. In the clinic, the decision support can be provided based on personalized algorithm analytics that can identify similar patients and understand which procedures were most effective for this group of patients and suggest these treatments at the point of delivery. Post-visit, the process repeats itself, while continually updating the intelligent integrative databases that will then be able to help this individual and others in the future. Overall, this environment would promote precision oral health and allow us to build powerful predictive models and hopefully improve dental care delivery and patient outcomes. This strategy would allow the provider to better employ different translational projects that also

identify underlying pathophysiological determinants of oral health. Through the cyclical nature of learning healthcare systems, dentists will have the potential for feedback rapidly, even in real time to guide care-related decisions.

We must acknowledge the big data limitations originated from dental EHR in developing a precision oral health system. One of which is our current approach to oral health documentation and data capture. The inherent limitation to EHR and other records is their purpose to document the encounter, for communication between providers, and as a legal document.^{117,118} The primary focus of clinical documentation is not research but continuity of guideline-concordant patient care. EHR limitations may lead to miscoded, under-coded, or other issues with documentation of signs and symptoms. These deficiencies can be partially overcome by the electronic phenotyping as previously discussed, but cannot be totally overcome without implementing next generation of intelligent EHR.

Conclusion

The new definition of oral health promulgated by the World Dental Federation¹¹⁹ supports a comprehensive vision that reflects multifaceted interaction between physiological, social, and psychological attributes of dental care that are germane for the quality of life. Intelligent integrative infrastructure for big oral health data will provide a foundational framework for implementing this vision into reality. Following major tenets of a learning healthcare system, big data will be increasingly used to support delivery of precision oral health by determining the best care every time and with each care experience with continuous knowledge discovery. This approach will translate heterogeneous big data into a cycle of care improvement while identifying gaps in care, advancing quality, and improving patient satisfaction and safety. Ultimately, ever-expanding evidence will support personalized patient-centered care, optimal wellness, individualized decision support, and targeted therapies. It can promote creative thinking, dental education and innovation for care delivery and reimbursement models, leading to vast improvements in public's oral health.

References

- Dzau V, Ginsburg G, Chopra A, Goldman D, Green E, Leonard D, McClellan M, Plump A, Terry S, Yamamoto K. Realizing the full potential of precision medicine in health and health care: a vital direction for health and health care. *NAM Persp*. 2016;6(9):1–13. <https://doi.org/10.31478/201609k>.
- Weyant RJ. Evidence-based dentistry: the foundation for modern dental practice. *Dent Clin North Am*. 2019;63(1):ix–x. <https://doi.org/10.1016/j.cden.2018.09.001>.
- Gillette J, Matthews JD, Frantsve-Hawley J, Weyant RJ. The benefits of evidence-based dentistry for the private dental office. *Dent Clin N Am*. 2009;53(1):33–45. <https://doi.org/10.1016/j.cden.2008.09.002>.
- Slavkin HC. From high definition precision healthcare to precision public oral health: opportunities and challenges. *J Public Health Dent*. 2018. <https://doi.org/10.1111/jphd.12296>.
- Joda T, Waltimo T, Pauli-Magnus C, Probst-Hensch N, Zitzmann NU. Population-based linkage of big data in dental research. *Int J Environ Res Public Health*. 2018 Oct 25;15(11):1–5.
- Nanayakkara S, Zhou X, Spallek H. Impact of big data on oral health outcomes. *Oral Dis*. 2019 Jul;25(5):1245–52. <https://doi.org/10.1111/odi.13007>.
- Meyer SL. Toward precision public health. *J Public Health Dent*. 2019. <https://doi.org/10.1111/jphd.12315>.
- Joda T, Waltimo T, Probst-Hensch N, Pauli-Magnus C, Zitzmann NU. Health data in dentistry: an attempt to master the digital challenge. *Public Health Genomics*. 2019; 22(1–2):1–7. <https://doi.org/10.1159/000501643>.
- Kusiak J, Somerman M. Data science at the National Institute of Dental and Craniofacial Research. *J Am Dent Assoc*. 2016;147(8):597–9. <https://doi.org/10.1016/j.adaj.2016.06.005>.
- Stohler C. Editorial: the next frontier: digital disease detection in cyberspace. *J Oral Facial Pain Headache*. 2014;28(2):105. <https://doi.org/10.11607/ofph.2014.2.e>.
- Stohler C. Editorial: orofacial pain and the prospects of precision medicine. *J Oral Facial Pain Headache*. 2015;29(4):321.
- Smiley CJ, Tracy SL, Abt E, Michalowicz BS, John MT, Gunsolley J. Systematic review and meta-analysis on the nonsurgical treatment of chronic periodontitis by means of scaling and root planing with or without adjuncts. *Br Dent J*. 2015;219(5):215–5. <https://doi.org/10.1038/sj.bdj.2015.684>.
- Canadian Agency for Drugs and Technologies in Health. *Dental scaling and root planing for periodontal health: a review of the clinical effectiveness, cost-effectiveness, and guidelines*. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health; 2016.
- Beil H, Preisser JS, Rozier RG. Accuracy of record linkage software in merging dental administrative data sets. *J Public Health Dent*. 2012;73(2):89–93. <https://doi.org/10.1111/j.1752-7325.2012.00343.x>.
- Liu K, Acharya A, Alai S, Schleyer T. Using electronic dental record data for research. *J Dent Res*. 2013;92(7):S90–6. <https://doi.org/10.1177/0022034513487560>.
- Marshall TA, Straub-Morarend CL, Qian F, Finkelstein MW. Perceptions and practices of dental school faculty regarding evidence-based dentistry. *J Dent Educ*. 2013;77(2):146–51.
- Afrashtehfar KI, Mansour KA. From dental science to clinical practice: knowledge translation and evidence-based

- dentistry principles. *Saudi Dental J.* 2017;**29**(3):83–92. <https://doi.org/10.1016/j.sdentj.2017.02.002>.
18. Newlands R, Duncan EM, Prior M, Elouafkaoui P, Elders A, Young L, Clarkson JE, Ramsay CR. Translation research in a dental setting (TRiaDS) research methodology group. Barriers and facilitators of evidence-based management of patients with bacterial infections among general dental practitioners: a theory-informed interview study. *Implement Sci.* 2016;**11**(11):1–12. <https://doi.org/10.1186/s13012-016-0372-z>.
 19. Meursinge Reynders R, Ronchi L, Ladu L, Di Girolamo N, de Lange J, Roberts N, Mickan S. Barriers and facilitators to the implementation of orthodontic mini-implants in clinical practice: a protocol for a systematic review and meta-analysis. *Syst Rev.* 2016;**5**(5):22. <https://doi.org/10.1186/s13643-016-0198-4>.
 20. Budrionis A, Bellika JG. The learning healthcare system: where are we now? A systematic review. *J Biomed Inform.* 2016;**64**:87–92. <https://doi.org/10.1016/j.jbi.2016.09.018> Epub 2016 Sep 28.
 21. Engineering a Learning Healthcare System. *A look at the future: workshop summary*. Washington, DC: The National Academies Press; 2011. <https://doi.org/10.17226/12213>.
 22. Mandl KD, Kohane IS, McFadden D, Weber GM, Natter M, Mandel J, Schneeweiss S, Weiler S, Klann JG, Bickel J, Adams WG, Ge Y, Zhou X, Perkins J, Marsolo K, Bernstam E, Showalter J, Quarshie A, Ofili E, Hripcsak G, Murphy SN. Scalable collaborative infrastructure for a learning healthcare system (SCILHS): architecture. *J Am Med Inform Assoc.* 2014 Jul-Aug;**21**(4):615–20. <https://doi.org/10.1136/amiainl-2014-002727>.
 23. Houry MJ, Galea S. Will precision medicine improve population health? *JAMA.* 2016;**316**(13):1357–8. <https://doi.org/10.1001/jama.2016.12260>.
 24. Dolley S. Big Data's role in precision public health. *Front Public Health.* 2018;**6**:68. <https://doi.org/10.3389/fpubh.2018.00068>.
 25. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci.* 2010 Sep 20;**5**:69. <https://doi.org/10.1186/1748-5908-5-69>.
 26. von Bültzingslöwen I, Östholm H, Gahnberg L, Ericson D, Wennström JL, Paulander J. Swedish quality registry for caries and periodontal diseases - a framework for quality development in dentistry. *Int Dent J.* 2019 Oct;**69**(5):361–8. <https://doi.org/10.1111/idj.12481>.
 27. Walji MF, Kalenderian E, Stark PC, White JM, Kookal KK, Phan D, Tran D, Bernstam EV, Ramoni R. BigMouth: a multi-institutional dental data repository. *J Am Med Inform Assoc.* 2014;**21**(6):1136–40. <https://doi.org/10.1136/amiainl-2013-002230>.
 28. Gilbert GH, Williams OD, Korelitz JJ, Fellows JL, Gordan VV, Makhija SK, Meyerowitz C, Oates TW, Rindal DB, Benjamin PL, Foy PJ, National Dental PBRN collaborative group. Purpose, structure, and function of the United States National Dental Practice-Based Research Network. *J Dent.* 2013 Nov;**41**(11):1051–9. <https://doi.org/10.1016/j.jdent.2013.04.002>.
 29. Stark PC, Kalenderian E, White JM, Walji MF, Stewart DCL, Kimmes N. Consortium for Oral health-related informatics: improving dental research, education, and treatment. *J Dent Educ.* 2010;**74**(10):1051–65.
 30. Rao DP, McFaull S. Tooth 'aches': injuries related to toothbrush use. *Paediatr Child Health.* 2019;**24**(1):e40–4. <https://doi.org/10.1093/pch/pxy073> Epub 2018 May 31.
 31. Suni J, Vähänikkilä H, Pääkilä J, Tjäderhane L, Larmas M. Review of 36,537 patient records for tooth health and longevity of dental restorations. *Caries Res.* 2013;**47**(4):309–17. <https://doi.org/10.1159/000346691>.
 32. Käkilehto T, Salo S, Larmas M. Data mining of clinical oral health documents for analysis of the longevity of different restorative materials in Finland. *Int J Med Inform.* 2009 Dec;**78**(12):e68–74. <https://doi.org/10.1016/j.ijmedinf.2009.04.004>.
 33. Raedel M, Hartmann A, Bohm S, Priess HW, Samietz S, Konstantinidis I, Walter MH. Four-year outcomes of restored posterior tooth surfaces—a massive data analysis. *Clin Oral Investig.* 2017;**21**(9):2819–25. <https://doi.org/10.1007/s00784-017-2084-4>.
 34. Lee JH, Jeong SN, Choi SH. Predictive data mining for diagnosing periodontal disease: the Korea National Health and nutrition examination surveys (KNHANES V and VI) from 2010 to 2015. *J Public Health Dent.* 2018;**79**(1):44–52. <https://doi.org/10.1111/jphd.12293>.
 35. Chan C, You H, Lian H, Huan C. Patients receiving comprehensive periodontal treatment have better clinical outcomes than patients receiving conventional periodontal treatment. *J Formos Med Assoc.* 2016;**115**:152e162. <https://doi.org/10.1016/j.jfma.2015.10.017>.
 36. Su WW, Su CW, Chang DC, Chuang SL, Chen SL, Hsu CY, Yen AM, Chiu SY, Fann JC, Lee YH, Jeng YC, Lee YC, Chiu HM, Chen TH, Wang CP, Chen MK. Impact of varying anatomic sites on advanced stage and survival of oral cancer: 9-year prospective cohort of 27 717 cases. *Head Neck.* 2019 May;**41**(5):1475–83. <https://doi.org/10.1002/hed.25579>.
 37. Nalliah RP, Da Silva JD, Allareddy V. The characteristics of hospital emergency department visits made by people with mental health conditions who had dental problems. *J Am Dent Assoc.* 2013;**144**(6):617–24.
 38. Thyvalikakath TP, Padman R, Vyawahare K, Darade P, Paranjape R. Utilizing dental electronic health records data to predict risk for periodontal disease. *Stud Health Technol Inform.* 2015;**216**:1081.
 39. Rai NK, Carey C, Brunson D, Tiwari T. Increasing dental students' understanding of population surveillance through data mining. *J Dent Educ.* 2019 Mar;**83**(3):281–6. <https://doi.org/10.21815/JDE.019.031>.
 40. Filker PJ, Cook N, Kodish-Stav J. Electronic health records: a valuable tool for dental school strategic planning. *J Dent Educ.* 2013 May;**77**(5):591–7.

41. Boland MR, Hripcsak G, Albers DJ, Wei Y, Wilcox AB, Wei J, Li J, Lin S, Breene M, Myers R, Zimmerman J, Papapanou PN, Weng C. Discovering medical conditions associated with periodontitis using linked electronic health records. *J Clin Periodontol*. 2013 May;**40**(5):474–82. <https://doi.org/10.1111/jcpe.12086>.
42. Kalenderian E, Tokede B, Ramoni R, Khan M, Kimmes N, White J, Vaderhobli R, Yansane A, Feilzer A, Walji M. Dental clinical research: an illustration of the value of standardized diagnostic terms. *J Public Health Dent*. 2016; **76**(2):152–6. <https://doi.org/10.1111/jphd.12124>.
43. Tiwari T, Rai N, Brow A, Tranby EP, Boynes SG. Association between medical well-child visits and dental preventive visits: a big data report. *JDR Clin Trans Res*. 2019 Jul;**4**(3):239–45.
44. Huber J, Woods T, Fushi A, Duong MT, Eidelman AS, Zalal AR, Urquhart O, Colangelo E, Quinn S, Carrasco-Labra A. Social media research strategy to understand clinician and public perception of health care messages. *JDR Clin Trans Res*. 2019. <https://doi.org/10.1177/2380084419849439>.
45. Helmi M, Spinella MK, Seymour B. Community water fluoridation online: an analysis of the digital media ecosystem. *J Public Health Dent*. 2018;**78**(4):296–305. <https://doi.org/10.1111/jphd.12268>.
46. Liu K, Acharya A, Alai S, Schleyer TK. Using electronic dental record data for research: a data-mapping study. *J Dent Res*. 2013;**92**(7 Suppl):90S–6S.
47. Shah T, Rabhi FA, Ray PK. Investigating an ontology-based approach for big data analysis of inter-dependent medical and oral health conditions. *Cluster Comput*. 2014;**18**:351–67. <https://doi.org/10.1007/s10586-014-0406-8>.
48. Shetty V, Yamamoto J, Yale K. Re-architecting oral healthcare for the 21st century. *J Dent*. 2018;**74**(Suppl 1): S10–4. <https://doi.org/10.1016/j.jdent.2018.04.017>.
49. Polverini PJ, Krebsbach PH. Research and discovery science and the future of dental education and practice. *J Dent Educ*. 2017;**81**(9):eS97–eS107. <https://doi.org/10.21815/JDE.017.040>.
50. Lee CH, Yoon H-J. Medical big data: promise and challenges. *Kidney Res Clin Pract*. 2017;**36**(1):3–11.
51. Glick M. Taking a byte out of big data. *J Am Dent Assoc*. 2015 Nov;**146**(11):793–4.
52. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013;**309**(13):1351–2. <https://doi.org/10.1001/jama.2013.393>.
53. Yu X-T, Zeng T. Integrative analysis of Omics big data. *Methods Mol Biol*. 2018;**1754**:109–35. https://doi.org/10.1007/978-1-4939-7717-8_7.
54. O'Connor PJ. Normative data: their definition, interpretation, and importance for primary care physicians. *Fam Med*. 1990;**22**(4):307–11.
55. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014;**2**(1):3.
56. McKenzie KW, Pretty IA. Connected toothbrushes: bridging the gap to personalized oral health. *Am J Dent*. 2018;**31**(3):115–20.
57. Rooban T, Vernon L, Kumar PM, Mohandoss A, Walls T. Using smartphone video “selfies” to monitor change in toothbrushing behavior after a brief intervention: a pilot study. *Indian J Dent Res*. 2016;**27**(3):268–77. <https://doi.org/10.4103/0970-9290.186241>.
58. Inada E, Saitoh I, Yu Y, Tomiyama D, Murakami D, Takemoto Y, Morizono K, Iwasaki T, Iwase Y, Yamasaki Y. Quantitative evaluation of toothbrush and arm-joint motion during tooth brushing. *Clin Oral Investig*. 2014;**19**(6):1451–62. <https://doi.org/10.1007/s00784-014-1367-2>.
59. Levitin SA, Grbic JT, Finkelstein J. Completeness of electronic dental Records in a student clinic: retrospective analysis. *JMIR Med Inform*. 2019 Mar 21;**7**(1):e13008.
60. Riley GF. Administrative and claims records as sources of health care cost data. *Med Care*. 2009 Jul;**47**(7 Suppl 1): S51–5. <https://doi.org/10.1097/MLR.0b013e31819c95aa>.
61. Huang X, Palmer SR, Ahn S-J, Richards VP, Williams ML, Nascimento MM, Burne RA. A highly *Arginolytic streptococcus* species that potently antagonizes *Streptococcus mutans*. *Appl Environ Microbiol*. 2016;**82**(7): 2187–201. <https://doi.org/10.1128/aem.03887-15>.
62. Camelo-Castillo A, Benitez-Paez A, Belda-Ferre P, Cabrera-Rubio R, Mira A. *Streptococcus dentisani* sp. nov., a novel member of the mitis group. *Int J Syst Evol Microbiol*. 2014;**64**(Pt 3):1073–3. <https://doi.org/10.1099/ij.s.0.062802-0>.
63. Chen T, Yu W-H, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)*. 2010;**2010**:1–10. <https://doi.org/10.1093/database/baq013>.
64. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, Nelson KE, White O, Methé BA, Huttenhower C. The human microbiome project: a community resource for the healthy human microbiome. *PLoS Biol*. 2012;**10**(8):e1001377. <https://doi.org/10.1371/journal.pbio.1001377>.
65. Morrison J, Laurie CC, Marazita ML, Sanders AE, Offenbacher S, Salazar CR. Genome-wide association study of dental caries in the Hispanic communities health study/study of Latinos (HCHS/SOL). *Hum Mol Genet*. 2016;**25**(4):807–16.
66. Haworth S, Shungin D, van der Tas JT, Vucic S, Medina-Gomez C, Yakimov V. Consortium-based genome-wide meta-analysis for childhood dental caries traits. *Hum Mol Genet*. 2018;**27**(17):3113–27.
67. Zou D, Ma L, Yu J, Zhang Z. Biological databases for human research. *Genomics Proteomics Bioinformatics*. 2015;**13**(1):55–63.
68. Zhang F, Finkelstein J. The relationship between single nucleotide polymorphisms and dental implant loss: a

- scoping review. *Clin Cosmet Investig Dent*. 2019;**11**: 131–41. <https://doi.org/10.2147/CCIDE.S207445>.
69. Bader JD, Vollmer WM, Shugars DA, Gilbert GH, Amaechi BT, Brown JP, Laws RL, Funkhouser KA, Makhija SK, Ritter AV, Leo MC. Results from the xylitol for adult caries trial (X-ACT). *J Am Dent Assoc*. 2013;**144**(1):21–30. <https://doi.org/10.14219/jada.archive.2013.0010>.
 70. Ritter AV, Preisser JS, Chung Y, Bader JD, Shugars DA, Amaechi BT, Makhija SK, Funkhouser KA, Vollmer WM, X-ACT Collaborative Research Group. Risk indicators for the presence and extent of root caries among caries-active adults enrolled in the xylitol for adult caries trial (X-ACT). *Clin Oral Investig*. 2011;**16**(6):1647–57. <https://doi.org/10.1007/s00784-011-0656-2>.
 71. Giannobile W. Improving clinical trials in dentistry. *J Dent Res*. 2015;**94**(3):6S–7S. <https://doi.org/10.1177/0022034515569308>.
 72. Papageorgiou SN, Antonoglou GN, Sándor GK, Eliades T. Randomized clinical trials in orthodontics are rarely registered a priori and often published late or not at all. *Plos One*. 2017;**12**(8):e0182785. <https://doi.org/10.1371/journal.pone.0182785>.
 73. Smaïl-Faugeron V, Fron-Chabouis H, Durieux P. Clinical trial registration in oral health journals. *J Dent Res*. 2014;**94**(3):8S–13S. <https://doi.org/10.1177/0022034514552492>.
 74. Institute of Medicine. *Sharing clinical research data: workshop summary*. Washington, DC: The National Academies Press; 2013. <https://doi.org/10.17226/18267>.
 75. Aida J, Kondo K, Yamamoto T, Saito M, Ito K, Suzuki K, Osaka K, Kawachi I. Is social network diversity associated with tooth loss among older Japanese adults? *PLOS ONE*. 2016;**11**(7):e0159970. <https://doi.org/10.1371/journal.pone.0159970>.
 76. Chan A, Antoun J, Morgaine K, Farella M. Accounts of bullying on twitter in relation to dentofacial features and orthodontic treatment. *J Oral Rehabil*. 2017;**44**(4):244–50. <https://doi.org/10.1111/joor.12487>.
 77. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;**216**:574–8.
 78. Mertz E, Bolarinwa O, Wides C, Gregorich S, Simmons K, Vaderhobli R, White J. Provider attitudes toward the implementation of clinical decision support tools in dental practice. *J Evid Based Dent Pract*. 2015;**15**(4):152–63. <https://doi.org/10.1016/j.jebdp.2015.09.001>.
 79. Mendonca E. Clinical decision support systems: perspectives in dentistry. *J Dent Educ*. 2004;**68**(6):589–97.
 80. Russell SL, Greenblatt AP, Gomes D, Birenz S, Golembeski CA, Shelley D, McGuirk M, Eisenberg E, Northridge ME. Toward implementing primary care at chairside: developing a clinical decision support system for dental hygienists. *J Evid Based Dent Pract*. 2015;**15**(4): 145–51. <https://doi.org/10.1016/j.jebdp.2015.08.003>.
 81. Abel N, Contino K, Jain N, Grewal N, Grand E, Hagans I, Hunter K, Roy S. Eighth joint National Committee (JNC-8) guidelines and the outpatient management of hypertension in the African-American population. *N Am J Med Sci*. 2015;**7**(10):438–45. <https://doi.org/10.4103/1947-2714.168669>.
 82. Machado JP, Lam XT, Chen J-W. Use of a clinical decision support tool for the management of traumatic dental injuries in the primary dentition by novice and expert clinicians. *Dent Traumatol*. 2018;**34**(2):120–8. <https://doi.org/10.1111/edt.12390>.
 83. Montini T, Schenkel AB, Shelley DR. Feasibility of a computerized clinical decision support system for treating tobacco use in dental clinics. *J Dent Educ*. 2013;**77**(4): 458–62.
 84. Rindal DB, Rush WA, Schleyer TK, Kirshner M, Boyle RG, Thoele MJ, Asche SE, Thyvalikakath T, Spallek H, Durand EC, Enstad CJ, Huntley CL. Computer-assisted guidance for dental office tobacco-cessation counseling. *Am J Prev Med*. 2013;**44**(3):260–4. <https://doi.org/10.1016/j.amepre.2012.10.023>.
 85. Yoon S, Odlum M, Lee Y, Choi T, Khronish IM, Davidson KW, Finkelstein J. Applying deep learning to understand predictors of tooth mobility among urban latinos. *Stud Health Technol Inform*. 2018;**251**:241–4.
 86. Yoon S, Choi T, Odlum M, Mitchell DA, Kronish IM, Davidson KW, Finkelstein J. Machine learning to identify behavioral determinants of oral health in inner city older hispanic adults. *Stud Health Technol Inform*. 2018;**251**:253–6.
 87. Hickin MP, Shariff JA, Jennette PJ, Finkelstein J, Papapanou PN. Incidence and determinants of dental implant failure: a review of electronic health records in a U.S. dental school. *J Dent Educ*. 2017;**81**(10):1233–42.
 88. Kordbacheh Changi K, Finkelstein J, Papapanou PN. Peri-implantitis prevalence, incidence rate, and risk factors: a study of electronic health records at a U.S. dental school. *Clin Oral Implants Res*. 2019 Apr;**30**(4):306–14. <https://doi.org/10.1111/clr.13416> Epub 2019 Mar 10.
 89. Levitin SA, Jeong IC, Finkelstein J. Mining electronic dental records to identify dry socket risk factors. *Stud Health Technol Inform*. 2019 Jul 4;**262**:328–31.
 90. Boehm A, Jeong IC, Finkelstein J, Whalen S, Graham R. Using big data to uncover patient determinants of care utilization compliance in a student dental clinic. *Stud Health Technol Inform*. 2019 Jul 4;**262**:324–7.
 91. Sen S, Giamberardino LD, Moss K, Morelli T, Rosamond WD, Gottesman RF, Beck J, Offenbacher S. Periodontal disease, regular dental care use, and incident ischemic stroke. *Stroke*. 2018;**49**(2):355–62. <https://doi.org/10.1161/strokeaha.117.018990>.
 92. Chinn CH, Levine J, Matos S, Findley S, Edelstein BL. An Interprofessional collaborative approach in the development of a caries risk assessment mobile tablet application: my smile buddy. *J Health Care Poor Underserved*. 2013;**24**(3):1010–20. <https://doi.org/10.1353/hpu.2013.0114>.

93. Robinson PN. Deep phenotyping for precision medicine. *Hum Mut.* 2012;**33**(5):777–80. <https://doi.org/10.1002/humu.22080>.
94. Feres M, Louzoun Y, Haber S, Faveri M, Figueiredo LC, Levin L. Support vector machine-based differentiation between aggressive and chronic periodontitis using microbial profiles. *Int Dent J.* 2017;**68**(1):39–46. <https://doi.org/10.1111/idj.12326>.
95. Morelli T, Moss KL, Preisser JS, Beck JD, Divaris K, Wu D, Offenbacher S. Periodontal profile classes predict periodontal disease progression and tooth loss. *J Periodontol.* 2018;**89**(2):148–56. <https://doi.org/10.1002/jper.17-0427>.
96. Papantonopoulos G, Gogos C, Housos E, Bountis T, Loos BG. Prediction of individual implant bone levels and the existence of implant “phenotypes”. *Clin Oral Implants Res.* 2016;**28**(7):823–32. <https://doi.org/10.1111/clr.12887>.
97. Kornman K, Polverini P. Clinical application of genetics to guide prevention and treatment of oral diseases. *Clin Genet.* 2014;**86**(1):44–9. <https://doi.org/10.1111/cge.12396>.
98. Adam J, Bore M, Childs R, Dunn J, Mckendree J, Munro D, Powis D. Predictors of professional behaviour and academic outcomes in a UK medical school: a longitudinal cohort study. *Med Teach.* 2015;**37**(9):868–80.
99. Niessen AS, Meijer RR, Tendeiro JN. Predicting performance in higher education using proximal predictors. *PLoS One.* 2016 Apr 13;**11**(4):e0153663.
100. Bill Cope B, Kalantzis M. Big data comes to school: implications for learning, assessment, and research. *AERA Open.* 2016 April-June;**2**(2):1–19. <https://doi.org/10.1177/2332858416641907>.
101. Schleyer TK, Corby P, Gregg AL. A preliminary analysis of the dental informatics literature. *Adv Dent Res.* 2003 Dec;**17**:20–4.
102. Song M, Liu K, Abromitis R, Schleyer TL. Reusing electronic patient data for dental clinical research: a review of current status. *J Dent.* 2013 Dec;**41**(12):1148–63.
103. Weyant RJ. Teaching evidence-based practice: considerations for dental education. *Dent Clin North Am.* 2019 Jan;**63**(1):97–117.
104. Gordon SM, Camargo GA, Mejia GC, Sutherland JN. Use of the dental electronic health record for research: assessing demographic and oral health characteristics data for clinic patients. *J Dent Educ.* 2018 Dec;**82**(12):1249–57.
105. Brocklehurst P, Morris P, Tickle M. Social marketing: an appropriate strategy to reduce oral health inequalities? *Int J Health Promot Educ.* 2012;**50**(2):81–91–1. <https://doi.org/10.1080/14635240.2012.661969>.
106. Lin K, Finkelstein J. Patient empowerment in online support group for temporomandibular disorder. *Proc of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 2017;**2017**:1569–74. <https://doi.org/10.1109/BIBM.2017.8217897>.
107. Hispanic Heritage Month 2018. <https://www.census.gov/newsroom/facts-for-features/2018/hispanic-heritage-month.html>. Published September 2018. Accessed January 1, 2019.
108. Han C. Oral health disparities: racial, language and nativity effects. *SSM Popul Health.* 2019;**8**:100436. <https://doi.org/10.1016/j.ssmph.2019.100436>.
109. Ramos-Gomez F, Cruz GD, Watson MR, Canto MT, Boneta AE. Latino oral health: a research agenda toward eliminating oral health disparities. *J Am Dent Assoc.* 2005;**136**(9):1231–40.
110. Razak PA, Richard KMJ, Thankachan RP, Hafiz KAA, Kumar KN, Sameer KM. Geriatric oral health: a review article. *J Int Oral Health.* 2014;**6**(6):110–6.
111. Kotronia E, Wannamethee SG, Papacosta AO, Whincup PH, Lennon LT, Visser M, Weyant RJ, Harris TB, Ramsay SE. Oral health, disability and physical function: results from studies of older people in the United Kingdom and United States of America. *J Am Med Dir Assoc.* 2019;**20**(12):1654.e1–1654.e9. <https://doi.org/10.1016/j.jamda.2019.06.010>.
112. Gansky SA, Shafik S. At the crossroads of oral health inequities and precision public health. *J Public Health Dent.* 2019. <https://doi.org/10.1111/jphd.12316>.
113. Spallek H, Song M, Polk D, Bekhuis T, Frantsve-Hawley J, Aravamudhan K. Barriers to implementing evidence-based clinical guidelines: a survey of early adopters. *J Evid Based Dent Pract.* 2010;**10**(4):195–206. <https://doi.org/10.1016/j.jebdp.2010.05.013>.
114. Domingo-Ferrer J, Kriegel H, Schubert M, Embley D, Li X. KDD pipeline. *Encyclopedia of Database Systems.* 2009;**2009**:1586–7. https://doi.org/10.1007/978-0-387-39940-9_1134.
115. Effland T, Lawson A, Balter S, Devinney K, Reddy V, Waechter H, Gravano L, Hsu D. Discovering foodborne illness in online restaurant reviews. *J Am Med Inform Assoc.* 2018;**25**(12):1586–92. <https://doi.org/10.1093/jamia/ocx093>.
116. Friedman CP, Allee NJ, Delaney BC, Flynn AJ, Silverstein JC, Sullivan K, Young KA. The science of learning health systems: foundations for a new journal. *Learn Health Syst.* 2016;**1**(1):e10020.
117. Ho Y, Gadd C, Kohorst K, Rosenbloom S. A qualitative analysis evaluating the purposes and practices of clinical documentation. *Appl Clin Inform.* 2014;**5**(1):153–68. <https://doi.org/10.4338/ACI-2013-10-RA-0081>.
118. Mangalmurti S, Murtagh L, Mello M. Medical malpractice liability in the age of electronic health records. *Surv Anesthesiol.* 2011;**55**(6):317–9. <https://doi.org/10.1097/sa.0b013e3182379506>.
119. Glick M, Williams DM, Kleinman DV, Vujicic M, Watt RG, Weyant RJ. A new definition for oral health developed by the FDI World Dental Federation opens the door to a universal definition of oral health. *J Public Health Dent.* 2017 Dec;**77**(1):3–5.

How to cite this article: Finkelstein J, Zhang F, Levitin SA, Cappelli D. Using big data to promote precision oral health in the context of a learning healthcare system. *Journal of Public Health Dentistry.* 2020;**80**:S43–S58. <https://doi.org/10.1111/jphd.12354>