

Leveraging Electronic Dental Record data for clinical research: Proof-of-concept study

NIDCR Protocol Number: 15-023

NIDCR Grant Number: U19-DE-22516

NIDCR Grant Principal Investigator (GPI):

Gregg H. Gilbert, DDS, MBA

Study Principal Investigator:

Thankam Paul Thyvalikakath, DMD, MDS, PhD

Co-Investigators:

Titus Schleyer, DMD, PhD

William Duncan, PhD

Institution:

Indiana University School of Dentistry, Regenstrief Institute

NIDCR Program Officials:

Dena Fischer, DDS, MSD, MS

Version Number: 2.0

09 February 2017

STATEMENT OF COMPLIANCE

The study will be conducted in accordance with the Code of Federal Regulations on the Protection of Human Subjects (45 CFR Part 46), and the NIDCR Clinical Terms of Award. All personnel involved in the conduct of this study have completed human subjects protection training.

SIGNATURE PAGE

The signature below constitutes the approval of this protocol and the attachments, and provides the necessary assurances that this study will be conducted according to all stipulations of the protocol, including all statements regarding confidentiality, and according to local legal and regulatory requirements and applicable US federal regulations and guidelines.

Grant Principal Investigator/South Central Regional Director:

Signed: _____ Date: _____

Name: Gregg H. Gilbert, DDS, MBA

Title: Professor and Chair

Study Principal Investigator:

Signed: _____ Date: _____

Name: Thankam Paul Thyvalikakath, DMD, MDS, PhD

Title: Associate Professor & Director of Dental Informatics

Study Co-Investigators:

Signed: _____ Date: _____

Name: Titus Schleyer, DMD, PhD

Title: Director, Center for Biomedical Informatics

Signed: _____ Date: _____

Name: William Duncan, PhD

Title: Assistant Research Scientist

Regional Directors:

Signed: _____

Date: _____

Name: Jeffrey Fellows

Title: Regional Director, Western Region

Signed: _____

Date: _____

Name: Brad Rindal

Title: Regional Director, Midwest Region

Signed: _____

Date: _____

Name: Tom Oates

Title: Regional Director, Southwest Region

Signed: _____

Date: _____

Name: Valeria Gordan

Title: Regional Director, South Atlantic Region

Signed: _____

Date: _____

Name: Cyril Meyerowitz

Title: Regional Director, Northeast Region

SIGNATURE PAGE- NETWORK STAFF

A copy of this page is to be signed by all Steering Committee members, Regional Coordinators, and other National Dental PBRN staff members responsible for conducting any portion of the study (if not already designated to sign the protocol above). The signature page should be printed, signed, then scanned into a PDF document and submitted to the Coordinating Center (NDPBRN-helpdesk@westat.com) for storage on the Internal Website.

The signature below constitutes:

- 1) acknowledgement of having read this protocol version (as indicated in the upper right corner of this page) and the attachments, and
- 2) an assurance that this individual will conduct all of his or her assigned study tasks according to all stipulations of the protocol, including all statements regarding confidentiality, and according to local legal and regulatory requirements and applicable US federal regulations and guidelines.
- 3) an assurance that this individual will read and follow all study plans applicable to his/her role on the study (e.g. Regional Coordinators will read and follow the Practice Reference Sheet and other applicable plans developed in the future).

Signed: _____ Date: _____

Name: _____

Title: _____

TABLE OF CONTENTS

	PAGE
STATEMENT OF COMPLIANCE	1
SIGNATURE PAGE	2
SIGNATURE PAGE- NETWORK STAFF	4
TABLE OF CONTENTS	5
LIST OF ABBREVIATIONS	7
PROTOCOL SUMMARY	8
1 KEY ROLES AND CONTACT INFORMATION.....	2
2 INTRODUCTION: BACKGROUND INFORMATION AND SCIENTIFIC RATIONALE.....	5
2.1 Background Information	5
2.2 Rationale	6
2.3 Potential Risks and Benefits.....	7
2.3.1 Potential Risks.....	7
2.3.2 Potential Benefits.....	8
3 OBJECTIVES.....	9
3.1 Study Objectives	9
3.2 Study Outcome Measures.....	9
4 STUDY DESIGN	12
4.1 Overview of data assessment, cleaning and transformation.....	12
4.2 Ramp-up Phase	13
5 STUDY ENROLLMENT AND WITHDRAWAL	14
5.1 Practitioner Inclusion Criteria	14
5.2 Strategies for Recruitment and Retention	15
5.2.1 Practitioner Recruitment and Retention.....	15
5.3 Withdrawal	15
5.3.1 Reasons for Withdrawal	15
5.3.2 Handling of Withdrawals.....	15
5.4 Premature Termination or Suspension of Study	15
6 STUDY SCHEDULE	17
6.1 Practice Screening/Enrollment (Stage 1)	17
6.2 Practice Communication (Stage 2 - Duration approximately 1 year)	18
7 STUDY PROCEDURES/EVALUATIONS	19
7.1 Develop query	19
7.2 Run queries on archived or copied practice database (to be completed by respective EDR vendors).....	19
7.3 Data Cleaning & Preparation.....	20
7.4 Extension of OHD Ontology to incorporate data elements relevant to PCRs, RCTs and tooth loss.....	20
7.5 Extract and transform data of interest into OHD-structured data, and use this representation to generate data sets for statistical analysis.....	20
8 ASSESSMENT OF SAFETY	22
8.1 Specification of Safety Parameters	22

8.1.1	Unanticipated Problems	22
8.2	Reporting Procedures	22
9	STUDY OVERSIGHT	24
10	STATISTICAL CONSIDERATIONS	25
10.1	Study Hypotheses	25
10.2	Sample Size Considerations	25
10.3	Analysis Plan.....	26
11	QUALITY CONTROL AND QUALITY ASSURANCE.....	28
12	ETHICS/PROTECTION OF HUMAN SUBJECTS	29
12.1	Ethical Standard.....	29
12.2	Institutional Review Board.....	29
12.3	Informed Consent Process.....	29
12.4	Exclusion of Women, Minorities, and Children (Special Populations).....	30
12.5	Participant Confidentiality.....	30
13	DATA HANDLING AND RECORD KEEPING.....	32
13.1	Data Management Responsibilities.....	32
13.2	Data Capture Methods	32
13.3	Types of Data.....	32
13.4	Schedule and Content of Reports	33
13.5	Study Records Retention	33
13.6	Protocol Deviations	33
14	PUBLICATION/DATA SHARING POLICY	35
15	LITERATURE REFERENCES	36
	APPENDICES.....	39
	APPENDIX A: SCHEDULE OF EVENTS	40
	APPENDIX B: DRAFT LIST OF STUDY VARIABLES.....	41
	APPENDIX C: SAMPLE CLASSES IN THE HEALTH AND DISEASE ONTOLOGY	44
	APPENDIX D: EXAMPLE OF THE ORAL HEALTH AND DISEASE ONTOLOGY REPRESENTING CLINICIANS, PATIENTS, MEDICAL HISTORY, TEETH AND THEIR ANATOMY, PROCEDURES.....	47

LIST OF ABBREVIATIONS

ADA	American Dental Association
CC	Coordinating Center
CFR	Code of Federal Regulations
DMP	Data Management Plan
EDR	Electronic Dental Records
FFR	Federal Financial Report
GCP	Good Clinical Practice
GPI	Grant Principal Investigator
HIPAA	Health Insurance Portability and Accountability Act
ICF	Informed Consent Form
IRB	Institutional Review Board
IU	Indiana University
N	Number (typically refers to participants)
National Dental PBRN	National Dental Practice-Based Research Network
NIDCR	National Institute of Dental and Craniofacial Research, NIH, DHHS
NIH	National Institutes of Health
OCTOM	Office of Clinical Trials Operations and Management, NIDCR, NIH
OHD	Oral Health and Disease
PCR	Posterior Composite Restoration
PHI	Protected Health Information
QA	Quality Assurance
QC	Quality Control
RAS	Regional Administrative Sites
RC	Regional Coordinator
RDC	Regenstrief Data Core
RCT	Root Canal Treatment
RI	Regenstrief Institute
SPI	Study Principal Investigator
SQL	Standard Query Language
US	United States

PROTOCOL SUMMARY

- Title:** Leveraging electronic dental record (EDR) data for clinical research: Proof-of-concept study.
- Précis:** This study investigates the feasibility of electronic clinical data reuse for research purposes.
It is a proof-of-concept retrospective cohort study on the utility of electronic dental records data to assess outcomes of posterior composite restorations (PCR) and tooth loss following root canal treatment (RCT) (regardless of whether the tooth received dental restorations following RCT or not). An important second goal of this study is to determine the extent to which electronic dental records data may support obtaining reliable study outcomes for National Dental Practice-Based Research Network (PBRN) studies.
- Objectives:** The primary objective of this study is to explore the extent to which National Dental PBRN EDR data can be used to assess: a) PCR longevity and b) tooth loss following RCT, with relevant data in the National Dental PBRN practitioners' EDR systems.
- The secondary objectives of this study are to:
1. Extend the Oral Health and Disease ontology (OHD) to integrate heterogeneous data sources from different systems to perform data analysis;
 2. Explore the extent to which practice/practitioner-, patient-, tooth-, and treatment-related factors are available with relevant data in National Dental PBRN practitioners' EDR systems;
 3. Assuming there will be adequate data, explore the effect of practice/practitioner-, patient-, tooth-, and treatment-related factors on PCR longevity and tooth loss following RCT.
- Population:** The study participants will be approximately 100 US dental practitioners from the National Dental PBRN who have used EDR, Dentrix or EagleSoft for at least 5 years, to maintain electronic clinical information of at least existing conditions and treatment performed.

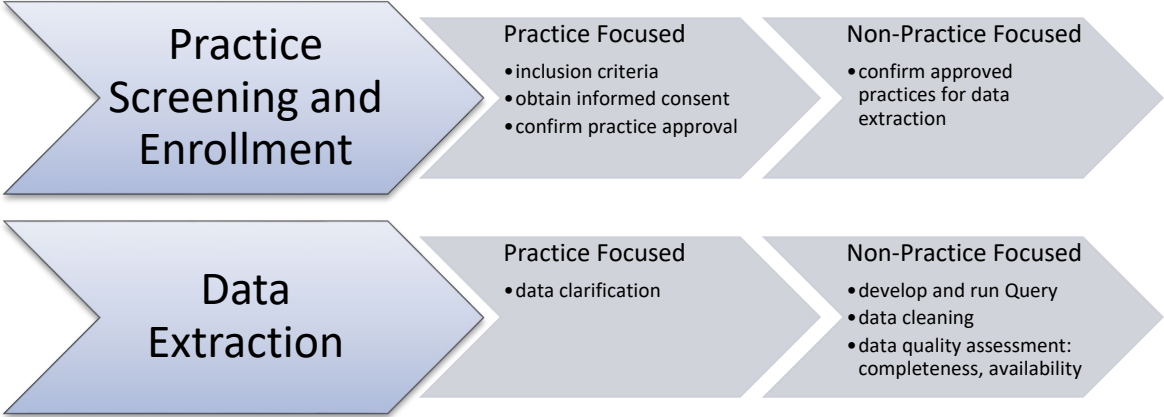
Number of Sites: Approximately 100 US dental practitioners from the National Dental PBRN.

Study Duration: Approximately 3 years

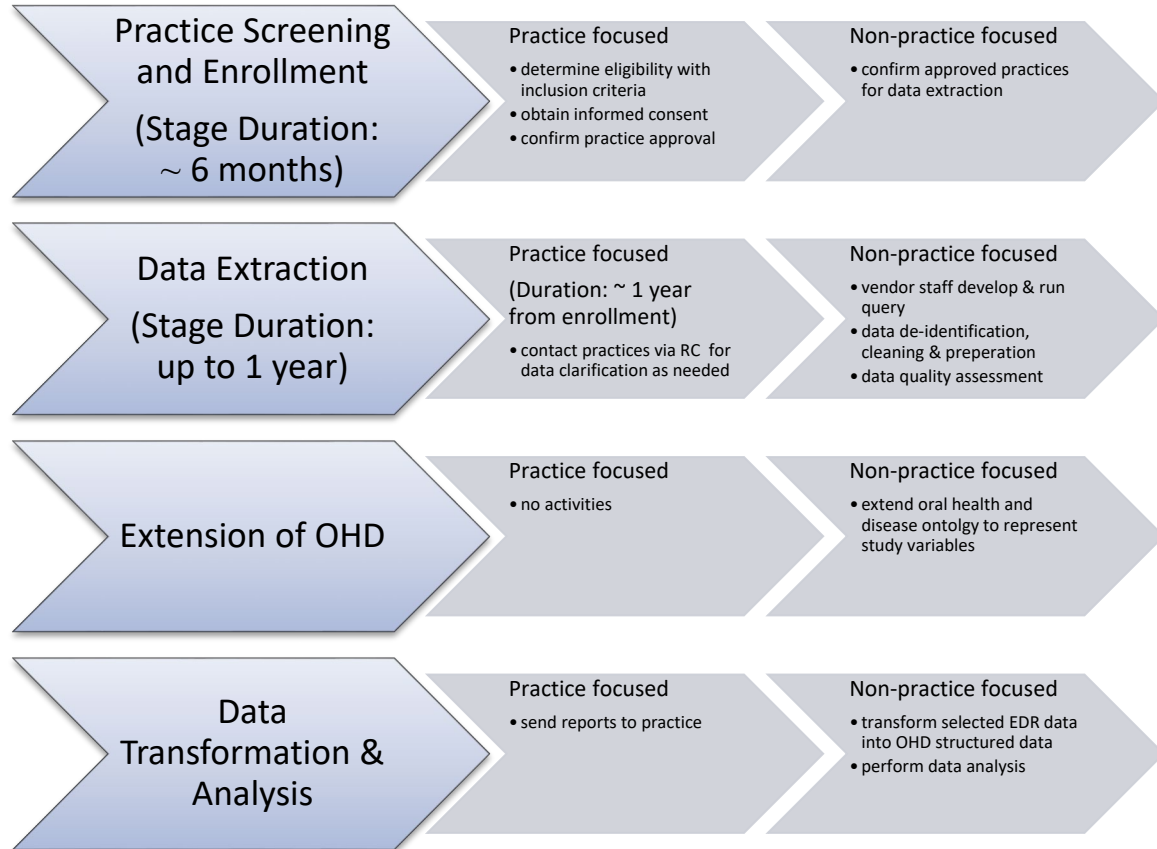
Subject Participation Duration: Approximately 1 year

Estimated Time to Complete Enrollment: Approximately 6 months

Schematic of Study Design: Ramp-up Phase (Duration ~ 3 months)



Schematic of Study Design: Full Study



1 KEY ROLES AND CONTACT INFORMATION

Grant Principal Investigator:

Gregg H. Gilbert, DDS, MBA
Professor and Chair
University of Alabama at Birmingham
1720 Second Ave. South
School of Dentistry, SDB 109
Birmingham, AL 35294-0007
Phone: 205-975-8886
Fax: 205-975-0603
Email: ghg@uab.edu

Study Principal Investigator:

Thankam Paul Thyvalikakath, DMD, MDS, PhD
Associate Professor & Director of Dental Informatics Core
Research Scientist, Regenstrief Institute, Inc.
Department of Restorative Dentistry, Indiana University School of Dentistry, IUPUI
1121 W Michigan Street, Room 314,
Indianapolis, IN 46202
Phone: 317-274-5460
Email: tpt@iu.edu

Co-Investigators:

Titus Schleyer, DMD, PhD
Clem McDonald Professor of Biomedical Informatics
Director, Center for Biomedical Informatics
Regenstrief Institute, Inc.,
410 West 10th Street, Suite 2000,
Indianapolis, IN 46202-3012
Phone: 317-274-9204
Email: schleyer@regenstrief.org

William Duncan, PhD
Assistant Research Scientist
Regenstrief Institute, Inc.,
410 West 10th Street, Suite 2000,
Indianapolis, IN 46202-3012
Phone: 716-603-4769
Email: wdduncan@gmail.com

NIDCR Program Official:

Dena Fischer, DDS, MSD, MS
NIH/NIDCR/DER
6701 Democracy Boulevard, MSC 4878
Bethesda, MD 20892-4878
Phone: 301-594-4876

Email: dena.fischer@nih.gov

**Coordinating
Center:**

Regenstrief/Indiana Univeristy
410 West 10th Street, Suite 2000
Indianapolis, IN 46202
Thankam Paul Thyvalikakath, DMD, MDS, PhD
Phone: 317-274-5460
Email: tpt@iu.edu

Institutions:

Western Region (region #1)

Administratively based at the Kaiser Permanente Center for
Health Research, Portland Oregon
Lisa Waiwaiole, Regional Coordinator
Kaiser Permanente Center for Health Research
3800 N. Interstate Ave.
Portland, OR 97227-1110
Office: (503) 335-2454
Fax: (503) 335-6311
Email: Lisa.Ann.Waiwaiole@kpchr.org

Midwest Region (region #2)

Administratively based at the HealthPartners Institute for
Education and Research in Minneapolis, MN
Emily Durand, Regional Coordinator
HealthPartners Institute for Education and Research
8170 33rd Avenue South
MS: 21111R
Minneapolis, MN 55445
Office: (952) 967-7404
Fax: (952) 967-5022
Email: Emily.C.Durand@HealthPartners.com

Southwest Region (region #3)

Administratively based at the University of Texas Health
Science Center at San Antonio in San Antonio, TX
Stephanie C. Reyes, Regional Coordinator
7703 Floyd Curl Drive, MC 7894
San Antonio, TX 78229
Office: (210) 562-5654
Fax: (210) 562-4136
Email: reyess@uthscsa.edu

South Central Region (region #4)

Administratively based at the University of Alabama at
Birmingham in Birmingham, AL
Andrea Mathews, Program Manager
Department of Clinical and Community Sciences
School of Dentistry, SDB 114
1720 2nd Avenue South
Birmingham, AL 35294-0007
Office: (205) 934-2578
Fax: (205) 996-2172
Email: ahmathews@uab.edu

South Atlantic Region (region #5)

Administratively based at the University of Florida in
Gainesville, FL
Deborah McEdward, Regional Coordinator
University of Florida
P.O. Box 100415
Gainesville, FL 32610
Office: (352) 273-5848
Fax: (352) 273-7970
Email: dmcedward@dental.ufl.edu

Northeast Region (region #6)

Administratively based at the University of Rochester in
Rochester, NY
Pat Ragusa, Regional Coordinator
Eastman Institute for Oral Health
625 Elmwood Avenue, Box 683
Rochester, NY 14620
Phone: (585) 275-5780
Fax: (585) 273-1237
Email: Pat_Ragusa@URMC.Rochester.edu

**Other Key
Personnel:**

Dr. Tu Wanzhu - Biostatistician
Ms. Tracy Shea – Lead Regional Coordinator

2 INTRODUCTION: BACKGROUND INFORMATION AND SCIENTIFIC RATIONALE

2.1 Background Information

This study is designed to investigate the feasibility of electronic clinical data reuse for research purposes. The protocol describes a proof-of-concept, retrospective cohort study on the outcomes of posterior composite restorations (PCR) and tooth loss outcomes of root canal treatments (RCT) (regardless of whether the tooth received dental restorations following RCT or not). In the following sections, we discuss the value of using electronic patient records data for research and quality improvement purposes, and the rationale for choosing outcomes of posterior composite restorations and root canal treatments as topics for our proof-of-concept study.

To date, the predominant use of electronic data for dental outcomes or health services research has been analyzing insurance claims (Bailit 1996). The advantages of this approach include the high standardization of insurance data, their general availability for large populations, and well-established methodologies for aggregation and analysis. However, their drawbacks include the dearth of socioeconomic variables in claims, the absence of tooth condition/diagnostic codes, their focus on insured, not patient, populations, and lower validity compared to patient records (Bailit 2003, Hayden 1997, Demko et al. 2008). Other weaknesses are that claims data are primarily obtained for reimbursement, not research, purposes; there is an absence of information about items that are not billable or “bundled”; and there is potential skewing of information to maximize reimbursement.

Traditional clinical research, on the other hand, may fall short of the needs of clinicians, patients and funding agencies for many reasons (Kahn et al. 2007). They include the high cost of clinical trials, the time required to obtain results, difficult enrollment, often poor generalizability, and challenges in discovering clinical outcomes and side effects in a nuanced and clinically meaningful manner (Sung et al. 2003, Kahn et al. 2007, Olsen et al. 2007). Consequently, the use of electronic data collected during clinical care has received increased attention as a method for increasing our evidence base (Sung et al. 2003, Olsen et al. 2007, Prokosch and Ganslandt 2009, Terry et al. 2010, Pearson et al. 2011). Using EHR data for research can help increase efficiency (Sung et al. 2003, Kahn et al. 2007), lower costs (Kohane 2011), allow the study of patient rather than research participant populations, avoid certain selection biases, facilitate longitudinal studies, detect rare events, and discover drug side effects earlier than possible with traditional methods (Brownstein et al. 2007, Wilke et al. 2007, Seyfried et al. 2009).

EDR use is growing at a rapid pace, facilitating a growing corpus of EDR data available for research and quality assurance (QA). Beyond the limitations of insurance claim data, additional factors make the use of EDR data increasingly attractive in the context of research in the National Dental Practice-Based Research Network (PBRN). The experience during the PBRN’s initial funding period has shown that there are more relevant research questions than can be practically addressed using the typical PBRN study approach; that some important long-term research questions do not fit very well

into the timeframe of PBRN studies; and that costs and practice workflow issues limit the "throughput" of the PBRN system. As our recent survey of Dental Practice-Based Research Network (DPBRN) practitioners (Schleyer et al. 2013) has shown, approximately 75% use a computer to manage clinical information and 15% are paperless, making secondary analysis of EDR data increasingly feasible and compelling. These findings are evidence for a national trend toward complete digital record-keeping in clinical practice in the foreseeable future. Office computing, primarily focused on billing and insurance processing, has essentially saturated dental practice already. Chairside computing, in which practitioners use a computer in the operator, is undergoing a phase of rapid growth. Importantly, paperless operation, in which all dental records are computerized, is beginning to move beyond early adoption.

2.2 Rationale

Turning to our research questions, the literature already provides some information on the longevity and outcomes of PCRs. Major findings include: 1) among high caries risk patients, PCRs often fail earlier than amalgam restorations (Van Nieuwenhuysen et al. 2003, Burke and Lucarotti 2009, Sunnegardh-Gronberg et al. 2009, Opdam et al. 2010); 2) PCRs appear to be associated with fewer tooth fractures than, for instance, amalgam restorations (Forss and Widstrom 2004, Tyas 2005); and 3) a variety of factors, including restoration size, root surface involvement, and patient age, appear to be associated with restoration longevity (Levin et al. 2007, Simecek et al. 2009).

Key findings from the above studies illustrate the limitations to current insights on PCR longevity and outcomes. First, most studies, including those referenced above, have been conducted outside the US with geographically limited, often specialized populations and small sample sizes (on average, up to 2,000 restorations). Second, retrospective analyses based on insurance claims are subject to inherent biases due to patients' changes in insurance status, and the absence of diagnostic and other clinical information. Few studies have combined patient records review with clinical patient examinations to evaluate the status of the restorations (Kim et al. 2013, McCracken et al. 2013, Rho et al. 2013). Study results remain inconclusive due to the limited follow-up time, limited number of restoration failures and presence of confounding variables such as socio-economic factors, oral hygiene status and systemic diseases. Our proposed study will improve this understanding through a US-based, large-scale, retrospective analysis of EDR data with a longer follow-up time and multiple practices/practitioners.

An important goal of this study is to determine the extent to which EDR data may support obtaining reliable study outcomes for National Dental PBRN studies. To explore this goal, we will assess the outcome of tooth loss in EDR data since tooth loss is the primary outcome for a National Dental PBRN longitudinal study of outcomes following endodontic treatment. The proposed study will improve our understanding of the utility of EDR data for obtaining National Dental PBRN longitudinal study outcomes in a cost- and time-efficient manner.

In summary, there are five compelling reasons to develop our capabilities for using EDR data for clinical research.

1. Electronic dental records are valuable for research and their analysis can complement existing research methodologies (Song et al. 2013).
2. Electronic dental record use is growing, making a rapidly growing corpus of EDR data available for research and quality assurance (Schleyer et al. 2013a).
3. Use of electronic patient data for dental clinical research is in its infancy but highly promising (Song et al. 2013).
4. Numerous outcomes studied in dental PBRN studies are captured in EDRs (Liu et al. 2013).
5. Our current work developing the Oral Health and Disease (OHD) Ontology and structuring data from EDRs suggests it is a practical methodology for converting PBRN EDR data into data usable for research studies (Duncan et al. 2013, Schleyer et al. 2013b; see Appendix D).

This study lays the groundwork for making quality improvement a part of dental practice, as suggested by Olsen et al. (Olsen et al., 2007). This means that clinicians need tools that enable them to perform analyses of the data in their EDR for selected quality measures, implement appropriate interventions, if necessary, and repeat the analyses at a later date to determine the outcomes of the intervention. If a tool based on our findings is developed for PCRs and tooth loss due to RCT, extension to other types of analyses based on available practice data is very feasible. The National Dental PBRN is eminently suited to develop, implement and evaluate this approach.

We therefore, propose a proof-of-concept study to investigate the feasibility of using OHD-structured EDR data for research within the National Dental PBRN. We have chosen to study outcomes of posterior composite restorations and tooth loss outcomes following RCT regardless of whether the tooth received dental restorations following RCT.

2.3 Potential Risks and Benefits

2.3.1 Potential Risks

As with any study, there is the possibility of breach of confidentiality for participating practices, practitioners and patient data. We address this potential risk by removing identifiers for practices and practitioners and creating a data set from existing patient data in the EDR. The data gathered will be in compliance with the Health Insurance Portability and Accountability Act Privacy and Security rules <http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html>. Appropriate precautions will be taken and procedures will be followed to maintain confidentiality. These include use of unique study codes for practitioners, de-identification or obfuscating of patient protected health information (PHI), encryption of electronic data for transmission to Indiana University (IU)/Regenstrief Institute (RI) and the coordinating center (CC), and password-protected computers for data storage. Compliance with all IRB regulations concerning data collection, data analysis, data storage, and data destruction will be strictly observed.

2.3.2 Potential Benefits

There are only limited immediate benefits for the practitioners who participate in the study. This retrospective data analysis will not benefit the individual patients whose data is included in the study. However, participating practitioners will receive individual reports on their EDR documentation, such as completeness, availability, and currency of relevant data on completed medical history, progress notes, allergies, findings, conditions and completed dental procedures. If adequate data is available, they will also receive reports on the longevity of PCR and tooth loss following RCTs performed in their practices. They will also receive a report showing their practice data compared to other practices data in aggregate without any identifiers. A potential benefit of this study for the oral health research community is to determine the extent to which the EDR data may support obtaining reliable study outcomes.

3 OBJECTIVES

3.1 Study Objectives

The primary objective of this study is to explore the extent to which National Dental PBRN EDR data can be used to assess a treatment outcome. As a proof of concept, the study will assess outcomes of two treatments: a) PCR longevity and b) tooth loss following RCT, with relevant data in the National Dental PBRN practitioners' EDR systems.

The secondary objectives of this study are to:

1. Extend the OHD ontology to integrate heterogeneous data sources from different systems to perform data analysis;
2. Explore the extent to which practice/practitioner-, patient-, tooth-, and treatment-related factors are available with relevant data in National Dental PBRN practitioners' EDR systems;
3. Assuming there will be adequate data, explore the effect of practice/practitioner-, patient-, tooth-, and treatment-related factors on PCR longevity and tooth loss following RCT.

3.2 Study Outcome Measures

We will first assess the extent to which the extracted EDR data can be used to assess PCR and RCT outcomes. This assessment will be based on the availability, completeness, concordance, and currency of the variables documented in the EDR. We will identify the variables collected in previous study protocols that evaluated PCR and RCT outcomes and evaluate their presence or absence in the EDR. We will also determine the range of variables available in the EDR, its completeness and concordance. While we expect certain variables to be present explicitly, certain variables may be present only implicitly. For instance, a missing tooth would be documented explicitly through CDT codes or findings. A deep carious lesion may not be explicitly documented but could be inferred implicitly by looking for documentation on whether Dycal or vitrebond was placed before placing composite restoration.

Completeness will be measured as the proportion of dental records from participating practices that contain each study variable. Concordance will be measured as the proportion of available study variables identically recorded in free text (clinical findings) and coded (administrative) data in the EDR. Currency of the data will be measured to assess whether the study variables are relevant expressions at a given point in time. Based on the extent to which relevant study variables are available, we will select a subset of records with at least 60% complete and current data to assess posterior composite restoration longevity, time to tooth loss on which RCT was performed and potential factors influencing these two outcomes.

Once we have generated final data sets, we will again assess for completeness, availability and currency. We will achieve this through descriptive statistical analysis, such as the frequency of the data appearing in the EDR, number of providers by

amount of missing data, etc. During our data extraction and transformation processes, we will formally assess the validity of data to the degree possible. For instance, we will sample a number of tooth extraction events and attempt to validate whether a tooth extraction actually took place by comparison with progress notes. At this point, we will make a final assessment of the extent of missing data, and the feasibility of using the data to assess restoration longevity and tooth loss (as described below). Additionally, we will assess which records have sufficient data to address the secondary study objectives, described below. A final analysis plan will be developed based upon factors for which a defined level of completeness is available.

Restoration longevity

The primary outcome of PCR is “restoration longevity” or “restoration survival”, defined as the time from the date of restoration placement until the date of an “event”. An “event” is defined as the first time either: 1) a new fixed restoration (filling or crown) is placed on that surface; 2) caries/secondary caries is charted on that surface; 3) RCT is initiated on that tooth; or 4) extraction is performed on that tooth. Note: To the degree possible, we will attempt to verify that the event is related to the original restoration. For instance, an occlusal restoration separate from an existing occlusal restoration is not considered a related event. This scenario could, for instance, be detected by reviewing the corresponding progress note. Data will be censored if no event occurred by the end of the available data collection.

Tooth loss

The next primary outcome variable is the time to tooth loss following RCT (performed on the same tooth); tooth loss is indicated by tooth extraction in the dental record.

Factors associated with primary outcomes

For the secondary objectives, potential factors influencing the outcomes of PCR and tooth loss following RCT will be drawn from clinical and demographic factors (independent variables) available from EDR. Tables 1 and 2 show a sample list.

Table 1: Sample list of clinical factors that will be compiled from the National Dental PBRN practitioners' EDR systems (as available).

Information category	Sample factors
Demographics	age, sex, race, ethnicity, socioeconomic variables (e.g. type of insurance)
Reason for visit	chief complaint, history of current problem, replacement of old restorations, loss of retention, restoration fracture, overhang, pulpal problems including pulpal exposure, hypersensitivity and pulpal pain, discoloration of restorations
Medical and dental history	presence or absence of medical conditions (such as diabetes, stroke etc.), medication history, last dental visit, types of dental treatment received
Social history	smoking or alcohol habits, oral care habits
Tooth characteristics	tooth type, cavity classification, caries, tooth fracture, crack, attrition, abrasion, erosion, presence or absence of root surface
Restorative treatment	procedure, restorative material, surfaces restored, restoration class (I, II, V), subsequent treatment including extraction, endodontic treatment, prosthodontics treatment
Endodontic treatment	number of visits required to complete RCT, antibiotics prescribed, restoration type, post replaced, periodontal involvement, presence or absence of pain/soft tissue tenderness, presence or absence of periapical radiolucency, number of canals

Table 2: Practice and dentist characteristics that will be compiled from EDR systems (as available)

Practice/practitioner characteristics	Sample factors
Practice	Number of dentists and hygienists, type and quantity of procedures each day, number of patients per day
Dentist	Cumulative count of procedures of different types, workload, years in the practice, number of encounters with patient

The clinical factors discussed in tables 1 and 2 will be obtained from procedure codes, chart notes, odontograms, etc. It is anticipated that some data will be easier to obtain (e.g. procedures, which are structured data) than others which may be available in free text progress notes (social history, medication history). Natural language processing and text mining methods will be utilized to identify the presence of relevant clinical factors in progress notes and other free text notes in the EDR. The practitioner level data will be obtained from the National Dental PBRN Enrollment Questionnaire that all Network practitioner members are required to complete.

4 STUDY DESIGN

- This is a proof-of-concept retrospective, cohort study to investigate the:
 - feasibility of electronic data reuse for research purposes and
 - utility of EDR data to assess outcomes of PCR and tooth loss outcomes of RCT (regardless of whether the tooth received dental restorations following RCT or not).
- The practitioner study participants will be dentists from the National Dental PBRN regions who have used EDR software for at least 5 years (Dentrix or EagleSoft) to maintain electronic clinical information of at least existing conditions and treatment performed.
- The study will include approximately 100 US dental practitioners who placed at least one PCR per patient on at least 100 patients or performed at least one RCT on permanent tooth per patient on at least 50 patients and:
 - have follow-up electronic data available for at least 2 years
 - have performed these procedures between January 1, 2000 and October 31, 2015.
- RC and RAS will recruit, enroll and educate practitioners on study details.
- RCs will inform participating practices about their minimal involvement with mining data from their EDR and that the assigned vendor (EagleSoft or Dentrix) personnel will perform data mining remotely on historical or archived data made available to them.
- Upon practitioners agreeing and signing informed consent on behalf of the practices, RCs will inform the vendor personnel and IU/RI study team.
- Vendor personnel will contact practices to confirm their study participation. Next, they will run data extraction programs on the practice's archived data base.
- Vendor personnel will perform data extraction based on the study procedures described in section 7. They will assign unique study codes for practices and practitioners, de-identify or obfuscate patient PHI, and encrypt electronic data for transmission to Indiana University (IU)/Regenstrief Institute (RI). This process will have three benefits: (1) reduce technical overhead for the participating practice; (2) reduce the probability of security and/or confidentiality breaches; and (3) increase practice acceptance by reusing a process that is already used during regular customer support by the vendor.
- Upon receiving the encrypted data, the IU/RI personnel will run a scrubber program to remove any potential patient and practitioner identifiers present in free text notes and confirm absence of any patient or practitioner specific identifiers in the entire data.
- After data is received by IU/RI, study team personnel may contact participating practices via RC if any data clarification is needed. This may occur during the one-year period following the enrollment of practice in the study.

4.1 Overview of data assessment, cleaning and transformation

- First, based on existing literature, the Study Principal Investigator (SPI) or

designee will compile a list of EDR data elements representing information about PCRs, RCTs, tooth loss, and clinical practice factors influencing these clinical elements. In addition to clinical variables, we will also construct a list of relevant practice and practitioner characteristics that can be determined from the EDR.

- Second, we will analyze the degree to which these data elements of interest are stored in practitioners' EDR. As mentioned previously, we will employ text mining and natural language processing methods to identify and analyze relevant data from free text notes that cannot be retrieved using standard data query methods. We will measure the availability, completeness, concordance and currency of the variables documented in the EDR. We will identify the variables collected in previous studies that evaluated PCR and RCT outcomes and evaluate their presence or absence in EDR data. We will also determine the range of variables available in the EDR, its completeness and concordance.
- Third, we will extend the OHD ontology to incorporate data elements relevant to PCRs, RCTs, and tooth loss.
- Fourth, we will extract and transform data of interest into OHD-structured data, and use this representation to generate data sets for statistical analysis.
- As the final step, data analysis will be performed and results will be finalized.
- RI/IU personnel will also prepare reports for practitioners.

4.2 Ramp-up Phase

The Study Team will pilot the study with approximately 4 practices. The purposes of the ramp-up phase of the study are to evaluate the ability to extract longitudinal data from Dentrax and EagleSoft systems, understand the extent/breadth of data available, understand the completeness/depth of the data and streamline the data mining and transformation processes. During this phase, we will also create an iterative process to determine data quality and completeness and also refine the definition of data quality and completeness.

The duration of the ramp-up phase will be approximately three months to allow for extraction of data and initial analyses to refine the extent/breadth and completeness/depth measures.

5 STUDY ENROLLMENT AND WITHDRAWAL

5.1 Practitioner Inclusion Criteria

In order to be eligible to participate in this study, a practitioner must meet all of the following criteria:

- Enrolled in the National Dental PBRN as a practitioner who is licensed in the United States to treat patients and treats patients in the United States on a recurring basis.
- Is a “full” participation level member, or willing to change from “information only” or “limited” to “full”.
- Complete study-specific training, this will be based on local IRB requirements.
- Attended or viewed a National Dental PBRN orientation session or has attended at least one annual regional meeting of practitioners.
- Has used the Dentrax or EagleSoft EDR system in the practice for at least 5 years.
- Maintain electronic clinical information of at least existing conditions and treatment performed.
- Have placed at least one PCR on permanent teeth per patient in at least 100 patients or performed at least one RCT on permanent tooth per patient in at least 50 patients (based on practitioner self-report) and:
 - Have electronic follow-up data available for at least 2 years
 - Have performed these procedures between January 1, 2000 and October 31, 2015.
- Willing to share practice level data with the study team. The vendor technical staff will retrieve data from the practice EDR system by following the study procedure described in Section 7.2.
- Practitioner has the authority to share practice-level data.

5.2 Strategies for Recruitment and Retention

5.2.1 Practitioner Recruitment and Retention

All National Dental PBRN dentist practitioners at full participation level, or willing to change to full participation level, and who use EDR software (Dentrix or EagleSoft) will be approached for participation in the study. The study team will rely on the RCs from the Regional Administrative Sites (RAS) to help recruit current practitioner participants into the study.

The study participants will be approximately 100 US dental practitioners from the National Dental PBRN who use EDR software, Dentrix or EagleSoft, to maintain electronic clinical information of at least existing conditions and treatment performed.

Actual enrollment will be based on interest and enrollment into the Network, and likely will vary per region.

Practitioners will be offered an honorarium \$500.00 payment for participating and sharing their data. We will also share the outcomes report on PCRs and RCTs from their patient data.

During data cleaning and processing, practices may be contacted for data clarification. The study team will rely on the RC from the RAS to contact the practitioner and either query the practitioner on behalf of the study team or facilitate communication with the study team.

5.3 Withdrawal

5.3.1 Reasons for Withdrawal

Practitioners are free to withdraw from participation in the study at any time upon request. Patient data from these practitioners will not be used for study analysis from the time they withdraw from the study.

5.3.2 Handling of Withdrawals

If a practitioner withdraws from the study during the study enrollment period, he/she will be replaced by another practitioner. If the practitioner withdraws from the study after the enrollment period, he/she will not be replaced.

5.4 Premature Termination or Suspension of Study

This study may be suspended or prematurely terminated if there is sufficient reasonable cause. Written notification, documenting the reason for study suspension or termination, will be provided by the suspending or terminating party. If the study is prematurely terminated or suspended, the SPI will promptly inform the IRB and will provide the reason(s) for suspension or termination.

Circumstances that may warrant termination include, but are not limited to:

- Determination of unexpected, significant, or unacceptable risk to patients.
- Insufficient adherence to protocol requirements.
- Data that are not sufficiently complete and/or evaluable.
- Determination of futility.

6 STUDY SCHEDULE

Those National Dental PBRN practitioners who opt to participate will be sent information and instructions pertaining to the study. These instructions will provide information for the practitioner(s) and staff who will help to execute the study. In addition, RCs will conduct in-person or remote meetings with office staff to provide further information about the study and to make sure the practitioner and their office staff understand the study procedure(s).

The study will proceed in stages:

- 1) We plan to recruit practitioners based on interest and enrollment into the Network, and inclusion criteria and therefore, may vary per region;
- 2) Practitioners will undergo informed consent procedures as required by the IRB. Individual Investigator Agreements, the University of Alabama at Birmingham (UAB) Master Service Agreement and similar documents will also be obtained prior to the start of the study as required; and

The study team, along with the RAS and RCs, will coordinate the launch of the study. Once the RC has performed informed consent procedures, the study team will coordinate with the assigned EagleSoft and Dentrix technical personnel to confirm the practice is approved for data extraction. Due to concerns regarding system impacts and confidentiality, Dentrix and EagleSoft technical personnel will run the queries on archived databases and perform pseudonymization on necessary data elements. Both Dentrix and EagleSoft will run queries on non-production data bases, which will eliminate impact on real-time production systems for providers. This process will have three benefits: (1) reduce technical overhead for the practice in participating; (2) reduce the probability of security and/or confidentiality breaches; and (3) increase practice acceptance minimizing impact on their live production systems. Dentrix and EagleSoft personnel who are part of the study team will extract data from Dentrix and EagleSoft practices respectively. If necessary IU/RI personnel will run query scripts developed by Dentrix and EagleSoft to extract data from participating practices archived databases.

An overview of study procedures to be completed at each data extraction instance can be found in Appendix A.

6.1 Practice Screening/Enrollment (Stage 1)

- Verify practitioner inclusion criteria.
- RCs obtain and document informed consent from practitioner participant according to regional IRB requirements.
- RCs will contact study team, who will coordinate with EDR technical personnel to confirm practitioner approval for data extraction (for details regarding data extraction process refer to section 7.2).
- RC will provide confirmation of upcoming data extraction to the participating practitioners.

6.2 Practice Communication (Stage 2 - Duration approximately 1 year)

During data cleaning and transformation following data extraction, practices may be contacted for clarification of data as needed. The study team will rely on the RC from the RAS to contact the practitioner and either query the practitioner on behalf of the study team or facilitate communication with the study team.

7 STUDY PROCEDURES/EVALUATIONS

To complete the objectives of this research, we will perform the following steps:

7.1 Develop query and assess data completeness

- Compile a list of EDR data elements representing information about PCRs, RCTs, tooth loss, and clinical and practice factors influencing these clinical entities. Based on existing literature, we will compile a list of clinical factors that influence PCRs (Kim et al. 2013, McCracken et al. 2013), RCTs (Ng et al. 2011, Ng et al. 2011, Bernstein et al. 2012) and tooth loss subsequent to RCT. Appendix B provides a preliminary list based on several studies. The list will then be cross-checked with Dentrix's data dictionary and the data list generated from EagleSoft (Duncan et al. 2013) to confirm the National Dental PBRN practitioners' ability to document the data elements in Dentrix and EagleSoft. In addition to clinical variables, we will also construct a list of relevant practice and practitioner characteristics that can be determined from the EDR (see Table 2; Section 3.2).

7.2 Run queries on archived or copied practice database (to be completed by respective EDR vendors)

- EagleSoft and Dentrix will have data agreements in place between participating practices and the vendor prior to data extraction.
- EagleSoft and Dentrix personnel will coordinate with participating practices for data extraction from each practice.
- Identify patient records with at least one PCR on permanent teeth placed between January 1, 2000 through October 31, 2015 and have follow-up electronic data available for at least 2 -years.
- Identify patient records with at least one RCT performed on permanent teeth between January 1, 2000 through October 31, 2015 and have follow-up electronic data available for at least 2 -years.
- Extract data from practice database and generate individual data sets, one data set per practice of records meeting the enrollment criteria.
- Create mapping file linking the real patient identifier to a unique pseudonymized patient identifier.
- Pseudonymize appropriate data elements and prepare data set for secure transfer.
- Securely transmit study data from EDR vendor to /Regenstrief Institute.

7.3 Data Cleaning & Preparation

- IU/Regenstrief Institute staff will receive individual data sets from the EDR vendor.
- Run scrubbing process on free text components of the data set, this will remove any potentially identifiable data that may be present.
- Assess completeness, availability and currency of relevant data.
- Mine data relevant to clinical factors from patient records

Once we have generated final data sets, we will assess for completeness, availability and currency. We will achieve this through descriptive statistical analysis, such as the frequency of the data appearing in the EDR, number of providers by amount of missing data, etc. During our data extraction and transformation processes, we will formally assess the validity of data to the degree possible. For instance, we will sample a number of tooth extraction events and attempt to validate whether a tooth extraction actually took place by comparison with progress notes. At this point, we will make a final assessment of the extent of missing data, and the feasibility of using the data to assess restoration longevity and tooth loss (as described in the outcome measures section above). Additionally, we will assess which records have sufficient data to address the secondary study objectives, described below. A final analysis plan will be developed based upon factors for which a defined level of completeness is available.

7.4 Extension of OHD Ontology to incorporate data elements relevant to PCRs, RCTs and tooth loss

In this step, we will extend the OHD using the clinical and practice/practitioner data compiled in section 7.2. Appendix D shows an example of the OHD, which models patients, medical history, teeth and their anatomy, findings, and procedures and the roles of their participants. Extensions of the ontology would involve reviewing our findings from section 7.2, adding classes that represent any entities necessary to answer queries but do not already exist in OHD or related ontologies, and documenting these terms in a manner that is easy to understand. Where necessary, we will consult with the practitioner for additional clarity about a term. As part of this process, we will document the number of patients with sufficient data to complete the ontology, again using descriptive frequencies.

7.5 Extract and transform data of interest into OHD-structured data, and use this representation to generate data sets for statistical analysis

In the fourth step, we will transform selected EDR data into OHD-structured data, and query the OHD-structured data to generate data sets for statistical analysis. Transforming EDR data involves first understanding the schemas of the EDR systems. Although the schemas offer us insights into what data to expect, we expect to see variations in the actual data recorded across dental practices. For instance, while one system may record a patient's medical history as 'existing medical conditions', another practice may record it

as 'medical alerts'. These variations will be represented as a single concept, medical history in the OHD so that it supports different data formats represented in the EDR systems.

8 ASSESSMENT OF SAFETY

8.1 Specification of Safety Parameters

Safety monitoring for this study will focus on unanticipated problems involving risks to participants.

8.1.1 Unanticipated Problems

The Office for Human Research Protections (OHRP) considers unanticipated problems involving risks to patients or others to include, in general, any incident, experience, or outcome that meets **all** of the following criteria:

- Unexpected in terms of nature, severity, or frequency given (a) the research procedures that are described in the protocol-related documents, such as the IRB-approved research protocol and informed consent document; and (b) the characteristics of the patient population being studied;
- Related or possibly related to participation in the research (possibly related means there is a reasonable possibility that the incident, experience, or outcome may have been caused by the procedures involved in the research); and
- Suggests that the research places patients or others at a greater risk of harm (including physical, psychological, economic, or social harm) than was previously known or recognized.

8.2 Reporting Procedures

Incidents or events that meet the OHRP criteria for unanticipated problems require the creation and completion of an unanticipated problem report form. OHRP recommends that investigators include the following information when reporting an adverse event, or any other incident, experience, or outcome as an unanticipated problem to the IRB:

- Appropriate identifying information for the research protocol, such as the title, investigator's name, and the IRB project number;
- A detailed description of the adverse event, incident, experience, or outcome;
- An explanation of the basis for determining that the adverse event, incident, experience, or outcome represents an unanticipated problem;
- A description of any changes to the protocol or other corrective actions that have been taken or are proposed in response to the unanticipated problem.

To satisfy the requirement for prompt reporting, unanticipated problems will be reported to the IRB and to NIDCR within 2 weeks of the investigator becoming aware of the problem.

- All unanticipated problems should be reported to appropriate institutional officials (as required by an institution's written reporting procedures), the supporting agency head (or designee), and OHRP within one month of the IRB's receipt of the report of the problem from the investigator.

All unanticipated problems will be reported to NIDCR's centralized reporting system via Rho Product Safety:

- Product Safety Fax Line (US): 1-888-746-3293
- Product Safety Fax Line (International): 919-287-3998
- Product Safety Email: rho_productsafety@rhoworld.com

9 STUDY OVERSIGHT

The GPI, SPI and co-Investigators at UI/RI will be responsible for study oversight, including monitoring safety, ensuring that the study is conducted according to the protocol and ensuring data integrity. The SPI and co-Is will ensure that steps are in place to minimize loss of patient and practitioner confidentiality and will review the study procedures and data at regular intervals for confidentiality concerns. The SPI will be responsible for ensuring that practitioner enrollment and protocol deviations will be reported to NIDCR in a timely manner. To ensure data integrity, the SPI and co-Is will adhere to quality management processes (see Section 13).

The IU/RI will maintain the research records for this study, using the principles of GCP and complying with regulatory and institutional requirements for the protection of confidentiality of patients. The IU/RI will permit authorized representatives of NIDCR and regulatory agencies to examine (and when required by applicable law, to copy) research records for the purposes of QA reviews, audits, and evaluation of the study safety, progress and data validity.

10 STATISTICAL CONSIDERATIONS

10.1 Study Hypotheses

Since the proposed research is exploratory, we will not be testing specific study hypotheses.

10.2 Sample Size Considerations

We will recruit a purposive sample of 100 US dental practitioners from the National Dental PBRN based on the inclusion criteria each with an estimated average of >2,300 active dental patients per year (ADA 2008 Survey of Dental Practice). Over 5 years, if we conservatively assume that there will be an average of 2,500 unique patients per practice in the electronic dental records, and that 25% of the patients will have an average of 1PCR each, and 10% of the patients will have 1 RCT each, we can expect to have around 125,000 PCRs in 62,500 patients and 25,000 RCTs in 25,000 patients in the database (Simecek et al. 2009, Bernstein 2012, McCracken 2013). We assume a five-year failure rate of 24% among composite restorations (Kim et al. 2013) and a tooth loss rate of 6% (Ng et al. 2011 showed about 5.6% in 4 years). Based on these assumptions, we expect 30,000 failed restorations after PCR (125,000 x 0.24) and 1,500 tooth losses after RCT (25,000 x 0.06).

The power for detecting any predictor of a given effect size will be larger for PCR outcomes than for the RCT outcomes, due to the smaller number of patients with RCT and the lower rate of tooth loss. Hence, we focus the power for detecting risk factors for tooth loss following RCT. Since there will be multiple patients per provider, we account for intra-class correlation between patients nested within a provider by adjusting the sample size. With an average of 250 RCTs nested within a dentist, the effective sample size could be reduced from the actual sample size. In an extreme case, where all patients of each dentist belong to only one of 2 levels of an evenly split binary predictor, e.g. if half of all patients are female and treated by half of the dentists, and the males are treated by the other dentists, an assumed intra-class correlation of 0.1 would result in an effective sample size of 962 from the 25,000 RCTs. But we expect each of the predictors we investigate to be somewhat evenly distributed over a large number of dentists, so the effective sample size will be much less attenuated. For RCT, the effective total sample size can be reasonably expected to be >3,400, in which case we will have 90% power for detecting a difference in rate of tooth loss between 5% and 7% after 5 years, based on a logrank test at 5% significance level. Thus, we should have adequate power to detect all clinically significant predictors in both aims.

10.3 Analysis Plan

Statistical analysis to determine the factors influencing the outcome of PCRs and tooth loss outcomes of RCT will be performed using methods below.

Primary Objectives

For the primary objective we will first assess the extent to which the extracted EDR data can be used to assess PCR and RCT outcomes. This assessment will be based on the availability, completeness, concordance, and currency of the variables documented in the EDR. We will identify the variables collected in previous study protocols that evaluated PCR and RCT outcomes and evaluate the presence or absence of these data in EDR. We will also determine the range of variables available in the EDR, its completeness and concordance. Completeness will be measured as the proportion of dental records from participating practices that contain each study variable. Concordance will be measured as the proportion of available study variables identically recorded in free text (clinical findings) and coded (administrative) data in the EDR. Currency of the data will be measured to assess whether the study variables are relevant expressions at a given point in time.

A further investigation will be made of the ability to define the primary study outcomes of restoration longevity and tooth loss from EDR. For the former, both (1) the date of restoration and occurrence and (2) either the date of an “event” (new restoration, caries/secondary caries, RCT, or extraction) or the date of last examination establishing the continuing viability of the tooth. For the latter, evidence of the RCT and the tooth extraction must be present, preferably with dates for both events.

As part of data quality checking, we will compare rates of restoration failure and tooth loss with our assumptions based on current literature (24% failure and 6% loss).

Secondary Objectives

Based on the extent to which relevant study variables are available, we will select a subset of records with at least 60% complete and accurate data to assess posterior composite restoration longevity, time to loss of tooth on which RCT was performed and potential factors influencing these two outcomes. Records with at least 60% complete data will be determined based on the data quality assessment results and three researchers. Two researchers will independently review the results and select the records. The records selected by the two researchers will be included. Those records where the two researchers disagreed will be reviewed by the third researcher and reach a consensus on including or excluding the record from the analysis.

Statistical analysis to determine the factors influencing the outcome of PCR and tooth loss outcomes of RCT will be performed using methods appropriate for time-to-event data. These include estimation of survival using Kaplan-Meier curves and Cox proportional hazards regression. Each identified procedure will be followed forward in time by identifying all subsequent dental procedures, clinical and status codes for those teeth. Because multiple procedures may occur in the same individual we will use frailty

models (Therneau and Grambsch. 2000), an extension of the Cox regression model that accommodates correlated outcomes that result from the non-independence within person. For PCR, the method by Ha et al.(Ha et al. 2012) can be used to handle two-level nesting in the sample. We will examine the association of the clinical, demographic, medical history and dental care factors of interest with restoration survival using univariate models. We will then use the results of these models within each domain to determine the most important independent factors related to survival. For example, we will determine the most important clinical factors and the most important factors in the medical history. These can then be combined into a multiple regression model to assess the importance of each factor. Interactions among the variables will also be examined.

11 QUALITY CONTROL AND QUALITY ASSURANCE

Informed consent procedures will be performed by RCs, and the RAS will perform QA and quality control (QC) procedures of the consenting process. Documentation of practitioner consent will be maintained at the RAS.

For the QA/quality control (QC) activities associated with data extraction and processing, the data management plan (DMP) developed by IU/RI specifies data QA/QC procedures. In particular, the DMP details the following: Identification of data quality measures, development and execution of said measures within the database(s) and the processes related to the manual review of data, discrepancy management, data updates, data verification and approval, and database audit.

12 ETHICS/PROTECTION OF HUMAN SUBJECTS

12.1 Ethical Standard

The investigator will ensure that this study is conducted in full conformity with the principles set forth in The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, as drafted by the US National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (April 18, 1979) and codified in 45 CFR Part 46.

12.2 Institutional Review Board

This protocol will be reviewed by the National Dental PBRN Central Institutional Review Board (IRB). The UAB IRB for Human Use serves as the National Dental PBRN Central IRB.

Once the local institution has decided to use the National Dental PBRN Central IRB review, the National Dental PBRN Central IRB is the IRB responsible for the review of the protocol. The National Dental PBRN Central IRB then performs all future continuing protocol and informed consent reviews and amendment (new protocol version) reviews. Additionally, when recruitment and/or educational materials are to be used by all participating local institutions, the National Dental PBRN Central IRB will review and approve the language. The Central IRB also reviews unanticipated problems distributed by the Administrative Unit to local institution PIs.

Once the National Dental PBRN Central IRB has approved a protocol, local institutions are able to perform a Local Context Review (e.g. administrative review). If a RAS or other local institution elects not to use the National Dental PBRN Central IRB, the protocol, consent form(s), recruitment materials and all participant materials will be submitted to the RAS or other local institution IRB for review and approval.

Approval (either centrally for those regions who agree to central approval, or regionally for those who do not) of both the protocol and the consent form must be obtained before any participant is enrolled. Any amendment to the protocol will require review and approval by the IRB before the changes are implemented in the study.

12.3 Informed Consent Process

Informed consent is a process that is initiated prior to the individual agreeing to participate in the study and continues throughout study participation. The RCs will explain the research study to the practitioner participant, discuss risks and possible benefits of study participation, and answer any questions that may arise. If required by the responsible IRB, a consent form describing in detail the study procedures and risks will be given to the participant. Consent forms will be IRB-approved, and the participant is required to read and review the document. The participant will sign the informed consent document prior to any study-related assessments or procedures. Participants may withdraw consent at any time throughout the course of the study. A copy of the informed consent document will be given to participants for their records.

The consent process will be documented at the RAS.

12.4 Exclusion of Women, Minorities, and Children (Special Populations)

Practitioners of any gender or racial/ethnic group, women and minorities may participate. Patient data will be included in this study if they meet study inclusion criteria irrespective of age, gender or racial/ethnic group.

12.5 Participant Confidentiality

Practice and practitioner participant confidentiality and patient data are strictly held in trust by the investigators, study staff, and the sponsor(s) and their agents. The study protocol, documentation, data, and all other information generated will be held in strict confidence. No information concerning the study or the data will be released to any unauthorized third party without prior written approval of the sponsor.

Practices, practitioners and patients will be assigned unique identification numbers, which will be used to maintain study records and organize data transcripts. During the data extraction process from dental practices, the vendors will assign unique identification numbers to individual patient records. A document linking the unique identification number with patient record number used by practices will be shared with the respective RAS. This file will be kept in a password-protected file on the RAS's computer or filed in a locked cabinet, separate from all other research records.

After data extraction, practices may be contacted for clarification of data. The study team will rely on the RCs from the RAS to contact the practitioner and either query the practitioner on behalf of the study team or facilitate communication with the study team.

The data gathered will be in compliance with the Health Insurance Portability and Accountability Act Privacy and Security rules

<http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html>.

RI personnel have been trained in the importance of protecting participant confidentiality and will ensure that the investigators and staff maintain appropriate training in this area. Regenstrief employs best practice systems, service, and architectures to protect its network from internal and external threats. Elements from the border firewalls to intrusion detection software provide checks and balances for incoming and outgoing traffic. Externally facing resources are placed in firewalled demilitarized zone (DMZ) and as needed utilize application proxies to inspect and route traffic to internal resources. Inside the border, logical segments are created to limit access and create specific networks as access requirements dictate. File transfers to and from remote connections are encrypted using protocols such as secure shell (SSH), secure file transfer protocol (SFTP), or Internet Protocol Security (IPsec) VPNs with trusted partners.

Regenstrief security includes restricted access to Regenstrief sites, system security procedures and passwords which limit access to the least privileged, backup procedures that ensure against data loss, and secure file transfer which encrypts the transmission of data between Regenstrief and its clients.

13 DATA HANDLING AND RECORD KEEPING

Only study personnel (i.e., GPI, SPI, Co-I's, RCs, CC personnel, RI data core personnel and graduate students) will have access to the study data elements in the study database as described in Section 15.3 Types of Data. Study personnel will include those who are on the approved IRB study protocol and study personnel assigned work with the SPI and Co-I's on the project. All study personnel will have completed the required training elements for human subjects research certification.

The data gathered will be in compliance with the Health Insurance Portability and Accountability Act Privacy and Security rules

<http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html>.

13.1 Data Management Responsibilities

The IU/RI data core team Data Manager under the leadership of SPI and Co-Is, will be responsible for identifying relevant study variables from structured and unstructured data, assessing the completeness and availability of relevant data variables, extending the OHD ontology framework, and preparing the analytical datasets.

During the study period, when interim data analysis is needed, the RI Data Manager will coordinate the activities with the RCs and the Statistician. The interim datasets will be provided with the data collected as of the specified date. The data in those datasets will be cleaned if possible but may contain pending issues which will be provided to the Statistician if requested. The datasets will be provided to the Statistician via secure data transfer method.

The data extraction program will be developed and executed by Dentrix/EagleSoft personnel under SPI's guidance. Pseudonymization of patient identifier will be done at the time of extraction; de-identified data/databases from participating practices will be extracted. The volume of data extracted will be reduced for example, by excluding records of patients with less than one visit or treatment, short history of treatment at the practice, patient's age, etc.

13.2 Data Capture Methods

Dentrix and EagleSoft personnel will extract data from Dentrix and EagleSoft practices respectively. There is also a tentative plan that IU/RI personnel might run the program developed by Dentrix and EagleSoft to extract data from participating practices. The extracted data will then be transferred to IU/RI data core team who will perform initial data cleaning and preprocessing in collaboration with IU study personnel.

13.3 Types of Data

Data for the present study consist of the following:

- Practitioner level data from the National Dental PBRN enrollment questionnaire;
- Practice and practitioner characteristics data; and

- Patient demographics, reason for visit, medical and dental history, social history, tooth characteristics, and treatment.

13.4 Schedule and Content of Reports

Reports to monitor practitioner enrollment will be produced on a monthly basis during the study participant enrollment period, until enrollment targets are attained and enrollment is closed. These reports will contain a section for accrual information in aggregate and by region.

Reports to assess study progress will be produced annually, after the enrollment, data extraction and data analysis following data extraction. The study progress reports will contain frequencies and descriptive statistics for data quality metrics. The identification of key questions will be determined by the SPI and other study team members.

All of the reports listed above will be produced by the project manager under SPI's guidance and provided to the CC. The CC will distribute these reports to the National Network Director and NIDCR.

The procedure for locking the database prior to final analysis will be detailed in the study DMP. Briefly, the study data will be locked and the final datasets will be generated at the end of the study. Prior to locking the database, the designated IU/Regenstrief Institute staff will ensure all data is complete and clean. Then, the IU/Regenstrief Institute staff will obtain approval from the SPI to proceed with the data lock. The IU/Regenstrief Institute staff will then lock the database. The date and time of database lock will be documented. All team members will receive written notification from the IU/Regenstrief Institute staff when the database lock is complete.

No masking or coding is anticipated for this study.

13.5 Study Records Retention

Study records will be maintained for at least three years from the date that the grant federal financial report (FFR) is submitted to the NIH or longer as dictated by local IRB or state laws/regulations.

As outlined by IRB regulations, data will be destroyed in an appropriate and safe way after three years (e.g., and files will be securely deleted from computers).

The file linking the unique identification number with patient record number used by practices will be kept in a password-protected file or locked file cabinet by the respective RAS for a minimum of three years after conclusion of the study, in accordance with IRB regulations, before being securely destroyed.

13.6 Protocol Deviations

A protocol deviation (PD) is any noncompliance with the clinical study protocol or GCP principles. The noncompliance may be on the part of the patient, practitioner, SPI, or study staff. Any corrective actions resulting from deviations will be implemented promptly. All deviations from the protocol must be addressed in study subject source documents.

PDs must be reported to NIDCR at the timeframe within which PDs must be reported to IRBs. If the IRB overseeing the study protocol requires annual reporting of PDs to their IRB, that reporting frequency is acceptable to NIDCR.

14 PUBLICATION/DATA SHARING POLICY

This study will comply with the [NIH Public Access Policy](#), which ensures that the public has access to the published results of NIH funded research. It requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive [PubMed Central](#) upon acceptance for publication. All study personnel are required to read in its entirety and agree to abide by the network's "Data Analysis, Publications, and Presentations Policies" document. The current version of this policy is always kept at the network's public web site at <http://nationaldentalpbrn.org/publication.php>.

15 LITERATURE REFERENCES

- Bailit, H. (1996). "Health services research at Aetna Health Plans." *Med Care Res Rev* 53 Suppl: S77-81.
- Bailit, H. L. (2003). "Health services research." *Adv Dent Res* 17: 82-85.
- Bernstein, S. D., et al. (2012). "Outcomes of endodontic therapy in general practice: a study by the Practitioners Engaged in Applied Research and Learning Network." *J Am Dent Assoc* 143(5): 478-487.
- Brownstein, J. S., et al. (2007). "The tell-tale heart: population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction." *PLoS One* 2(9): e840.
- Burke, F. J. and P. S. Lucarotti (2009). "How long do direct restorations placed within the general dental services in England and Wales survive?" *Br Dent J* 206(1): E2; discussion 26-27.
- Demko, C. A., et al. (2008). "Concordance of chart and billing data with direct observation in dental practice." *Community Dent Oral Epidemiol* 36(5): 466-474.
- Duncan, W., et al. (2013). Representing Intracoronal Tooth Restorations in the Ontology for Oral Health and Disease. *International Conferences On Biomedical Ontologies 2013 Proceedings*: 109-112.
- Forss, H. and E. Widstrom (2004). "Reasons for restorative therapy and the longevity of restorations in adults." *Acta Odontol Scand* 62(2): 82-86.
- Goraieb, L. (2010). R-SPARQL.
- Ha, I. D., et al. (2012). "frailtyHL: A Package for Fitting Frailty Models with H-likelihood." *R Journal* 4(2): 28-37.
- Hayden, W. J. (1997). "Dental health services research utilizing comprehensive clinical databases and information technology." *J Dent Educ* 61(1): 47-55.
- Hommeaux, E. and A. Seaborne (2009). SPARQL query language for RDF W3C Recommendation, W3C.
- James, D. A. and S. DebRoy (2012). "RMySQL."
- Kahn, M. G., et al. (2007). "Configuration challenges: implementing translational research policies in electronic medical records." *Acad Med* 82(7): 661-669.
- Kim, K. L., et al. (2013). "The effect of clinical performance on the survival estimates of direct restorations." *Restor Dent Endod* 38(1): 11-20.

- Kohane, I. S. (2011). "Using electronic health records to drive discovery in disease genomics." *Nat Rev Genet* 12(6): 417-428.
- Levin, L., et al. (2007). "Cross-sectional radiographic survey of amalgam and resin-based composite posterior restorations." *Quintessence Int* 38(6): 511-514.
- Liu, K., et al. (2013). "Using Electronic Dental Record Data for Research: A Data-mapping Study." *J Dent Res* 92 (7): S90-S96
- McCracken, M. S., et al. (2013). "A 24-month evaluation of amalgam and resin-based composite restorations: Findings from The National Dental Practice-Based Research Network." *J Am Dent Assoc* 144(6): 583-593.
- Ng, Y. L., et al. (2011). "A prospective study of the factors affecting outcomes of non-surgical root canal treatment: part 2: tooth survival." *Int Endod J* 44(7): 610-625.
- Ng, Y. L., et al. (2011). "A prospective study of the factors affecting outcomes of nonsurgical root canal treatment: part 1: periapical health." *Int Endod J* 44(7): 583-609.
- Olsen, L., et al. (2007). *The Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine)*. Washington, DC, The National Academies Press.
- Opdam, N. J., et al. (2010). "12-year survival of composite vs. amalgam restorations." *J Dent Res* 89(10): 1063-1067.
- Pearson, J. F., et al. (2011). "Potential for electronic health records and online social networking to redefine medical research." *Clin Chem* 57(2): 196-204.
- Prokosch, H. U. and T. Ganslandt (2009). "Perspectives for medical informatics. Reusing the electronic medical record for clinical research." *Methods Inf Med* 48(1): 38-44.
- Rho, Y. J., et al. (2013). "Longevity of direct restorations in stress-bearing posterior cavities: a retrospective study." *Oper Dent* 38(6): 572-582.
- Ross, I. and R. Gentleman (1996). "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics* 5(3): 299-314.
- Schleyer, T., et al. (2013a). "Electronic dental record use and clinical information management patterns among practitioner-investigators in The Dental Practice-Based Research Network." *J Am Dent Assoc* 144(1): 49-58.
- Schleyer, T., et al. (2013b). "An ontology-base method for secondary use of electronic dental record data." *Proceeding - AMIA Joint Summits on Translational Science 2013*: 243-238

Seyfried, L., et al. (2009). "Enhanced identification of eligibility for depression research using an electronic medical record search engine." *Int J Med Inform* 78(12): e13-18.

Simecek, J. W., et al. (2009). "An evaluation of replacement rates for posterior resin-based composite and amalgam restorations in U.S. Navy and marine corps recruits." *J Am Dent Assoc* 140(2): 200-209; quiz 249.

Song, M., et al. (2013)". Reusing electronic patient data for dental clinical research: A review of current status." *J Dent* 41 (2013) 1148-1163

Sung, N. S., et al. (2003). "Central challenges facing the national clinical research enterprise." *JAMA* 289(10): 1278-1287.

Sunnegardh-Gronberg, K., et al. (2009). "Selection of dental materials and longevity of replaced restorations in Public Dental Health clinics in northern Sweden." *J Dent* 37(9): 673-678.

Terry, A. L., et al. (2010). "Using your electronic medical record for research: a primer for avoiding pitfalls." *Fam Pract* 27(1): 121-126.

Tyas, M. J. (2005). "Placement and replacement of restorations by selected practitioners." *Aust Dent J* 50(2): 81-89; quiz 127.

Therneau TM, Grambsch PM. (2000). *Modeling survival data: extending the Cox model.* Springer-Verlag, New York, 2000:169-229.

Van Nieuwenhuysen, J. P., et al. (2003). "Long-term evaluation of extensive restorations in permanent teeth." *J Dent* 31(6): 395-405.

Wilke, R. A., et al. (2007). "Use of an electronic medical record for the identification of research subjects with diabetes mellitus." *Clin Med Res* 5(1): 1-7.

APPENDICES

Appendix A. Schedule of Events

Appendix B. Draft List of Study Variables

Appendix C. Sample classes in the Health and Disease Ontology

Appendix D. Example of the OHD Ontology representing clinicians, patients, medical history, teeth and their anatomy, procedures.

APPENDIX A: Schedule of Events

Procedures	Practitioner Screening/ Enrollment (Phase Duration: ~ 6 months)	Data Extraction (Total duration up to 1 year)
Inclusion Criteria Met	X	
Obtain Consent	X	
Confirm approval of participation	X	
Contact practice for clarification of data	X	X

APPENDIX B: Draft List of Study Variables

Factors affecting outcome of RCT and PCR

Provider Information

Type of Provider –General Practitioner/Specialist
Age of practitioner
Type of salary structure
Frequency of changing dentist

Patient Demographics

Age/ Older Patients
Sex
Race/Ethnicity
Socioeconomic status
Dental insurance coverage

Clinical Findings

Oral cleanliness/ Oral Hygiene
Tooth number
Tooth type- Incisor, Canine, Premolar, Molar
Type of dentition - Primary/Permanent/Mixed
Jaw -Upper/Lower
Type of caries -Presence of primary or secondary caries
Cavity type - Class I/II/III/IV/V
Number of canals- Single/Multiple
Trauma
Proximal contacts with adjacent tooth
Pain on percussion and pressure
Swelling
Temperature sensitivity
Tenderness to palpation
Discharging sinus
Periodontal involvement – Yes/no
Extent of periodontitis – probing depth, mobility
Crack or fracture of tooth, root
Extractions/Tooth loss
Iatrogenic perforation
Patency at apical terminus; extrusion of root fillings,
Developmental Anomalies
Use of a rubber dam during treatment
Pre- and postoperative lesion depth estimates

Presence of Medical/Systemic History

Restorations

Site/ Surfaces of restoration /cavity (occlusal, mesial, distal, buccal or facial, lingual or palatal, incisal)

Size of restoration - Large direct restorations

Restoration type- Full/partial crown or onlay

Previous Restoration

Restorative materials used (amalgam; directly placed resin composite; indirect resin composite; glass ionomer; ceramic or porcelain; cast gold or other metal; combined metal/ceramic; temporary restorative material)

Reasons for placement (caries or non-caries defect)

Reasons for restoration of non-carious defects (abfraction/abrasion/erosion; developmental defect or hypoplasia; cosmetic reason; restore endodontically treated tooth; tooth fracture; poor anatomic form, other)

Restoration Fracture

Number of restored surfaces

Lost Restoration

Use of base/liner

Restoration Failure

Visit required to complete endodontic therapy -

One/two/three or more

Post placed – Yes/no

Type of post

Retentive pins

Endodontic Failure

Diagnosis

Periapical pathosis/infection

Irreversible pulpitis

Pulp Exposure

Hypersensitivity

Necrotic pulp

Endodontic Retreatment

Pulp vitality

Periapical Surgery

Treatment Plan

Antibiotic prescribed – Yes/no

Use of specific diagnostic techniques (clinical assessment, radiographs, optical techniques)

Diet counseling
Caries susceptibility testing
Fluoride prescriptions
Caries risk note
Caries activity note

APPENDIX C: Sample Classes in the Health and Disease Ontology

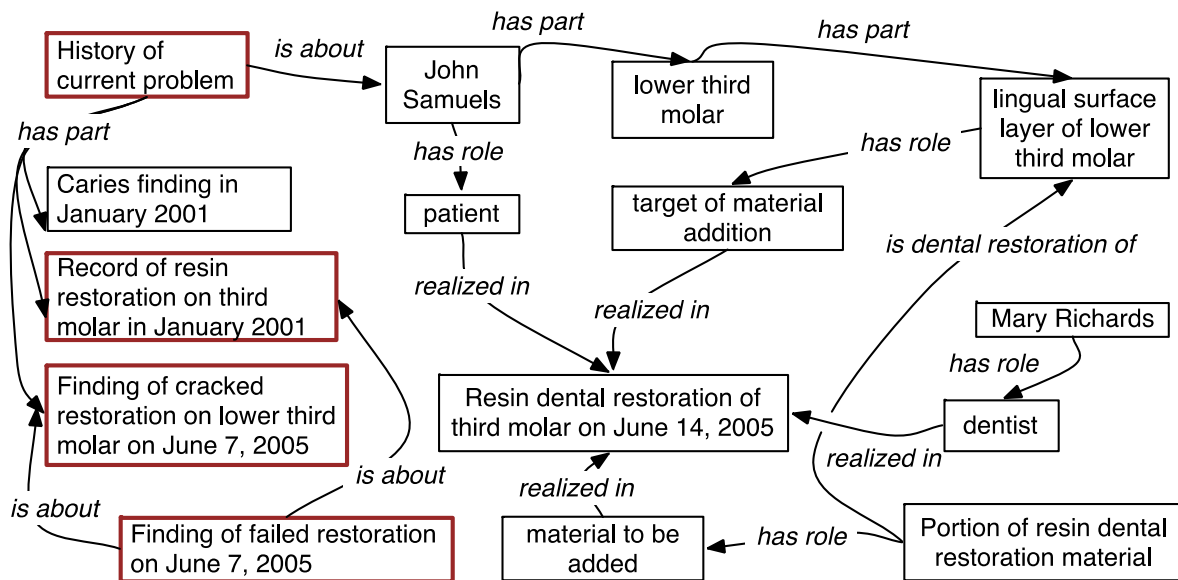
- Thing
 - BFO_0000007
 - planned process
 - health care process
 - health care encounter
 - clinical history taking
 - dental procedure
 - endodontic procedure
 - surgical dental procedure
 - tooth extraction procedure
 - tooth restoration procedure
 - crown restoration
 - inlay restoration
 - intercoronal tooth restoration
 - direct restoration
 - amalgam filling restoration
 - resin filling restoration
 - indirect restoration
 - outpatient encounter
 - dental visit
 - BFO_0000016
 - Disease
 - acquired genetic disease
 - constitutional genetic disease
 - dental caries
 - BFO_0000023
 - dental health care provider role
 - dental assistant role
 - dental hygienist role
 - dentist role
 - BFO_0000031
 - information content entity
 - CRID
 - current dental terminology code
 - billing code for diagnostic
 - billing code for clinical oral evaluations

- billing code for oral and maxillofacial surgery
 - billing code for extractions (includes local anesthesia, suturing, if needed, and routine postoperative care)
 - billing code D7110: single tooth
 - billing code D7120: each additional tooth
 - billing code D7140: extraction, erupted tooth or exposed root (elevation and/or forceps removal)
 - billing code for restorative
- data item
 - clinical finding
 - dental finding
 - caries finding
 - fractured tooth finding
 - missing tooth finding
 - missing tooth 1 finding
 - missing tooth 2 finding
 - no oral health issues reported
 - tooth clinically present finding
 - unerupted tooth finding
- BFO_0000040
 - Material anatomical entity
 - Anatomical structure
 - Cardinal organ part
 - Organ
 - Cavitated organ
 - Organ with organ cavity
 - Tooth
 - Molar tooth
 - Primary molar tooth
 - Secondary molar tooth
 - Lower secondary molar tooth
 - Lower first secondary molar tooth
 - Lower second secondary molar tooth
 - Lower third secondary molar tooth
 - Upper secondary molar tooth
 - Premolar tooth
 - processed material
 - ceramic dental restoration material

- metal dental restoration material
 - amalgam dental restoration material
 - gold dental restoration material
 - noble metal dental restoration material
- resin dental restoration material

Appendix D: Example of the Oral Health and Disease Ontology representing clinicians, patients, medical history, teeth and their anatomy, procedures.

Through previous work, we have developed an initial version of the OHD that represents many concepts of interest to the proposed study already (Duncan et al. 2013, Schleyer et al. 2013b). The purpose of the OHD is to provide a platform-neutral, generalizable mechanism for representing information about oral health and disease in humans. At this time, the exclusive source of the data the OHD represents are electronic dental records from actual clinical practices. Therefore, the OHD is grounded in the realities and limitations of records-keeping in clinical practice. Currently, the primary use of OHD-represented data is to advance our state of knowledge about oral health and disease, the efficacy of dental treatments, and related topics. We plan to extend the OHD to integrate data from different EDR systems and different providers. This appendix outlines the major types of information represented in the OHD at this time. They include entities such as dental care providers, patients, teeth and their status, surfaces of teeth, restoration materials, and restoration types. The OHD is the foundation for representing additional clinical data relevant to this study. Red borders indicate examples of extensions to be made under this study.



Note: To conserve space, the figure does not include all relevant entities and relationships available in the OHD.