Identifying Oral Health Concepts from Previously "OMOP-ified" Data

Robert Koski[1], Danielle Boyce[1], Benjamin Martin[1], Julie Cha[2], Maria Sanchez[1]
Johns Hopkins University[1], Georgia Institute of Technology[2]

Background

Electronic health record usage is ubiquitous in patient care. Medical systems, dental offices, insurance organizations, and other participants of the healthcare ecosystem almost universally utilize electronic records to capture patient information during the course of care. Patient notes, insurance claims, and registry data are common repositories of patient information. Many healthcare organizations recognize the value of standardizing medical records to a common data model, like the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), toward accomplishing observational research in their patient populations[1].

Despite the recent trend toward utilizing electronic data for observational research, dentistry has not widely adopted a common data model[2]. Siloed practices[3], challenging data structures, and terminology issues have been identified some of the top barriers to dentistry adopting a common data model for observational research. Few, if any, dental datasets have been converted to the OMOP-CDM or any other common data model. This makes conducting observational research at scale a Herculean effort that few have managed to overcome.

Now more than ever, oral health research needs novel approaches to conducting research and further uncovering the links between oral health and systemic disease. It is not feasible to wait to conduct oral health research until the barriers to wide adoption of a common data model are completely solved. The Observational Health Data Sciences and Informatics (OHDSI) Dentistry Workgroup has identified readily available already "OMOP-ified" (data that has been converted to the OMOP-CDM) data sets that contain dental use cases. Datasets with OMOP-ified data present an outstanding opportunity for oral health observational research to be readily conducted. Our group's goal is to identify additional dental cohorts and engage with eager collaborators who want to use their data to solve oral health problems. This study aims to summarize data profile counts and concept distribution patterns across different OHDSI datasets and investigate their impact on study feasibility to facilitate cross-institutional and network studies.

Methods

Two dentists and three individuals with significant expertise in the OMOP-CDM and cohort development were selected to identify oral health related OMOP-CDM concepts for use in identifying oral health use cases from existing OMOP-ified datasets. The goal of the concept search was to develop a cohort that OHDSI collaborators could utilize to identify if their OMOP instances contained any existing dental concepts that could subsequently be used in an observational study.

Concepts were selected with the following criteria:

1. Related to the delivery of oral healthcare (as identified by the dentists)
2. Existed in the OMOP-CDM (identified using Athena[4] by the OMOP-CDM experts)

The group primarily focused on procedure and condition concepts as these would be the most likely concepts to be encountered during the course of patient care. Concepts that were selected by the dentists were evaluated in Athena to investigate the parent concept(s). If the parent concept contained multiple oral health related concepts without other unrelated child concepts, the parent concept was retained and added to the dental discovery cohort.

When the group exhausted the potential concepts, the dental discovery cohort was used to generate a total patient count with the Johns Hopkins University OMOP instance and the Evidence network counts on the ATLAS demonstration platform[5].

Results

The resulting first iteration of the dental discovery cohort is fully described here. The cohort includes 2968 concepts relating to oral health conditions or procedures. Additional concepts were identified to narrow the cohort to specific demographic patient populations and patient populations with other systemic conditions, but these concept sets were not included in the final cohort. The cohort can be used by any collaborator with a dataset that has been converted to the OMOP-CDM.

The dental discovery cohort yielded a total of 172,584 patients and records in the Johns Hopkins University de-identified cohort discovery OMOP instance. Further defining the cohort to include events where a dental procedure occurred within 180 days of a dental condition resulted in 16,512 patients and 35,313 events, using the same concept sets refined for the discovery cohort. Demographic and clinical characterization of this initial cohort is provided in Table 1, below.

The ATLAS evidence network on the ATLAS demonstration instance counts yielded 44,947 patients and records.

We are investigating dental use cases that could potentially be included in these existing datasets. Potential use cases may be found in practice areas such as:

- Dental trauma or emergent dental conditions
- Oral maxillofacial surgery
- Orofacial pain
- Hospital dentistry and oral medicine
- Dental oncology
- Pediatric and special needs dentistry conducted under general anesthesia

Domain	Variable	Value
Gender	Male	53.5%
	Female	46.5%
Race	White	65.4%
	Black or African American	17.6%
	Asian	3.7%
	Other	13.2%
AgeGroup	0 - 4	10.0%
	5 - 9	5.6%
	10 - 14	5.7%
	15 - 19	6.0%
	20 - 24	3.6%
	25 - 29	2.6%
	30 - 34	3.4%
	35 - 39	3.7%
	40 - 44	4.1%
	45 - 49	4.2%
	50 - 54	5.6%
	55 - 59	8.0%
	60 - 64	8.6%
	65 - 69	8.9%
	70 - 74	8.6%
	75 - 79	6.0%
	80 - 84	4.0%
	85 - 89	1.3%
PriorObservationTime	observation time (days) prior to index	635.45
PostObservationTime	observation time (days) after index	754.73
VisitCountShortTerm	visit_occurrence count during day -30 through 0 relative to index	3.93

Table 1. Dental Cohort Clinical and Demographic Characterization

Conclusion

The dental discovery cohort demonstrates that mapped oral health concepts are present in existing OMOP-ified datasets. Further investigation of the present data is necessary to determine feasibility for study of specific cohorts. Additionally, the oral health concepts included in the discovery cohort may have inadvertently identified concepts associated with cohorts unrelated to oral health use cases. The OHDSI Dentistry Workgroup plans to refine and improve the dental discovery cohort so that OHDSI collaborators can identify their own existing oral health data and prompt further investigation and discovery of oral health cohorts.

References

1. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, DeFalco FJ, Londhe A, Zhu V, Ryan PB. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc. 2015;22(3):553–64. doi:10.1093/jamia/ocu023
2. Benoit B, Frédéric B, Jean-Charles D. Current state of dental informatics in the field of health information systems: a scoping review. BMC Oral Health. 2022;22(1):131. doi:10.1186/s12903-022-02163-9
3. Thyvalikakath TP, Duncan WD, Siddiqui Z, LaPradd M, Eckert G, Schleyer T, Rindal DB, Jurkovich M, Shea T, Gilbert GH. Leveraging Electronic Dental Record Data for Clinical Research in the National Dental PBRN Practices. Appl Clin Inform. 2020;11(2):305–314. doi:10.1055/s-0040-1709506
4. Athena. https://athena.ohdsi.org
5. Atlas OHDSI. https://atlas-demo.ohdsi.org