Towards the Reproducible Imaging Research: Implementation of Multi-modeal research in Alzheimer's integrating DICOM data with the OMOP CDM

Woo Yeon Park[1], Ben Martin[1], Gabriel Salvador[1], Blake Dewey[1], Teri Sippel Schmidt[1], Paul Nagy[1]

[1]Johns Hopkins University

Background

According to the 2023 OCED Health Statistic data, the United States conducted the highest number of clinical imaging scans in 2021 among OCED countries, with 369 CT, MRI, and PET exams combined per 1,000 population[1]. The abundance of clinical images presents a valuable opportunity for researchers to access a diverse and large dataset and mitigate embedded bias. Integrating clinical outcomes and other EHR-based data elements can further enhance imaging research. Our research team previously published a paper on developing an imaging extension model for the OMOP CDM to standardize integration of data for observational imaging research.

This study implemented a prototype of the OMOP CDM imaging extension and reproduced an imaging research study[2]. We used the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, a public dataset curated for Alzheimer's Disease imaging research which includes medical imaging data along with clinical codified measures. We reproduced the cohort definition of the research done by Boublay et al. on evaluating the correlation between brain volume and psychological symptoms[3].

Method

The study created an evaluation environment of a Postgres SQL database of OMOP CDM and Atlas instance to demonstrate the OMOP CDM imaging extension model for ADNI data. Our methodology involved three key phases.

1. **Extracted Digital Images in Communications in Medicine (DICOM) Standard**: In this phase, we harmonized DICOM terminology, including attributes and value sets, by extracting the standard data model according to Information Object Definitions (IODs) from DICOM standards Parts 3, 6, and 16. These were then added as custom concepts in the OMOP CDM, updating the concept, domain, and vocabulary tables accordingly.
2. **Data Transformation and Ingestion**: We downloaded and transformed patient demographic and neuropsychiatric inventory files to update the Person and Measurement tables. DICOM studies from the ADNI data archive were ingested. Using the DICOM tags and custom concepts created, we extracted patient, visit, and procedure data, formatting them into the person, visit occurrence, procedure occurrence, and image occurrence tables. Additional DICOM tag information, such as details about image acquisition parameters, was extracted and organized into the Image feature and Measurement tables. The Observation period table was updated using the combined data per patient.
3. **Phenotype Definition in Atlas**: The cohort definition was recreated using Atlas. The criteria included having done a T1-weighted Brain MRI scan, evaluated neuropsychiatric inventory score (NPI), and Alzheimer's disease diagnosis.

Result

We have extracted 2,824 non-long text type DICOM attributes and 5,223 DICOM Value Sets from the standard and added them as custom concepts in OMOP CDM. We ingested 100 DICOM studies, which included 948 DICOM series and 14,816 images for 35 patients. This resulted in 457,607 imaging features being organized in the Image feature and Measurement tables. There were 4,152 patient demographics, including sex, race, ethnicity, year, and month of birth, were added to the Person table. NPI scores included 12 subtotals per domain and a total score; in total, 88,819 were mapped to the Measurement table.

In Atlas, the cohort was generated using imaging features and clinical observations. Table 1 shows examples of concept IDs that were included in recreating the cohort. The figure illustrates how both imaging and EHR concept sets created the cohort. The cohort definition identified 8 patients with 26 DICOM image series.

Table 1. Integration of DICOM tags to OMOP CDM as custom concepts for imaging extension model

Source	Standard	Concept Name	Code/Tag	OMOP Concept ID
EHR	LOINC	Age	30525-0	3022304
EHR	LOINC	Sex	46098-0	3046965
EHR	SNOMED	Alzheimer's Disease	26929004	378419
EHR	SNOMED	Neuropsychiatric Inventory	1231514008	37157689
Image	DICOM	Inversion Time	(0018,0082)	2128000368
Image	DICOM	Repetition Time	(0018,0080)	2128000366
Image	DICOM	Echo Time	(0018,0081)	2128000367

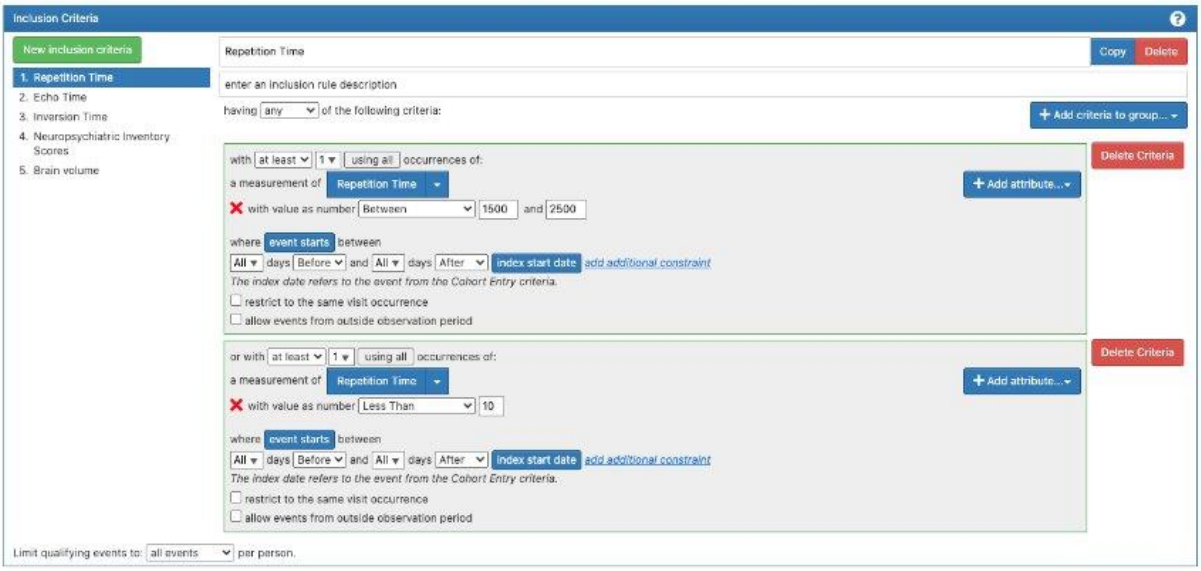

Figure 1. Screenshot of Atlas phenotype definition using imaging features.

Conclusion

Our study successfully demonstrated a working pilot implementation of the OMOP CDM imaging extension using the ADNI dataset, effectively integrating imaging data with clinical and demographic information. By harmonizing DICOM terminology and creating custom concepts within the OMOP CDM, we enhanced the standardization and interoperability of imaging data. This approach allowed

for the ingestion and transformation of a substantial amount of imaging and clinical data, facilitating the creation of a comprehensive dataset for observational research.

Data integrity and reproducibility present challenges in imaging research, impacting patients, providers, and researchers. Recent articles have highlighted integrity issues in Alzheimer's disease studies, ranging from photo editing to data manipulation[4]. This affects the reproducibility and generalizability of studies. Consequently, a reproducible and reliable evidence management framework is essential.

The results highlighted the feasibility and effectiveness of the OMOP CDM imaging extension model in addressing the challenges of data integrity and bias in imaging research. By integrating diverse and large datasets, researchers can mitigate embedded biases and improve the reproducibility and generalizability of their studies. Additionally, the phenotype definition created in Atlas demonstrated the practical application of this model in real-world research scenarios.

Our study provides a robust framework for managing and analyzing imaging data within the OMOP CDM, paving the way for more reliable and reproducible imaging research. This framework can be extended to other datasets and research questions, further enhancing the value and impact of imaging research in healthcare.

References

1. OECD (2023), Health at a Glance 2023: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/7a7afb35-en.
2. Park WY, Jeon K, Schmidt TS, et al. Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension. J Imaging Inform Med. 2024;37(2):899-908. doi:10.1007/s10278-024-00982-6
3. Boublay N, Bouet R, Dorey JM, et al. Brain Volume Predicts Behavioral and Psychological Symptoms in Alzheimer's Disease. J Alzheimers Dis. 2020;73(4):1343-1353. doi:10.3233/JAD-190612
4. Ault A. Integrity issues rampant in AD research, say investigators. Medscape. March 12, 2024. Accessed June 12, 2024. https://www.medscape.com/viewarticle/integrity-issues-rampant-alzheimers-research-say-2024a10004my.