Characterizing the Temporality of OMOP CDM Concepts in a Mastectomy Phenotype

Matthew Spotnitz, MD, MPH[1], Yechiam Ostchega, PhD, RN[1], Stephanie L. Goff, MD[2], Lakshmi Priya Anandan, MPH[1], Emily Clark, MPH[1], John Giannini, PhD[1], Lew Berman, PhD, MS[1]

[1]*All of Us* Research Program, Office of the Director, National Institutes of Health, Bethesda, Maryland

[2]Surgery Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

Introduction

Temporality measures whether data set values and elements have timestamps that follow an expected order [1]. It is an essential component of observational healthcare data quality assessments [2]. Also, using temporal signals may allow for effective phenotyping of complex medical conditions [3]. However, standardized approaches to characterizing temporal electronic health record (EHR) data are poorly established. In this analysis, we developed a method to describe the sequence of Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) concepts in a mastectomy phenotype.

Methods

Our phenotype was the first occurrence of any CPT4 or SNOMED code for a mastectomy procedure. With ATHENA(https://athena.ohdsi.org/search-terms/start), we selected OMOP CDM concepts for clinical measurements and interventions for use in our analysis. In the absence of having differentiation of neoadjuvant and adjuvant therapy in the concept names, we categorized our concepts according to 4 National Comprehensive Cancer Network (NCCN) phases: Workup and Diagnosis, Primary Treatment, Adjuvant Therapy, and Post-surgical Treatment. All analyses were conducted in the *All of Us* Researcher Workbench Controlled Tier Dataset v7 release. All programming and statistical analyses were performed using Microsoft Excel, Python version 3.7.12, and R version 4.1.1. Python and R analyses were implemented in Jupyter Notebook version 6.5.4.

Results

We plotted the proportions of patients with concepts in each category by sequence order (refer to Figure 1). Most patients had workup and diagnosis concepts first, followed by surgical treatment (e.g. mastectomy) concepts.
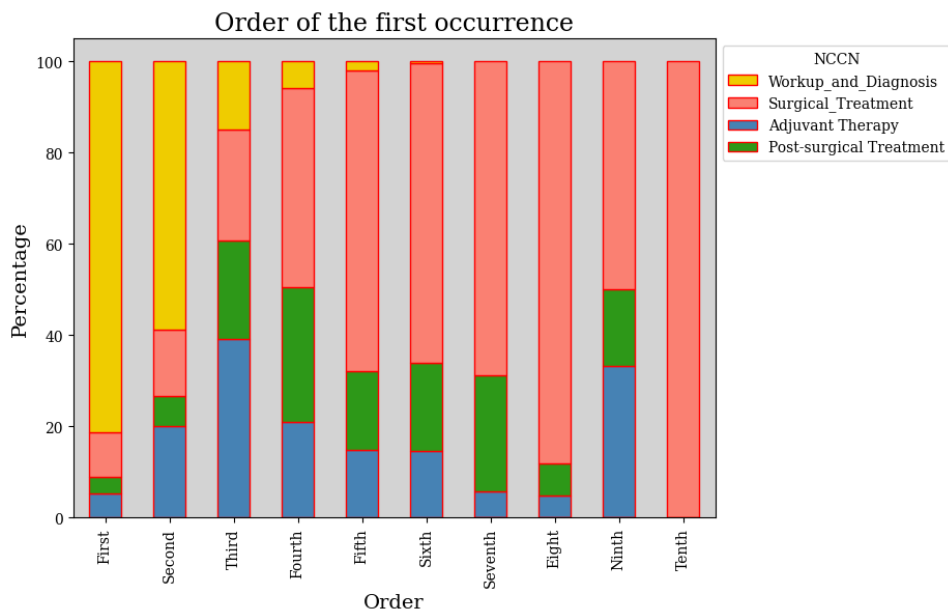

Figure 1: Stacked Bar Chart of OMOP CDM Mastectomy Concepts by National Comprehensive Cancer Network (NCCN) Phase and Sequence Order. Data Source: The *All of Us* Program.

The inability to differentiate neoadjuvant from adjuvant therapy based on concept names was a major limitation of our analysis. Consequently, we were unable to determine to what extent our temporal patterns were consistent with NCCN guidelines. Characterizing the time between surgery and radiation or pharmacological therapy may improve the granularity of a temporal analysis. In conclusion, we have presented a novel method for characterizing temporal patterns in OMOP CDM concepts. However, further refinement is warranted.

References

[1] Berman L, Ostchega Y, Giannini J, et. al. Application of a Data Quality Framework to Ductal Carcinoma in Situ Using Electronic Health Record Data from the All of Us Research Program. JCO Clinical Cancer Informatics. Submitted.

[2] Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC) 2016;4(1):1244. Published 2016 September 11. PMID: 27713905.

[3] Albers DJ, Elhadad N, Tabak E, Perotte A, Hripcsak G (2014) Dynamical Phenotyping: Using Temporal Analysis of Clinically Collected Physiologic Data to Stratify Populations. PLoS ONE 9(6): e96443. doi:10.1371/journal.pone.0096443