

Assessing Temporal Data Quality of Coronary Revascularization Procedure Patients in the All of Us Research Program

Matthew Spotnitz, MD, MPH¹, John Giannini, PhD¹, Jennifer Adjemian, PhD¹, Yechiam Ostchega, PhD, RN (retired)¹, Tamara R Litwin, PhD, MPH¹, Lewis Berman, PhD, MS (retired)¹, Geoffrey S. Ginsburg, MD¹

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

Background

Coronary artery disease (CAD) is prevalent in the United States and is a leading cause of cardiovascular morbidity and mortality [1,2]. Previously, we described a data quality framework that has been applied to multiple All of Us Research Program cohorts [3-7]. In this analysis, we used the same framework to evaluate participants who had coronary artery bypass grafting (CABG) or percutaneous coronary intervention (PCI).

One data quality dimension evaluated by this framework is temporality, which characterizes whether clinical events occur in plausible order and with interpretable timing. Temporal relationships can inform complex phenotyping and the selection of temporal anchors for observational research [8,9]. We therefore evaluated the ordering, timing, and completeness of clinically relevant events surrounding CABG and PCI.

Methods

We defined CABG and PCI cohorts using Current Procedural Terminology, Version 4 (CPT4), International Classification of Diseases (ICD), Healthcare Common Procedure Coding System (HCPCS), and Systematized Nomenclature of Medicine (SNOMED) codes. For each participant, we identified the earliest observed occurrence of five clinical concept sets: (1) lipid panel measurement, (2) lipid-lowering agent, (3) CAD diagnosis, (4) diagnostic catheterization, and (5) CABG or PCI procedure. These concepts represent risk-factor assessment and treatment, disease diagnosis, diagnostic evaluation, and coronary revascularization [10].

Participants with observations for all five concepts were considered temporality eligible. For each eligible participant, the five earliest observed events were ordered chronologically from earliest to latest; thus, no artificial “start” event was assigned. Same-day events were retained as zero-day transitions. We summarized the most frequent ordered pathways and calculated median transition times with interquartile ranges (IQRs) between consecutive events. We also quantified retention from the full cohort to the temporality-eligible subset and the proportion of eligible participants represented by the four most frequent pathways.

Results

The CABG and PCI cohorts included 3,766 and 9,172 participants, respectively. Complete data for all five concepts were available for 1,460 CABG participants (38.8%) and 3,147 PCI participants (34.4%). The four most frequent pathways accounted for 31.8% of temporality-eligible CABG participants and 28.9% of temporality-eligible PCI participants, demonstrating substantial heterogeneity in event ordering.

Table 1. Temporality-analysis retention and pathway concentration in the CABG and PCI cohorts.

Measure	CABG	PCI
Full cohort, N	3,766	9,172
Complete data for all 5 concepts, n (%)	1,460 (38.8%)	3,147 (34.4%)
Included in top 4 pathways, n (% of eligible)	465 (31.8%)	911 (28.9%)
Included in top 4 pathways, % of full cohort	12.3%	9.9%

In the most frequent CABG pathway (n=168), the sequence was lipid panel -> lipid-lowering agent -> CAD diagnosis -> diagnostic catheterization -> CABG. Median consecutive transition times were 768 (IQR 2,343), 910 (1,956), 156 (1,973), and 25 (348) days. In the most frequent PCI pathway (n=303), the sequence was lipid panel -> lipid-lowering agent -> CAD diagnosis -> PCI -> diagnostic catheterization; corresponding median transition times were 732 (2,217), 672 (1,736), 772 (2,953), and 0 (0) days. Thus, pathways differed substantially at the beginning but converged around diagnostic catheterization and revascularization near the end.

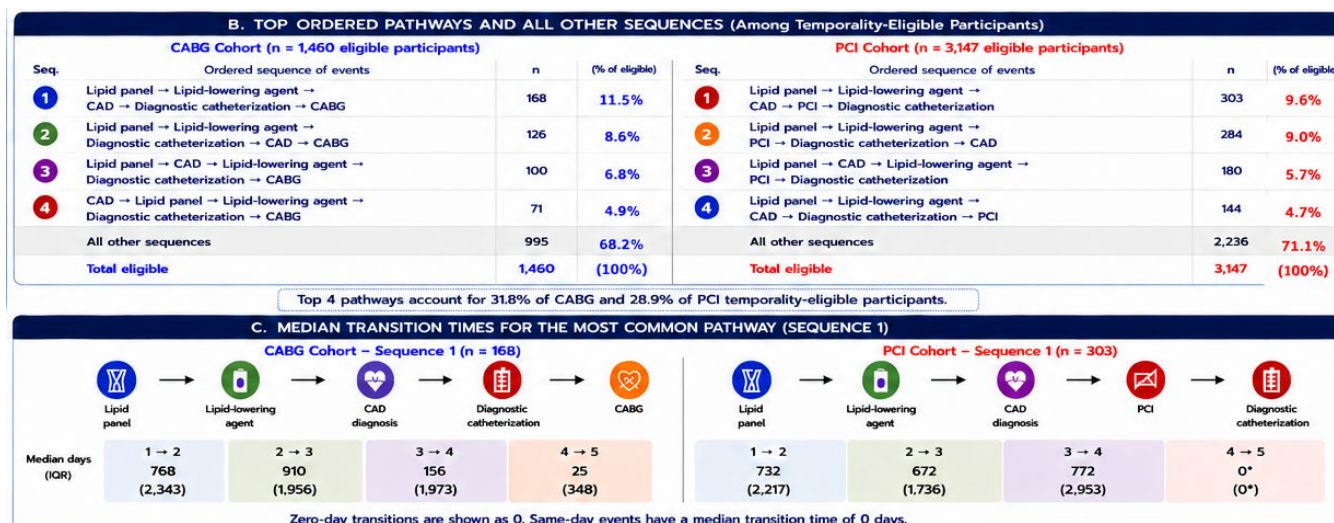


Figure 1. Top ordered pathways and median transition times for the most common pathway among temporality-eligible CABG and PCI participants. Values are median days with IQR in parentheses; same-day transitions are shown as 0 days. CABG, coronary artery bypass grafting; CAD, coronary artery disease; IQR, interquartile range; PCI, percutaneous coronary intervention.

Conclusion

We successfully applied a multidimensional data quality framework to evaluate temporality in CABG and PCI cohorts. Temporal pathways were heterogeneous, particularly for earlier risk-factor and disease-related events, but diagnostic catheterization and coronary revascularization were concentrated near the end of the observed sequences. Fewer than 40% of either full cohort had observations for all five concepts, indicating that temporal completeness is itself an important data quality consideration. These findings suggest that CABG and PCI procedure codes may be useful temporal anchors for events proximal to revascularization, while longitudinal completeness and pathway concentration should be evaluated before their use as anchors in phenotyping or downstream evidence generation. A reproducible temporality assessment incorporating event completeness, pathway ordering, and transition-time distributions could complement established OHDSI data quality tools and facilitate comparisons across data sources.

References

- [1] Duggan JP, Peters AS, Trachiotis GD et. al. Epidemiology of Coronary Artery Disease. *Surg Clin N Am* 102 (2022) 499–516. doi: 10.1016/j.suc.2022.01.007.
- [2] Palaniappan LA, Allen NB, Almarzooq ZI et. al. 2026 Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association. *Circulation*. 2026; 153(9):e275-906. doi: 10.1161/CIR.0000000000001412.
- [3] 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*. 2024;8:e2400052. doi: 10.1200/CCI.24.00052.
- [4] Spotnitz M, Giannini J, Ostchega Y, et. al. Assessing the Data Quality Dimensions of Partial and Complete Mastectomy Cohorts in the All of Us Research Program: Cross-Sectional Study. *JMIR Cancer* 2025;11:e59298. doi: 10.2196/59298
- [5] Spotnitz M, Giannini J, Clark E et. al. Assessing the Data Quality Dimensions of Surgical Oncology Cohorts in the All of Us Research Program. *JCO Clinical Cancer Informatics*, 2025 9:e2500078. doi: 10.1200/CCI-25-000
- [6] Spotnitz M, Giannini J, Clark E et. al. Assessing Data Quality of Rheumatoid and Psoriatic Arthritis Patients in the All of Us Research Program. *JAMIA Open* 2026, 9(2), ooag028. doi: 10.1093/jamiaopen/ooag028
- [7] Spotnitz M, Faye AS, Giannini J, et. al. Assessing Data Quality of Inflammatory Bowel Disease Patients in the All of Us Research Program. *JAMIA Open* 2026 9(3), ooag084. doi: 10.1093/jamiaopen/ooag084
- [8] 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.
- [9] 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.
- [10] Slipeczuk L, Blankstein R, Bucciarelli-Ducci C et. al. State of the Art: Evaluation and Medical Management of Nonobstructive Coronary Artery Disease in Patients With Chest Pain: A Scientific Statement From the American Heart Association. *Circulation*. 2025;152:e443–e466. doi: 10.1161/CIR.0000000000001394. []