Poster Code: B-257

When One Code Defines Millions: Sensitivity of Peripheral Artery Disease Cohorts to Operational Definition Choices in Real-World Data

Scan for poster + more

Scott L DuVall PhD,[1] Jared Kamauu BFA,[2] Aimee Harrison MFA,[2] Michael Buck PhD,[2] Craig Parker MD MS,[2] Allise Kamauu MS,[2] Aaron Kamauu MD MS MPH[2]

[1]PurpleLab Healthcare Analytics, USA; [2]Navidence Inc., USA

So What?

Operational definition choices can produce significant differences in cohort size – driven by a small number of high-impact diagnosis codes rather than overall code list size.

PART OF A 4-POSTER SERIES • Poster Session B — ◆ YOU ARE HERE ◆

The work in this poster stems from companion studies: Foundation ▸ Methodology ▸ Application

B-269	B-271	B-257	B-256
FOUNDATION	METHODOLOGY	APPLICATION • COHORT COVERAGE	APPLICATION • COHORT MAKE-UP
Identifying Operational Definition Variation of ASCVD	**Quantifying Cohort Variation in Real-World Data**	**Sensitivity of PAD, MI, Stroke & TIA Cohorts to Code List Choices**	**Cohort Characterization Variation in Real-World Data by Code List Choices**
Scientific Upstream – how ASCVD is defined across the spectrum of diagnosis codes.	*Methodology Bridge – different ASCVD definitions produce dramatically different study cohorts.*	*This Poster – small shifts in operational definition result in large cohort impact.*	*Downstream Impact – cohort characterization differences by small differences in code lists.*

BACKGROUND

Context

Composite phenotypes such as peripheral artery disease (PAD) are commonly defined in real-world research (RWR) using diagnosis code lists derived from prior studies. Published code lists vary widely in size and composition

- Peripheral Artery Disease (PAD): from 52-351 ICD-10-CM codes
- Myocardial Infarction (MI): from 22-25 ICD-10-CM codes
- Stroke: 118-171 ICD-10-CM codes
- Transient Ischemic Attack (TIA).: 2-11 ICD-10-CM codes

The Gap / Problem

While variability in these operational definitions is recognized (see *companion posters **B-269, B-271 & B-256***), the population-level impact of individual codes is rarely quantified. It is unclear whether cohort size is driven by the number of codes or by which specific codes are included.

OBJECTIVES

Primary Objective

To assess the sensitivity of PAD cohort size to operational definition choices in real-world data, and to identify the diagnosis codes that most strongly drive cohort inclusion.

Secondary Objectives

- Quantify code-level contributions to PAD cohort inclusion
- Identify individual high-impact ICD-10-CM codes that disproportionately affect cohort size
- Evaluate generalizability via parallel sensitivity assessments for MI, Stroke and TIA

LIMITATIONS

Findings are based on PurpleLab® CLEAR Claims; results from other RWD sources may vary.

Code-level characterizations of other sub-populations planned as next step.

Computable Operational Definitions (CODefs™)

Study elements — exposures, outcomes, and covariates — were defined using Navidence CODefs™: rigorous, data source-agnostic algorithms that ensure precision and reproducibility across RWR.

METHODS

Study Design & Data Source

Cross-sectional cohort replication study. Multiple published real-world cohort definitions for PAD, MI, stroke and TIA, with explicit ICD-10 code lists were identified from the literature, then precisely replicated (matching codes exactly) in PurpleLab® CLEAR Claims, US claims database.

Figure 1: PAD Cohort Replication Workflow

Targeted Literature Review
Published PAD cohort definitions with explicit ICD-10-CM code lists

↓

6 PAD Cohort Definitions Selected
52–351 ICD-10-CM codes each

↓

Exact Replication in PurpleLab® CLEAR Claims
US national claims database

↓

Compared Cohorts + Identified High-Impact Codes
Code- and patient-level overlap; per-code contribution

DISCLOSURES & ACKNOWLEDGMENTS

A Kamauu, J Kamauu, M Buck & C Parker: Employees of Navidence Inc. A Kamauu, C Parker: Co-Founders of Navidence Inc. A Harrison: Employee of Navidence Inc. during this work. S DuVall: Employee of PurpleLab.
No conflicts of interest to disclose.

RESULTS

6	52–351	5.8–8.8M
PAD Cohort *Definitions Replicated*	**ICD-10-CM Codes** *per PAD Cohort Definition*	**Patients Identified** *across the 6 PAD cohorts*

Figure 2: PAD Code List Size vs. Cohort Size

~6.8× more codes → ***only*** **~1.5×** more patients

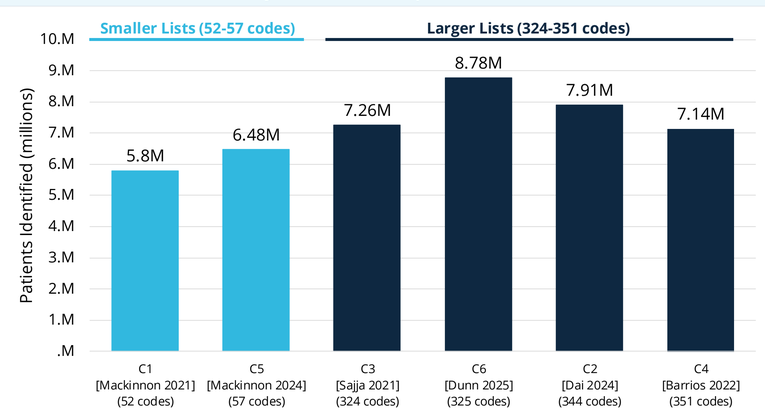
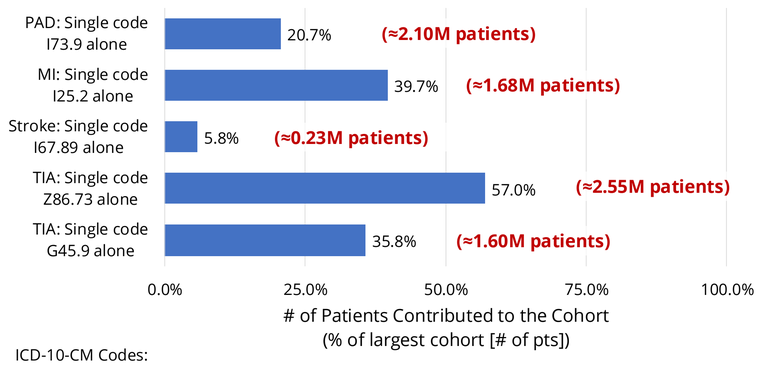


Figure 3: Single-Code Impact

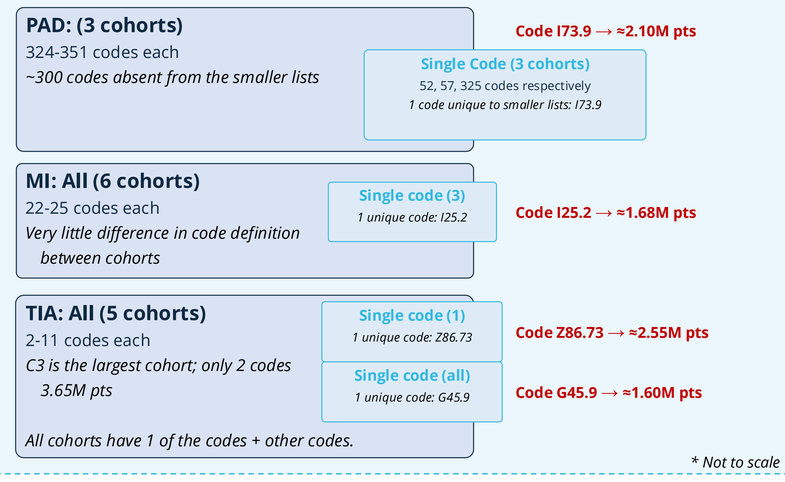
A single ICD-10-CM code can drive cohort inclusion



ICD-10-CM Codes:

- I73.9 - Peripheral vascular disease, unspecified
- I25.2 - Old myocardial infarction
- I67.89 - Other cerebrovascular disease
- Z86.73 - Personal history of transient ischemic attack (TIA), and cerebral infarction without residual deficits
- G45.9 - Transient cerebral ischemic attack, unspecified

Figure 4: Code List & Code Impact Schematic



REFERENCES

1. FDA. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. 2023.
2. Kamauu J, et al. Defining ASCVD in RWR: An Analysis of Conceptual and Operational Definition Variability. ISPOR 2026; **Poster B-269**.
3. Kamauu A, et al. Impact of Variance in Cohort Phenotype Definitions on RWR: ASCVD Assessment. ISPOR 2026; **Poster B-271**.
4. Mackinnon ES, et al. Increasing Prevalence and Incidence of ASCVD in Adult Patients in Ontario, Canada. CJC Open. 2021;4(2):206-213.
5. Dai D, et al. Multimorbidity in ASCVD and Its Associations With Adverse CV Events and Healthcare Costs. J Health Econ Outcomes Res. 2024;11(1):75-85.
6. Sajja A, et al. A simplified approach to identification of risk status in patients with ASCVD. Am J Prev Cardiol. 2021;7:100187.
7. Barrios V, et al. Treatment patterns and use of healthcare resources of patients with ASCVD. Front Cardiovasc Med. 2022;9:966049.
8. Mackinnon ES, et al. Real-World Risk of Recurrent CV Events in ASCVD Patients. Cardiol Ther. 2024;13(1):205-220.

CONCLUSIONS

Key Findings

- Six PAD cohort definitions (52-351 ICD-10-CM codes) identified 5.80–8.78M patient in PurpleLab® CLEAR Claims.
- Three larger lists (324-351 codes) identified only ~1.3× more patients than the two smaller lists (52, 57 codes) despite ~6× difference in code list size.
- A single code (I73.9 “Peripheral vascular disease, unspecified”) absent from three larger lists contributed **≈ 2.10M patients to the smaller cohorts** and one larger cohort.
- Parallel assessments of MI, stroke, and TIE show similar sensitivity.

Clinical Research Implications

- Code list size is not a proxy for phenotype completeness – which codes are included matters more than how many.
- Data-informed phenotype design aligned with study intent is essential for RWE generation.
- Researchers should publish complete code lists and code-level contribution to cohort definition.

Contact: aaron@navidence.com