Poster Code: B-256

A Data-Driven Evaluation of Myocardial Infarction Definitions and the Weight of a Single Code

Scan for poster + more

Scott L DuVall PhD[1] Douglas Londono PhD[1] Allison R Brosso BA[1] Dan Lemberg BS[1] Stephan C Dunning MBA MGIS[1]
Jared Kamauu BFA[2] Aimee Harrison MFA[2] Michael Buck PhD[2] Craig Parker MD MS[2] Aaron WC Kamauu MD MS MPH[2]

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

So What?

Adding one historical diagnosis code, ICD-10-CM I25.2 (Old myocardial infarction), to an MI phenotype ***added 1.5M patients*** *and grew the cohort* ***66.2%****.*

The added patients were demographically similar to the patients already captured by acute MI codes, but introduced differences in geography, age, and race.

Researchers should determine code lists based on use case and desired target population, not added volume alone.

BACKGROUND

Context

Cohort definitions in observational research can vary subtly by the inclusion of historical diagnosis codes. These choices can drastically shift a study population's size and composition, potentially affecting the generalizability of real-world evidence.

Myocardial infarction (MI) is among the most frequently studied outcomes and covariates in pharmacoepidemiology, yet published claims-based MI definitions are not interchangeable.

The Gap / Problem

Two published MI definitions, Mackinnon 2021[1] and Dai 2024[2], differ solely by the inclusion of ICD-10-CM I25.2 "Old myocardial infarction." I25.2 records a history of MI rather than an acute event, so including it changes both how many patients qualify and what clinical state they are in at index.

HEOR / RWE Relevance

Codes that enlarge a cohort are an attractive way to gain statistical power but may come at the cost of a shifted source population. If added patients differ systematically in age, sex, race, ethnicity, or geography, the definition choice becomes a source of confounding and a risk to comparability across studies.

See companion posters ***B-269, B-271, & B257***

OBJECTIVES

Primary Objective

To evaluate the impact of including "Old Myocardial Infarction" (ICD-10-CM I25.2) in an MI phenotype, to see how a single code influences cohort volume and demographic characteristics.

Secondary Objectives

- Quantify the change in cohort size attributable to I25.2 alone
- Characterize the incremental population contributed by I25.2 and compare it directly against the population identified by acute MI codes alone.
- Compare the cohorts on age, sex, race, ethnicity, and state of residence.
- Determine whether the gain in sample size inadvertently introduces demographic or geographic bias that could confound future comparative effectiveness research.

LIMITATIONS

- This cross-sectional design compared cohorts on baseline characteristics only and were not linked to downstream outcomes, so this analysis quantifies *who* is added rather than *how* estimates change.
- No reference standard for true MI status exists in claims, so this study cannot establish which definition is more valid, only how they differ in the population they select.
- Race and ethnicity are incompletely captured so comparisons are conditional on recorded values.
- Only two published definitions were compared, isolating I25.2. Other MI phenotypes differ on care settings, diagnosis position, and washout, which were not varied here.

REFERENCES

1. Mackinnon ES, et al. Increasing Prevalence and Incidence of ASCVD in Adult Patients in Ontario, Canada. CJC Open. 2021;4(2):206-213.
2. 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.

METHODS

Study Design & Data Source

Cross-sectional comparative analysis using PurpleLab® CLEAR Claims, a large, representative U.S. all-payer claims database spanning commercial, Medicare, and Medicaid populations.

Because both definitions are applied to the same source data, every difference observed between the cohorts is attributable to the code list itself.

Study Population

Two published MI definitions were applied. They share an identical acute MI code list and differ solely by whether ICD-10-CM I25.2 "Old myocardial infarction" is included.

Because the Mackinnon 2021 cohort is fully nested within the Dai 2024 cohort, we isolated the incremental population contributed by I25.2 and compared it directly against the acute code only cohort. SMD compares the Mackinnon cohort with the incremental cohort only.

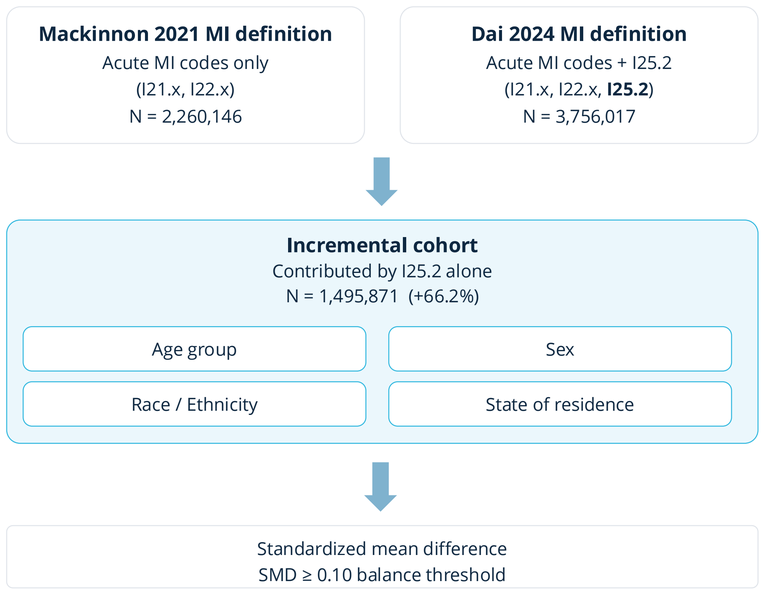

Figure 1. Study design. Both published definitions were applied to the same source population. Because they differ by exactly one code, the incremental cohort is a clean way to examine the effect of I25.2.

Analysis

Outcomes were defined as total cohort size and the distribution of age group, sex, race, ethnicity, and state of residence.

Distributions were compared using standardized mean difference. The conventional 0.10 balance threshold was used for similarity.

CONCLUSIONS

Key Findings

- The single diagnosis code I25.2 "Old myocardial infarction" added 1,495,871 patients to a base of 2,260,146, a 66.2% relative increase.
- The added patients were older (65+ 63.8% vs 59.5%) with modestly lower proportions female (40.9% vs 43.1%), African-American (10.1% vs 11.8%), and Hispanic (10.3% vs 11.5%).
- Geography showed the largest effect and the widest spread: 37% additional volume in Vermont, 122% in Maryland, a pattern more consistent with regional coding practice than with MI burden.
- Researchers should incorporate data-driven methods when defining study variables, weighing not only cohort volume, but also how each definition shifts the characteristics of the desired target population.

TRANSPARENCY & REPRODUCIBILITY

Both MI code lists were taken verbatim from their published sources listed under References, with no modification, so that the only difference between the cohorts is the presence of ICD-10-CM I25.2. Identical index-date logic, look-back, and eligibility requirements were applied to both definitions.

RESULTS

1,495,871
Patients Added By ICD-10-CM I25.2

+66.2%
Increase In Total Cohort Size

0.043 to 0.166
SMD Range

Characteristic	Mackinnon (acute only)	Added by I25.2	Dai 2024 (combined)	SMD
Patients, n	2,260,146	1,495,871	3,756,017	
Age group % age 65+	59.5	63.8	61.2	0.120
Sex % female	43.1	40.9	42.2	0.043
Race % African-American	11.8	10.1	11.1	0.082
Ethnicity % Hispanic	11.5	10.3	11.0	0.044
State % largest state	CA 11.9	CA 9.9	CA 11.1	0.166

Table 1. Cohort characteristics under the two published MI definitions and the incremental population contributed by I25.2. SMD is the multinomial form. The percentage shown is one level of that distribution.

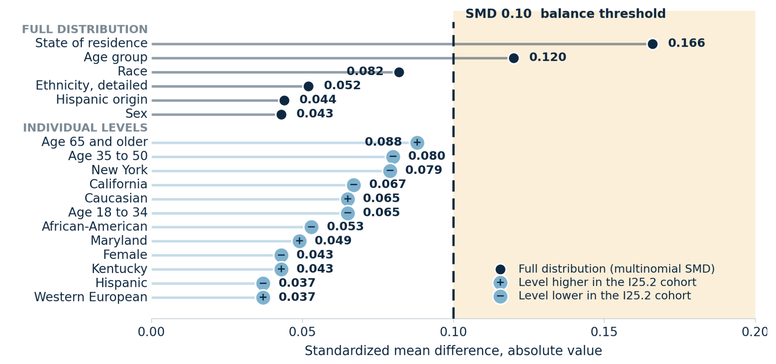

Figure 2. Standardized mean differences, acute code only cohort vs I25.2 cohort. Binary SMD for the 12 largest individual level shifts and multinomial SMD over the full distribution.

Figure 3. Patients added by I25.2 in each state as a percentage of that state's acute code only cohort.

DISCLOSURES & ACKNOWLEDGMENTS

DISCLOSURES: Scott L. DuVall, Douglas Londono, Allison R. Brosso, Dan Lemberg, and Stephan C. Dunning are employees of PurpleLab. Jared Kamauu, Michael Buck, Craig Parker, and Aaron Kamauu are employees of Navidence. Aimee Harrison was an employee of Navidence during this work.

Contact: aaron@navidence.com