Prevalence of Key Confounding Covariates in a Commercial Electronic Health Records Platform

Youjin Wang[1], Ashley I. Michnick[2], Jummai Apata[1], Nathan Kim[3], Maria Lewis[3], Wei Liu[1], Nora P. McElroy[3], Emma R. Hoffman[3]

[1]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, United States Food and Drug Administration (FDA), Silver Spring, MD
[2]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute (HPHCI), Boston, MA
[3]Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA

Presented at the International Society for Pharmacoepidemiology 41st Annual Meeting (Washington D.C., USA; 22-26 August 2025)

BACKGROUND

- Administrative claims data often insufficiently captures important clinical covariates that may confound the relationship between medical product exposures and health outcomes of interest
- Electronic health records (EHR) data may have better capture of such confounders
- The availability of large, commercially-available EHR platforms may provide an opportunity to address the limitations of confounding adjustment in medical product safety evaluations

OBJECTIVE

- To describe the annual prevalence of key confounding covariates within a large commercial EHR platform to determine its utility for medical product safety surveillance

METHODS

- **Population**: Individuals of all ages with evidence of a visit in the measurement year and in the year prior from 2017 – 2023
  - Identified in the TriNetX Live™ platform in the USA Network EHR database among health care organizations that provide data without shifting dates
- **Covariates**: Captured **body mass index (BMI)** and **diastolic/systolic blood pressure (DBP, SBP)** from clinical measurements; **obesity** and **smoking** from diagnosis, procedure, and medication code-based records
- **Statistical Analysis**: Measured annual prevalence as the proportion of all eligible individuals with evidence of the covariate of interest in the measurement year
  - No formal statistical comparisons were made since this analysis was descriptive
  - All counts are rounded up to the nearest 10 to protect patient privacy

RESULTS

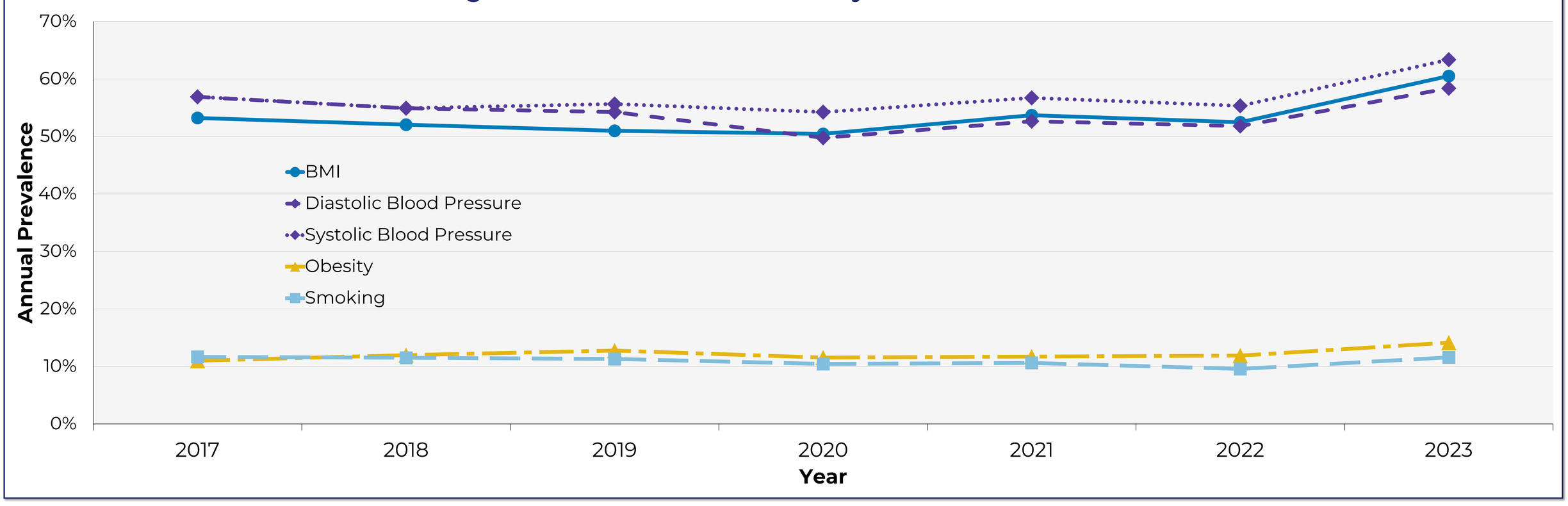
- Over the study period, the number of eligible individuals showed a general upward trend, peaking in 2022 at 17,240,150 before declining in 2023

Table 1. Number of Eligible Individuals, by Year

Year	Eligible Individuals
2017	11,046,660
2018	12,286,910
2019	13,230,670
2020	13,742,460
2021	15,727,080
2022	17,240,150
2023	15,163,230

- Clinical measurements' prevalence increased in 2023
  - **BMI**: 50.4 – 53.7% (2017-2022) → 60.5% (2023)
  - **SBP**: 54.2 – 56.9% (2017-2022) → 63.4% (2023)
  - **DBP**: 49.8%-56.9% (2017-2022) → 58.4% (2023)
- Prevalence of code-based covariates remained generally constant
  - **Obesity**: 11.0 – 14.1%
  - **Smoking**: 9.6 – 11.7%
- There were slight discrepancies in the prevalence of DBP vs. SBP starting in 2019

Figure 1. Annual Prevalence of Key Clinical Confounders



CONCLUSIONS

- The results showed lower than expected capture of certain code-based covariates. Obesity prevalence in 2023 was 34% in the CDC Behavioral Risk Factor Surveillance System (BRFSS) compared to 14.1% in our data, while smoking showed similar prevalence (12% *vs.* 11.6%, respectively)
- The inclusion of non-clinical visit types (e.g., virtual) may have biased prevalence estimates downward
- Differences in how clinical measurements are recorded may have led to variability in data completeness within the TriNetX common data model
- Additional evaluation of other data sources (including mechanisms of missing data) are warranted
- Data source characteristics should be carefully considered when planning any medical product safety evaluation

ACKNOWLEDGEMENTS/DISCLOSURES

- This project was supported by Task Order 75F40123F19009 under Master Agreement 75F40119D10037 from the U.S. FDA. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the U.S. Government.
- A.I.M., N.K., M.L., N.P.M., E.R.H. are employees of HPHCI, an organization which conducts work for government and private organizations, including pharmaceutical companies.
- Many thanks are due to TriNetX Live™, which provided the data used in this analysis.