Identification of Obesity and Tobacco Use in Claims Compared to Electronic Health Record (EHR) Data

Casie Horgan[1], Po-Yin Chang[2], Meg Her[1], Rebecca Hawrusik[1], Alexander Peters[1], Catherine Corey[2], David J. Graham[2], Judith C. Maro[1] and Noelle M. Cocoros[1]

[1] Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA
[2] U.S. Food and Drug Administration, Silver Spring, MD, USA

OBJECTIVES

- To examine claims and EHR-based indicators of obesity and tobacco use among sodium-glucose co-transporter-2 inhibitor (SGLT-2i) users in the FDA Sentinel System

- To understand how key characteristics of interest differ in EHR versus claims data

BACKGROUND

- Obesity and tobacco use are important covariates for many drug safety questions

- Claims data are expected to undercapture both obesity and tobacco use, leading to misclassification, while EHR data are expected to better identify these characteristics

- The Sentinel Common Data Model includes Vital Signs Clinical Data tables populated via EHR sources. Seven Integrated Delivery Systems populate these data, along with codes from administrative claims diagnoses and procedures

Figure 1. Sentinel Common Data Model's Clinical Data Tables

METHODS

- We identified new SGLT-2i users aged ≥20 years, March 2013 through June 2018, and assessed obesity and tobacco use in the 365 days prior to and including index dispensing

- EHR-based body mass index (BMI) was calculated using closest valid height and weight records to index and classified as:

Underweight/Normal: 15 ≤ BMI < 25

Overweight: 25 ≤ BMI < 30

Obese: 30 ≤ BMI < 40

Severely Obese: 40 ≤ BMI ≤ 90

- Claims-based BMI and obesity was defined using the closest diagnosis or procedure code to index. Separate "narrow" and "broad" claims-based obesity definitions were defined as:

Narrow Obesity: presence of a BMI-specific code for BMI ≥ 30

Broad Obesity: presence of a BMI-specific or general obesity code

- EHR-based tobacco use was defined by closest populated vital measure to index date using the following categories:

Current User

Never User

Former User

Other (includes passive exposure, environmental exposure, conflicting information, and record indicating the patient was not asked)

- Claims-based tobacco use was defined by a diagnosis code for tobacco use or a dispensing for a tobacco cessation product

RESULTS

- We identified 3,155 new SGLT-2i users

- 61% had an EHR-based BMI recorded

- 43% had any claims code for obesity (e.g. ICD-10 code E66.8 for "Other obesity") or BMI (e.g. ICD-10 code Z68.26 for "Body Mass Index (BMI) 26.0-26.9, adult")

- Only 14% of all EHR BMI measures had a *concordant* claims-based code when using the narrow definition of obesity; 44% had a *concordant* claims-based code when using the broad definition of obesity

- 70% had any EHR record for tobacco use; 44% of those records indicated current or former use

- 40% of EHR-based current and former tobacco users had a *concordant* claims code indicating tobacco use

Table 1. EHR Versus Claims-Based BMI/Obesity Data Availability in New SGLT-2i Users

		Presence of a Claims Code for Any BMI or General Obesity Code	
EHR Record for BMI		Yes	No
	Yes	932 (29.5%)	1,004 (31.8%)
	No	420 (13.3%)	799 (25.3%)

Table 2. EHR Versus Claims-Based Tobacco Use Data Availability in New SGLT-2i Users

		Presence of a Claims Code for Tobacco Use	
EHR Record for Tobacco Status		Yes	No
	Yes	419 (13.3%)	1,786 (56.6%)
	No	143 (4.5%)	807 (25.6%)

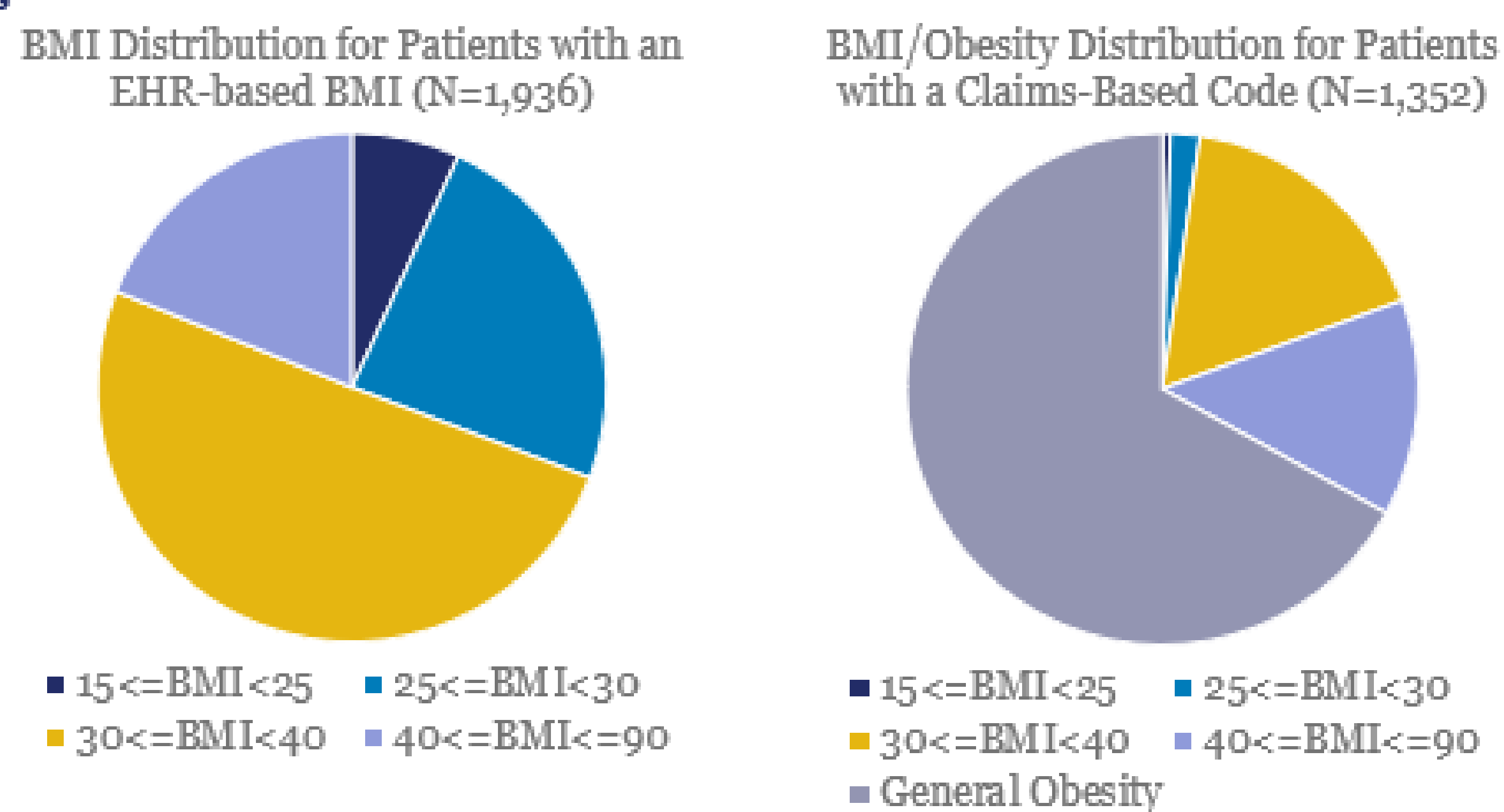**Figure 2. BMI and Obesity Estimates in Claims Versus EHR Data in New SGLT-2i Users**

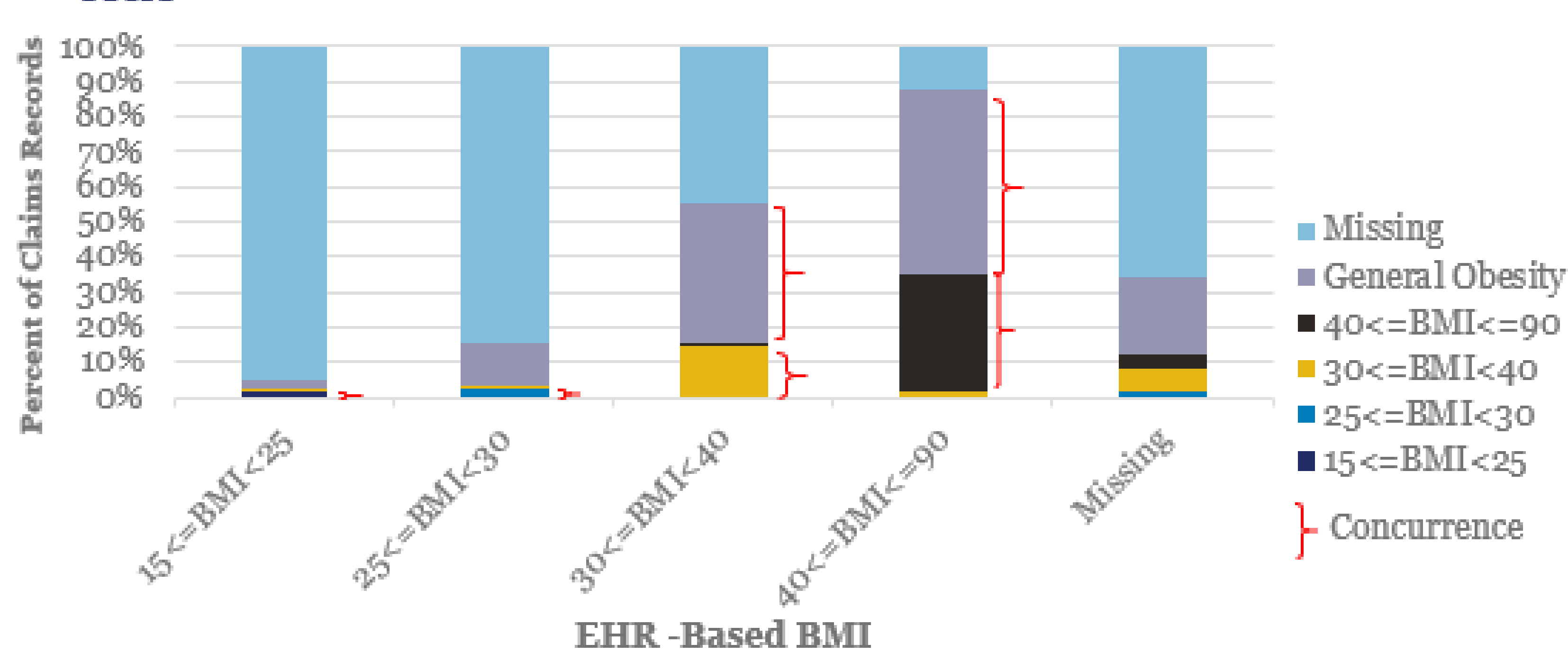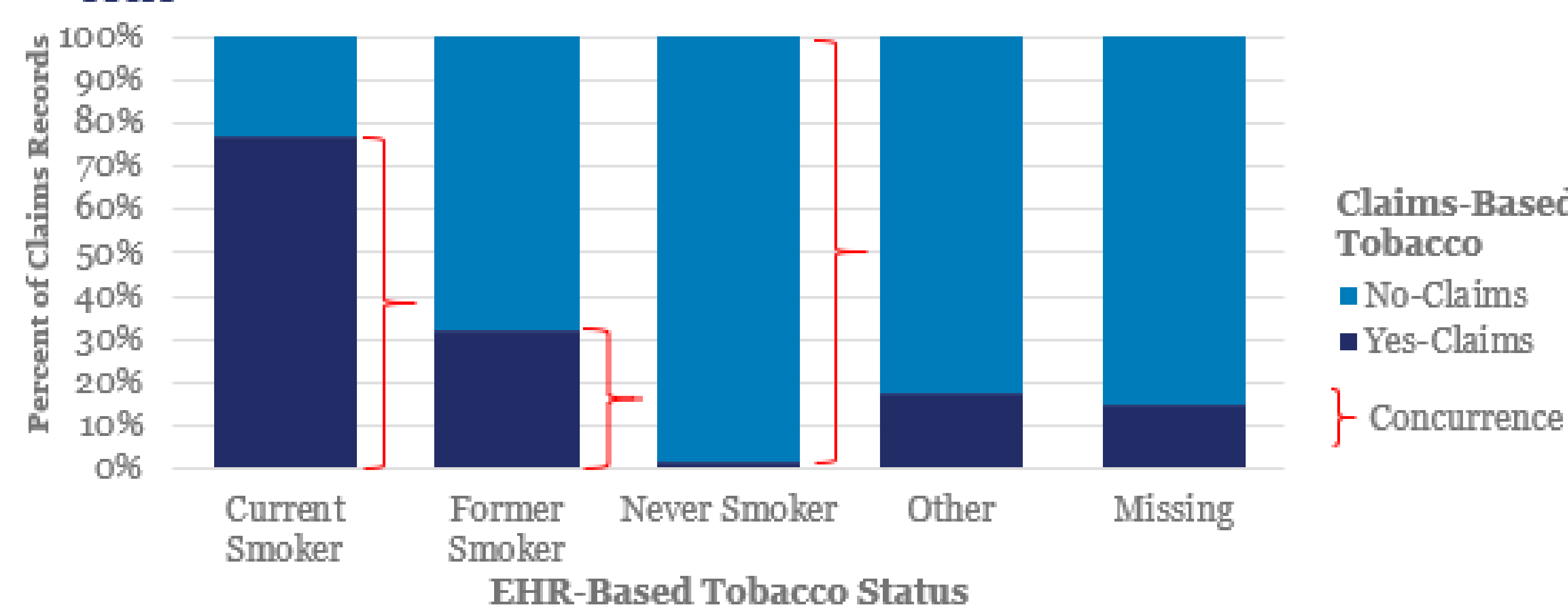**Figure 3. BMI and Obesity Concurrence across EHR and Claims Data in New SGLT-2i Users**

Figure 4. Tobacco Use Concurrence across EHR and Claims Data in New SGLT-2i Users

CONCLUSION

- SGLT-2i new users showed good concordance between claims and EHR for current tobacco use

- A more sensitive definition of obesity in claims data increased concordance with EHR data compared to a narrower definition that was BMI-specific

- Characterization of missingness and clinical measurement concordance for key confounders could inform quantitative bias analysis or imputation approaches for future work

ACKNOWLEDGEMENTS/DISCLOSURES