

Using Linked Data: Agreement in Diagnosis for Patients with Asthma or Diabetes

Halley Costantino, Robert J Taylor, Bridget L Balkaran, and Dena H Jaffe

Background

- Clinical phenotyping is used to identify undiagnosed patients and improve the sensitivity of diagnoses for real-world evidence (RWE).^{1,2}
- Diagnostic codes from electronic medical records (EHRs) and claims data are a primary source of disease phenotyping and are known to differ between sources.³
- The value of data linkage for observational studies is that each additional data source provides a more complete view of the patient’s health and care.

Objective

To evaluate the agreement between identification of patients with asthma or with diabetes using diagnosis codes from EHRs and claims data.

Methods

Study design: Retrospective cohort

Data Sources

Two data sources containing Datavant tokens were utilized:

- EHR: Data from the Oracle EHR Real-World Data (ORWD) released through January 2023 (105.2M patients).
- Claims: Data from a national US closed claims data set released through April 2023 (233.2M patients).

Study population

The study population was created in two stages:

Stage 1:

Patients fulfilling all of the following criteria:

- Patients in the EHR-claims linkable dataset
- Patients with 1+ EHR encounter from 01/07/2021 to 30/06/2022
- Patients with 1+ claim from 01/07/2021 to 30/06/2022
- Patients with continuous enrolment in claims data during the study period 01/07/2021 to 30/06/2022

Exclusion criteria: None

Stage 2:

Patients were identified for the following groups:

1. Asthma

Inclusion criteria:

- EHR: Patients with 2+ diagnoses of asthma (primarily ICD-9 493.x; ICD-10 J45.x; and other associated codes such as SNOMED, MEDCIN, and NCI) with a minimum 30 days from first (index) diagnosis
- Claims: Patients with 2+ diagnoses of asthma (primarily ICD-9 493.x; ICD-10 J45.x) with a minimum 30 days from first (index) diagnosis
- Patients between ages 2 to 89 years at time of first diagnosis (age reported from EHR)

Exclusion criteria: None

2. Diabetes

Inclusion criteria:

- EHR: Patients with 2+ diagnoses of type 2 diabetes (ICD-9 250.x0, 250.x2; ICD-10 E11.x, O24.1x; and other associated codes such as SNOMED, MEDCIN, and NCI) with a minimum 30 days from first (index) diagnosis
- Claims: Patients with 2+ diagnoses of type 2 diabetes (primarily ICD-9 493.x; ICD-10 J45.x) with a minimum 30 days from first (index) diagnosis
- Patients between ages 18 to 89 years at time of first diagnosis (age reported from EHR)

Exclusion criteria: None

Measures

Demographic variables were reported using EHR data for the following:

- Age at first diagnosis
- Sex
- Race
- Ethnicity
- Marital status
- Region of residence

Analysis

Descriptive analyses were conducted. Frequencies and percentages were used for categorical variables and means and standard deviations (SD) were used for continuous variables.

Results

Patient study populations were identified as those with diagnoses in both EHR and claims data (Tables 1 and 2).

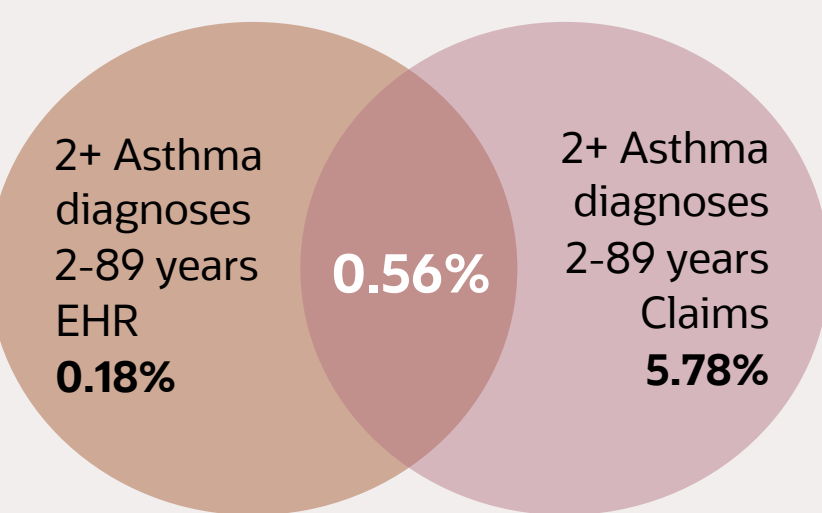
Asthma	EHR	Claims	Type 2 Diabetes	EHR	Claims
Total patients	105.2 million	233.2 million	Total patients	105.2 million	233.2 million
Linkable patients	35.4 million		Linkable patients	35.4 million	
Patients with 1+ claim and 1+ encounter during study period (1/7/2021 to 30/6/2022)	15,178,092		Patients with 1+ claim and 1+ encounter during study period (1/7/2021 to 30/6/2022)	15,178,092	
Patients with 2+ diagnoses for asthma in claims or EHR, aged 2 to 89 years	111,694	962,450	Patients with 2+ diagnoses for type 2 diabetes in claims or EHR, aged 18 to 89 years	253,669	1,699,360
Patients with 2+ diagnoses for asthma in both claims and EHR, aged 2 to 89 years	84,407		Patients with 2+ diagnoses for type 2 diabetes in both claims and EHR, aged 18 to 89 years	201,189	

Table 1: Attrition Table of Patients Aged 2 to 89 Years With 2+ Diagnoses for Asthma in EHR Data and 2+ Diagnoses for Asthma in Claims Data

Table 2: Attrition Table of Patients Aged 18 to 89 Years With 2+ Diagnoses for Type 2 Diabetes in EHR Data and 2+ Diagnoses for Type 2 Diabetes in Claims Data

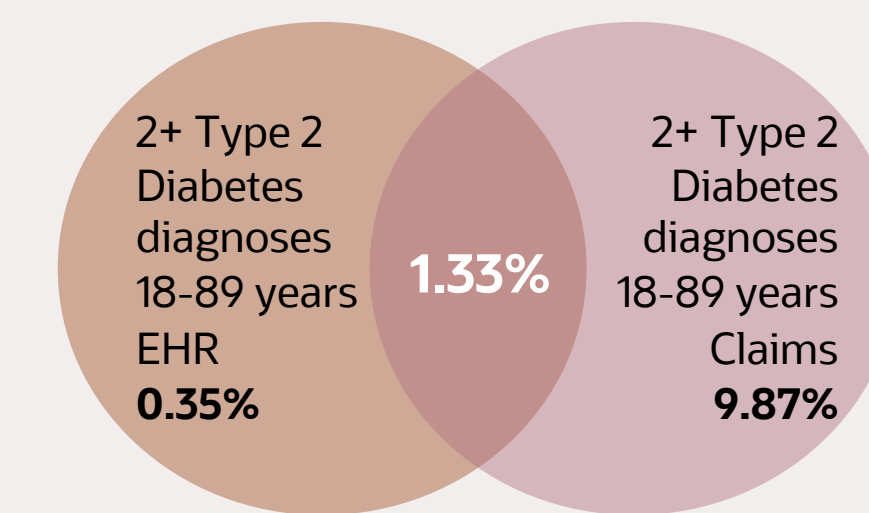
Diagnostic agreement between data sources was low for patients with asthma or diabetes (Figs 1 and 2).

Figure 1: Patients with 2+ Diagnoses of Asthma Aged 2 to 89 Years in Both EHR and Claims Data Among Linked EHR-Claims Patients (n=15,178,092)^a



^a Linked EHR-claims patients were those with at least one EHR encounter and one claim during the study period 1/7/2021 to 30/6/2022.

Figure 2: Patients with 2+ Diagnoses of Type 2 Diabetes Aged 18 to 89 Years in Both EHR and Claims Data Among Linked EHR-Claims Patients (n=15,178,092)^a



84,407 patients met the diagnostic and age criteria for asthma in both EHR and claims. This cohort represented:

- 0.56% of the total linked cohort,
- 75.57% of the patients with asthma identified in the EHR (84,407/111,694), and
- 8.77% of the patients with asthma identified in claims (84,407/962,450).

201,189 patients met the diagnostic and age criteria for type 2 diabetes in both EHR and claims. This cohort represented:

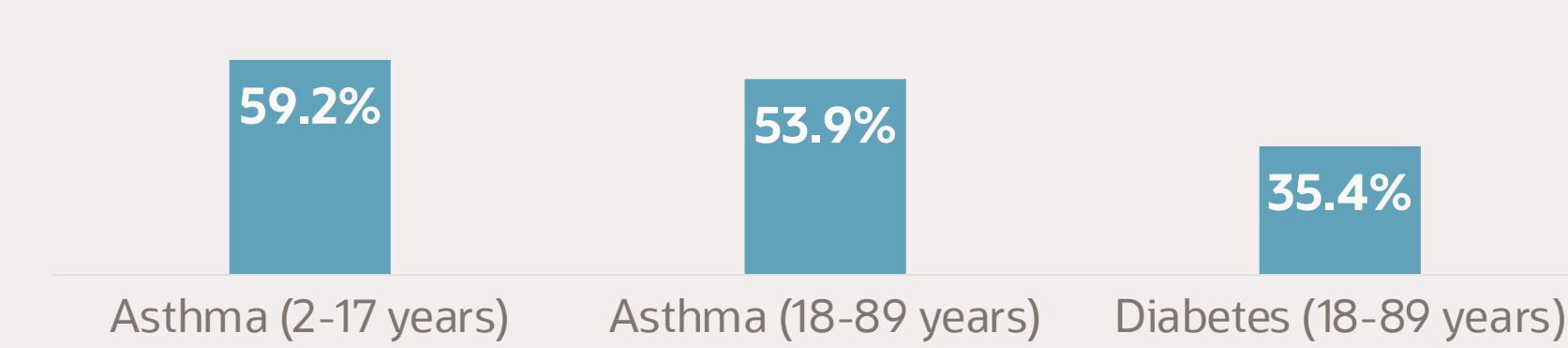
- 1.33% of the total linked cohort,
- 79.31% of the patients with type 2 diabetes identified in the EHR (201,189/253,669), and
- 11.90% of the patients with type 2 diabetes identified in claims (201,189/1,690,360).

Agreement with Diagnosis Dates

Agreement between diagnosis dates for each was assessed for index date (i.e., date first identified in the study period) and for any of the 2 diagnosis dates used to identify each cohort.

- Among patients identified in both data sources, agreement of either of the two diagnosis date was 74.7% (asthma, overall) and 51.1% (diabetes).
- Index date agreement (Fig 3):
 - Among patients with asthma, regardless of age group, the index diagnosis date identified in EHR and claims aligned for more than half of the patients.
 - Among patients with type 2 diabetes, agreement of the index diagnosis dates between the EHR and claims data was 35.4%.

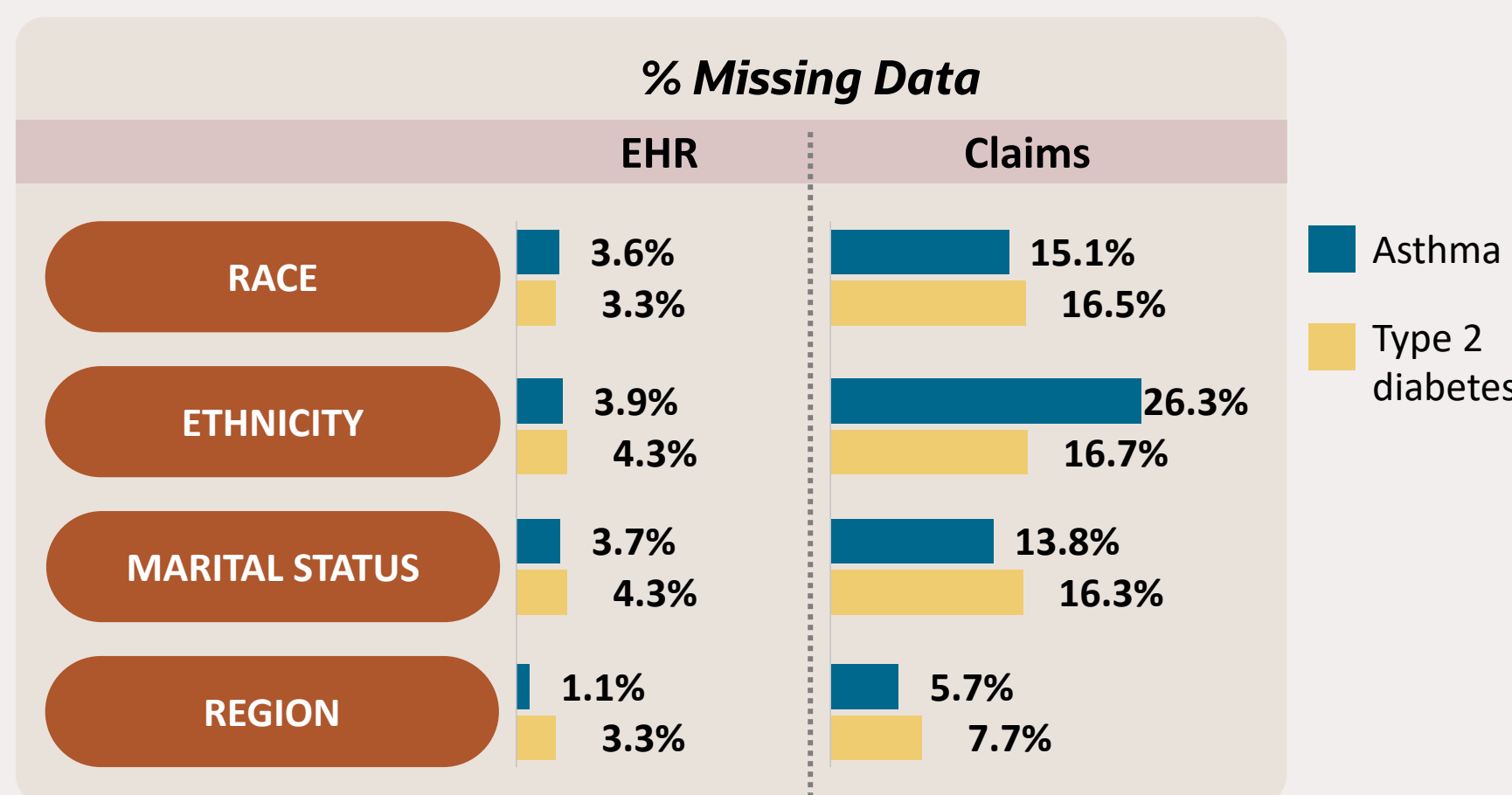
Figure 3: Percent Agreement of Index Diagnosis Date Between EHR and Claims



Patient Characteristics by Data Source

- The age distribution of adult patients was similar between the EHR and claims, whereas a higher proportion of children were present in the EHR than claims datasets.
 - For the asthma cohort, the average age in years was 33.12 (SD=22.22), 35.26 (SD=21.15), and 31.12 (SD=21.51) for patients identified from EHR, claims, or both data sources, respectively.
 - 35.3% compared with 28.0% of patients with asthma identified in the EHR and claims, respectively, were children 2-17 years.
 - Among adults 18+ years, there were slightly more patients with asthma aged 60+ years identified from the EHR (23.9%) compared to the claims' (21.0%) dataset.
 - For the type 2 diabetes cohort, the average age in years was 60.35 (SD=13.77), 59.34 (SD=14.00), and 59.90 (SD=13.76) for patients identified from EHR, claims, or both data sources, respectively.
 - In this cohort of adults 18+ years, there were slightly more patients with type 2 diabetes aged 60+ years identified from the EHR (53.8%) compared to the claims' (50.3%) dataset.
- Women represented approximately 61-63% of patients with asthma and 54% of patients with type 2 diabetes, across data sources.
- The percentage of patients with asthma who were Black/African American was greater in the EHR (27.5%) compared to the claims' dataset (20.6%), while for patients with type 2 diabetes, more moderate differences were observed (15.1% versus 13.7%, respectively).
- The percentage of patients who were Hispanic or Latino were similar across data sources for both asthma and type 2 diabetes cohorts (22-23% and 22-25%, respectively).
- Approximately two-thirds of patients with asthma (EHR=64.9%, claims=60.5%, both=67.8%) and less than one-third for those with type 2 diabetes (EHR=29.7%, claims =30.5%, both=30.5%) were single or never married.
- The geographic distribution of patients remained stable across each data source for patients with asthma and type 2 diabetes. A higher percentage of patients with asthma were identified in the South (29-34%), whereas more patients with type 2 diabetes were from the West (33-34%).
- Differences in demographic characteristics of patients identified from EHR and claims datasets were primarily regarding missing data, with percent missingness between 4-7 times higher in claims compared to EHR (Fig 4).

Figure 4: Percentage of Patients with Missing Demographic Data Among Patients with Asthma or Type 2 Diabetes Identified Using Diagnosis Records in EHR or Claims Data



Strengths and Limitations

- The ORWD is a large, national EHR-agnostic database that aggregates patient data from all EHR providers within a healthcare system. Linkage with a national claims dataset provides researchers with opportunities for more comprehensive and valid data.
- This study examined patients during a 1-year period, however depending on the study objective, longitudinal studies might consider patient eligibility in both data sets for a minimum period prior to and/or post index.
- The purpose of this study was to examine the variability of diagnoses found in these two data sources, although we note that most studies apply more rigorous methodology for disease phenotyping.

Conclusions

- Using patient diagnoses for disease phenotyping, we found over 6x's more patients with asthma or type 2 diabetes identified in claims than in EHR linked datasets. Notably, most patients identified using the EHR had diagnostic agreement with the claims dataset compared to the low agreement of those identified using claims who were also found in the EHR.
- Cohorts derived from linked data may be smaller than anticipated although represent more accurate patient crossover.
- Data linkage maximizes available information, however, requires careful examination of attrition and characteristics by data source independently and jointly.

References

- Pendergrass SA, Crawford DC. Using Electronic Health Records to Generate Phenotypes for Research. *Curr Protoc Hum Genet*. 2019;100(1).
- Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *Journal of the American Medical Informatics Association*. 2014;21(2):221-230.
- Lee S, Doktorchik C, Martin EA, et al. Electronic Medical Record–Based Case Phenotyping for the Charlson Conditions: Scoping Review. *JMIR Med Inform*. 2021;9(2):e23934.