Replicating Published Real-World Evidence Studies Using Electronic Medical Records From Community Hospitals

RWD131

Borda H,[1] Buchanan-Hughes A[1]

[1]Costello Medical, Boston, MA, US

Objective

To explore the strengths and limitations of the insightsDB electronic medical record database by replicating aspects of published studies conducted in other US healthcare databases.

Background

- Commercially-available real-world data (RWD) in the US are often sourced from employer-sponsored insurance plans or major provider networks, leading to underrepresentation of patients with lower socioeconomic status in real-world evidence studies.
- The insightsDB database, from the Institute for Health Metrics, collates electronic medical record (EMR) data from 98 community hospitals (88% governmental or voluntary non-profit) across 30 US states; over 5 million unique patients and 14 million encounters are captured, including linkage to social determinants of health (SDOH) from census tract data.[1]
- Further characterization of the database is needed to understand its relative strengths and limitations and to identify use cases.

Methods

- Characteristics of the insightsDB population from January 2017–June 2023 were described at entry into the database and compared to those of the total US population based on national census data.[2,3]
- Three case studies replicating published research were conducted to represent a broad range of disease areas and outcomes; selected studies were required to have used RWD from other US electronic healthcare databases and provided sufficient information for adequate replication of the methodology in insightsDB.
- Two published studies selected for replication used data from typical RWD databases: Optum Clinformatics and the Healthcare Cost and Utilization Project (HCUP), and one used a large regional EMR dataset.[4–6]

Results

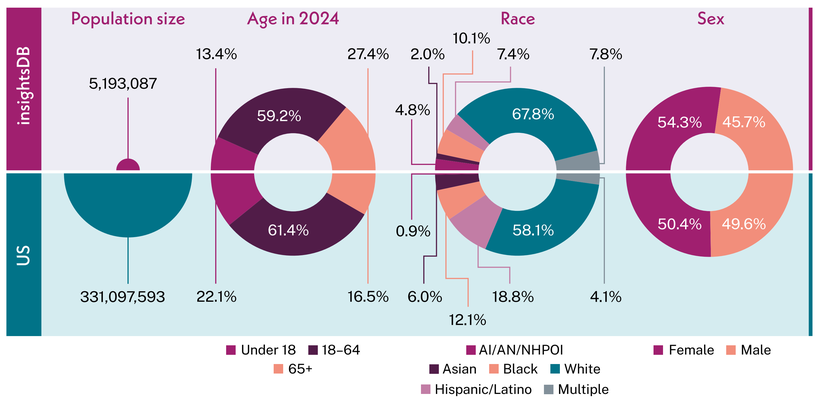
- Demographics of the insightsDB population were broadly representative of the general US population, including patient subgroups likely to be underrepresented in other commercially-available datasets, such as those aged ≥65 years, minority ethnic groups, or those living in areas of greater deprivation (**Figure 1**).
  - Over two-thirds of the insightsDB population live in a community with a household income below the national average.
- Key strengths and limitations of insightsDB identified in the replication of the respective study methods are outlined in **Table 1**.
  - In all case studies, replication of the analyses in insightsDB was successfully conducted; however, missing data in certain variables limited sample size of the study populations.
- A comparative summary of results against the original studies is also provided in **Table 1**.
  - **Case study 1**: Low use of iGlarLixi (insulin glargine 100 U/mL plus lixisenatide) was observed in the insightsDB population, suggesting that access to novel, expensive treatments is limited in patients treated in community hospital settings compared with populations captured in commercial insurance claims datasets.
  - **Case study 2**: Greater association of pre-existing chronic conditions or preterm birth with complications following delivery was reported in insightsDB than the replicated study, potentially due to challenges with managing these scenarios in less well-resourced hospital settings.
  - **Case study 3**: Prevalence of an inpatient admission within 30 days of a mastectomy procedure was twice as high in patients who received the procedure in an inpatient setting compared to those who underwent mastectomy in the outpatient setting. In patients having received an inpatient mastectomy, readmission rate at 30 days was similar to the original study.

Conclusion

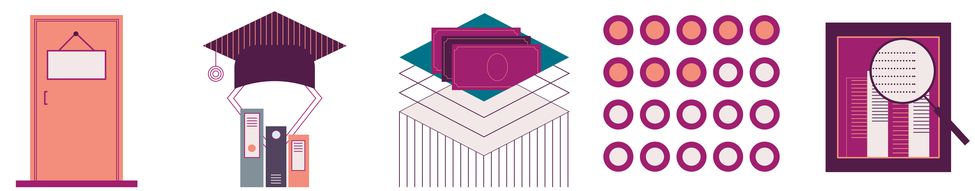
- Results across case studies highlighted differences in treatment access and outcomes for patients in community hospitals compared with the replicated studies, suggesting that insightsDB fills an important gap in the broad landscape of available RWD sources.
- As with any RWD source, analyses using data from community hospitals such as insightsDB should be carefully designed to maximize strengths and overcome limitations, such as limited data on tests and procedures carried out at external sites.
- Particular strengths of the database include capture of typically underrepresented groups and availability of linkage to census tract level SDOH data, which can be used to explore important health equity topics.

FIGURE 1

Overview of insightsDB compared to the 2022 US Population[2,3]



Proportion of individuals in insightsDB living in an area with greater deprivation than the US national estimate:[a]



48.8% live in an area with **above average severe housing issues**

46.2% live in an area with **below average academic attainment** (high school or higher)

67.3% live in an area with **below average household income**

39.4% live in an area with **above average SNAP benefit usage**

48.5% live in an area with **above average level of unemployment**

[a]US national estimates: Severe housing issues (at least one of the following problems: lack of complete kitchen facilities, lack of plumbing facilities, overcrowding or severely cost-burdened occupants): 16.7% (2016–2020);[3] High school education or higher, in population aged ≥18 years: 89.0% (2018–2022);[2] Median household income: $75,149 (2018–2022);[2] Receipt of food stamps/SNAP: 11.5% (2018–2022);[2] Unemployment in the last 12 months, in civilian labor force: 5.3% (2018–2022).[2]

TABLE 1

Case studies replicating published studies using insightsDB

	Type 2 Diabetes (T2D)	Maternal delivery complications	Inpatient admissions after mastectomy procedures
Study Overview	Pantalone et al. (2023) assessed treatment persistence after 12 months on a novel combination therapy drug for T2D, iGlarLixi, compared to basal-bolus insulin therapy. T2D patients were switched to one of these treatments when basal insulin alone was suboptimal.[4]	Huennekens et al. (2020) evaluated the association of demographic traits, comorbidities, and delivery characteristics with maternal delivery complications using multivariable logistic regression.[5]	Yu et al. (2020) analyzed the proportion of 30-day hospital readmission after uni-or bi-lateral mastectomy for breast cancer and explored both patient and operative factors associated with readmission.[6]
Date Source	Optum Clinformatics database.	A large Illinois health system EHR depository with data from 7 hospitals, including a large academic medical center and a referral center for women with complicated pregnancies.	2011 HCUP California State Inpatient Database.
Key Methodological Aspects	**Strengths** ♦ Captures use of medications prescribed outside the hospital setting and obtained over-the-counter that may not be visible in claims data. **Limitations** ♦ Of those initiating basal-bolus insulin, only 2.6% of patients compared to 35.7% in Pantalone et al. (2023) had a valid HbA1c reading, limiting sample size. ♦ Limited data for variables such as diagnoses or lab tests that were performed outside the hospital setting, e.g. in primary care settings.	**Strengths** ♦ InsightsDB contained 136,275 mothers who delivered children during the 6-year study period. ♦ Replication of methodology to consider the association of demographic traits, comorbidities, and delivery characteristics with delivery complications using multivariable logistic regression was feasible. **Limitations** ♦ Due to missing data on maternal characteristics, 50,358 individuals were included in the study; however, unstructured notes, which were not included in this analysis, may have been able to provide further data.	**Strengths** ♦ Records of mastectomy procedures in insightsDB captured patients undergoing the procedure both in the inpatient (n=1,876, 34.4%) and outpatient (n=3,572, 65.6%) setting. **Limitations** ♦ Patients who underwent the procedure in specialized oncology treatment centers would not be included in the community hospital records and may contribute to a smaller sample size.
Key Results	♦ In the original study, **~1:34** ratio of patients were taking a novel combination therapy vs standard basal-bolus insulin regimen compared to **~1:755** in insightsDB.	♦ Hypertensive disorders of pregnancy, placental disorders, and cesarean route of delivery were significantly (p<0.05) associated with higher odds of having a delivery complication in both studies. ♦ Levels of maternal BMI and race were significantly associated with delivery complications in Huennekens et al. (2020) but not in insightsDB. ♦ Pre-existing chronic conditions (OR 1.32; 95% CI 1.15–1.51) and preterm deliveries (OR 1.43; 95% CI 1.20–1.70) were associated with complications in insightsDB but not in Huennekens et al. (2020).	♦ Inpatient admission occurred within 30 days of an inpatient mastectomy procedure in **5.1%** of cases in insightsDB, similar to the proportion (**4.9%**) in the original publication. In patients undergoing an outpatient mastectomy in insightsDB, admissions within 30 days of the procedure occurred in **2.5%** of cases. ♦ Type of insurance and race significantly differed between the post operative admitted and non-admitted groups in the original study. In insightsDB, there was a higher frequency of patients utilizing Medicaid or Medicare in the group that was admitted within 30 days after their procedure. The percentage of Hispanic/Latino, AI/AN, Black, and White individuals was also higher in the admitted group compared to the non-admitted group.

Abbreviations: AI/AN: American Indian/Alaskan Native; BMI: body mass index; CI: confidence interval; EMR: electronic medical record; HCUP: Healthcare Cost and Utilization Project; NHPOI: Native Hawaiian/Other Pacific Islander; RWD: real-world data; SDOH: social determinants of health; SNAP: Supplemental Nutrition Assistance Program; OR: odds ratio; T2D: type 2 diabetes.

References: [1]insightsDB (2017). Available at: https://www.healthmetrics.org/insightsdb [Last accessed: 19 April 2024]; [2]US Census Bureau (2024). Available at : https://data.census.gov/ [Last accessed: 19 April 2024]; [3]Comprehensive Housing Affordability Strategy (CHAS) Data (2024). Available at: https://www.huduser.gov/portal/datasets/cp.html [Last accessed: 19 April 2024]; [4]Pantalone et al. Diabetes Spectr 2023;36(3):253–263; [5]Huennekens et al. Jt Comm J Qual Patient Saf 2020;46(11):623–630; [6]Yu et al. Breast J 2020;26(10):1966–1972. **Acknowledgements:** The authors thank Melanie Howes, Costello Medical, for graphic design assistance. We also thank Anita Karcz, Institute of Health Metrics; Anna Zolotor, Costello Medical; and Audrey Artignan, Costello Medical, for their contributions in the preparation of this poster. We thank IHM for the collection and provision of the deidentified data used for the research presented in this poster.

Presented at ISPOR International 2024 | Atlanta, GA, US | 5–8 May