ORACLE
Life Sciences

RWD131

Oracle Real-World Data: Analysis of the Characteristics and Representativeness of Oracle's Linked EHR-Claims Database

Seham Issa, **Hanaya Raad**, Fayssoil Fouad, Ramzi Agroubi, Stacey Purinton, Martina Furegato

Oracle, Paris, France

Objectives

Oracle Health Real-World Data (RWD), an Electronic Health Records (EHR) fueled by one of the largest learning health networks (LHN) in the United States (US), is linkable with Oracle Life Sciences Closed Claims Data. This study aims to determine the general characteristics of the linked datasets in comparison with the national US Census Bureau estimates.

Methods

Oracle EHR RWD and Claims* data were linked through patient-level tokenization. Demographics were assessed for individuals with at least one encounter between May 2024 and April 2025 (active patients). Population characteristics were compared with the publicly available Census bureau statistics (2023 estimates of age, sex, race and Hispanic origin; 2024 estimates for geographic distribution).

*All EHR RWD data current as of February 2025 and pulled from Oracle Health Data Intelligence; Not an exhaustive list. EHR RWD race and ethnicity percentages calculated from categories that align to CDC race concepts. U.S. Census data are based on July 1, 2023 Estimates of the Population by Race and Hispanic Origin for the United States. EHR RWD age as of February 2025 among patients with an encounter in the last 12 months. EHR RWD gender inclusive of only female and male categories for HIPAA deidentification purposes. Age and gender US Census data are based on U.S. Census 2023 Population by Age and Sex.

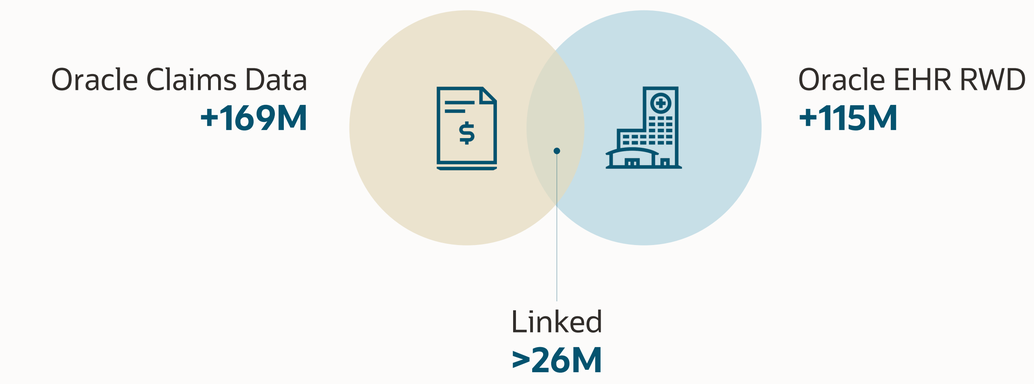
Results

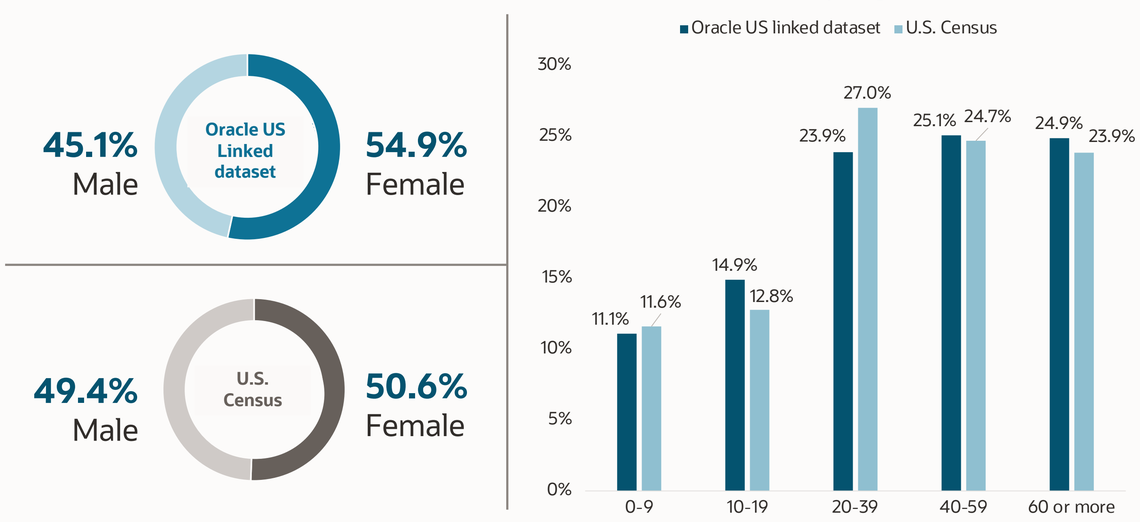
In total, the linked dataset included 26,274,088 unique individuals, of which 11,069,595 were active patients. The proportion of people aged 0-9, 10-19, 20-39, 40-59, 60 or more years was respectively 11.1%, 14.9%, 23.9%, 25.1% and 24.9%, versus 11.6%, 12.8%, 27.0%, 24.7% and 23.9% in the Census estimates.

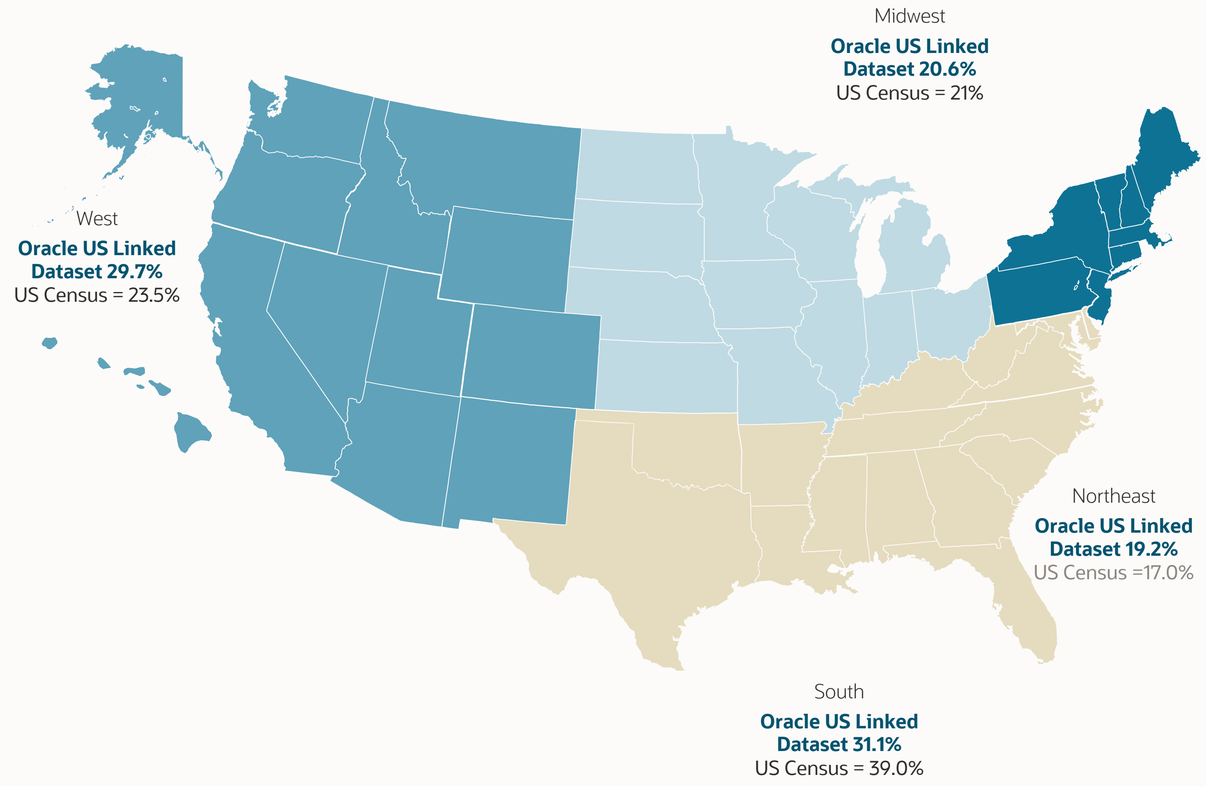
In total, 54.9% were female (versus 50.6%), 19.2% were located in the Northeast (versus 17.0%), 20.6% in the Midwest (versus 20.5%) , 31.1% in the South (versus 39.0%) and 29.7% in the West (versus 23.5%).

Race was distributed as follows: 75.0% White (versus 75.3%), 11.2% Black/African American (versus 13.7%), 3.3% Asian (versus 6.4%), and 9.5% had two or more races (versus 3.1%).

Among patients with complete ethnicity data, the distribution of non-Hispanics in the linked dataset compared to census data was 74.3% versus 80.5%.







*All EHR RWD data current as of February 2025 and pulled from Oracle Health Data Intelligence; Not an exhaustive list. EHR RWD race and ethnicity percentages calculated from categories that align to CDC race concepts. U.S. Census data are based on July 1, 2023 Estimates of the Population by Race and Hispanic Origin for the United States. EHR RWD age as of February 2025 among patients with an encounter in the last 12 months. EHR RWD gender inclusive of only female and male categories for HIPAA deidentification purposes. Age and gender US Census data are based on U.S. Census 2023 Population by Age and Sex.

Scan to download a copy of this poster

Conclusion

The linked Oracle RWD EHR-Claims data creates a uniquely powerful dataset enabling access to longitudinal reimbursement data with detailed clinical information. Compared to the US Census data, it provides an optimal demographic representativeness and a wide geographic presence, supporting precise understanding of patient care trajectories and strengthening the generalizability of research findings.

References

1. July 1, 2023 Estimates of the Population by Race and Hispanic Origin for the United States.