Characterization of a Large Scale Real-World Data Source: Impact of Linking Closed Claims with Electronic Health Record Data

RWD31

Amanda M. Moore, Marley Boyd, Steve Sison, E. Susan Amirian, Jeffrey S. Brown | TriNetX, LLC, Cambridge, MA

INTRODUCTION

Research using real-world data (RWD) commonly uses electronic health records (EHRs) or administrative claims; however, each data source has potential strengths and limitations to conder. When selecting a data source for research purposes, it is important to consider how well it can 1) address the study question, and 2) represent the underlying study population.[1,2] This study aimed to characterize two large data networks from TriNetX, the EHR-only Dataworks-USA and the Linked (EHR+Claims) networks, and describe how linking claims and EHR data together may fill information gaps.

Sickle cell disease (SCD) is a disorder known to be associated with a high degree of healthcare burden, including both acute and chronic complications. Here, SCD was used as a potential use case to describe how linking closed claims with EHR data can provide a more complete picture of the patient journey.

METHODS

TriNetX's Dataworks (DW)-USA is a federated EHR-only research network of de-identified EHR data sourced directly from 71 healthcare organizations with over 125 million patients in the United States (as of June 2025). Within DW-USA, the Linked (EHR+Claims) network includes a subset of approximately 21 million patients who have been tokenized and have linked closed claims and mortality data in addition to data from the EHRs.

As a use case, patients with an ICD-10-CM diagnosis code for SCD (D57.x; except D57.3) from 01 January 2022 to 31 December 2024 (first code in observation period = index date) were described in both the DW-USA and Linked networks. To establish a baseline period, patients must have had at least one health system encounter ≥12 months before index (both networks) and continuous insurance enrollment in the year before index (Linked only). Patients were followed until loss to follow-up, death, or the end of the dataset.

Patients with SCD were descriptively compared between the two networks. Demographics were assessed on the index date. Clinical characteristics, including vaso-occlusive events and other SCD-related complications, laboratory tests, procedures, medications, and healthcare resource utilization were described in the follow-up period. Effect size between networks was estimated by Cohen's *d*/Cohen's *h*.

	Dataworks-USA (N=35,909)	Linked (EHR+Claims) (N=9,940)	Cohen's d/h
	n (%)	n (%)	
Age (years)			
Mean (SD)	33 (20)	34 (17)	0.05
Sex, n (%)			
Female	20,540 (57.2)	5,780 (58.2)	0.02
Male	15,100 (42.1)	4,160 (41.8)	<0.01
Unknown	269 (0.7)	0 (0.0)	0.17
Race, n (%)			
Black/AA	30,099 (83.8)	7,030 (70.7)	0.32
White	2,040 (5.7)	668 (6.7)	0.04
Asian	251 (0.7)	28 (0.3)	0.06
NH/PI	25 (0.1)	10 (0.1)	0.01
AI/AN	82 (0.2)	15 (0.2)	0.02
Other	970 (2.7)	607 (6.1)	0.17
Unknown	2,442 (6.8)	1,582 (15.9)	0.29
Ethnicity, n (%)			
Hispanic or Latino	1,494 (4.2)	416 (4.2)	<0.01
Not Hispanic or Latino	28,084 (78.2)	6,492 (65.3)	0.29
Unknown	6,331 (17.6)	3,032 (30.5)	0.30

Table 1. Baseline demographics of patients with SCD in DW-USA and Linked.

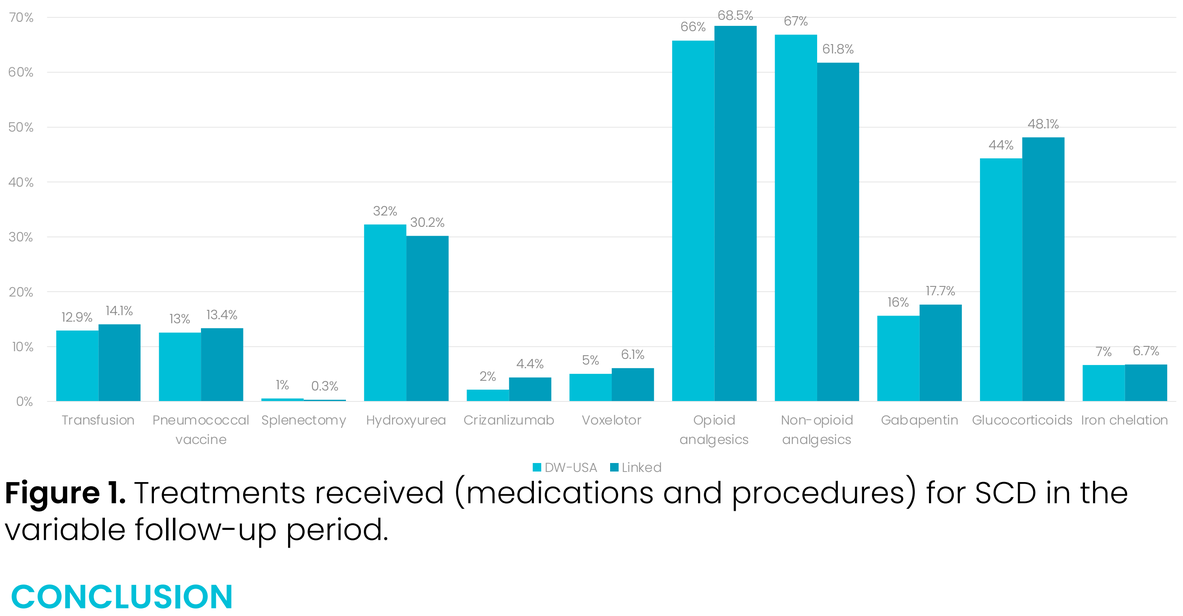


Figure 1. Treatments received (medications and procedures) for SCD in the variable follow-up period.

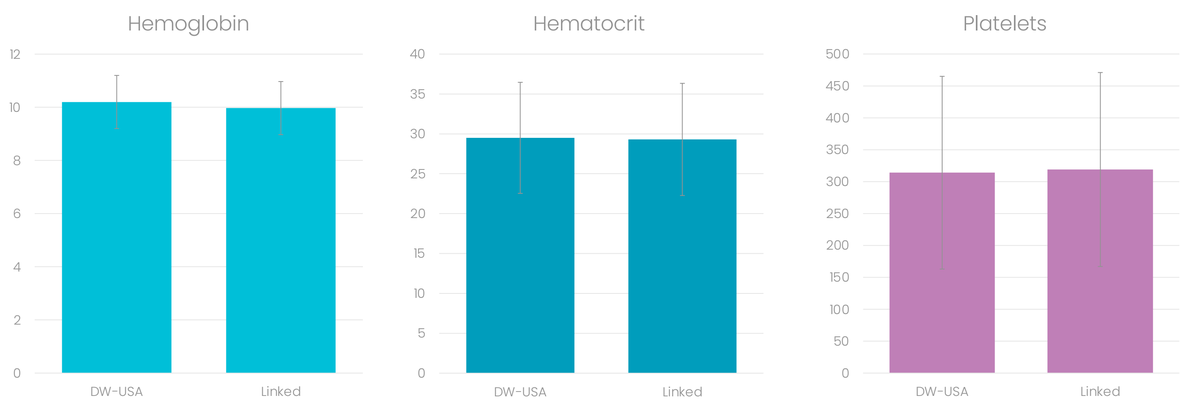
RESULTS

Among patients with SCD in DW-USA (n=35,909) and Linked (n=9,940), demographics at index were largely similar between networks, though a higher proportion of patients in DW-USA were Black or African American (83.8% v. 70.7%; h=0.32) (Table 1).

The proportions of patients with observable SCD-related procedures (e.g., blood transfusions) in the follow-up period were similar between the DW-USA and Linked cohorts. Likewise, proportions of patients treated with medications for SCD and its sequelae were similar between networks (Figure 1).

Results from laboratory tests in the follow-up period were similar between groups, including average values from complete blood counts (Figures 2-4).

Patients with SCD in DW-USA had a high degree of morbidity, including vaso-occlusive events (72.1%), anemia (30.6%), and dyspnea (22.0%); over half of the patients (52.9%) had an observed hospitalization in the follow-up period and 61% had an emergency visit. In the cohort with linked claims, a higher proportion of healthcare visits were observed (e.g., 77.9% with an emergency visit; h=0.37), subsequently resulting in a higher proportion of patients with comorbid diagnoses (e.g., 81.9% vaso-occlusive events; h=0.23) (Figure 5, Table 2).



Figures 2-4. Mean hemoglobin, hematocrit, and platelet counts in the variable follow-up period.

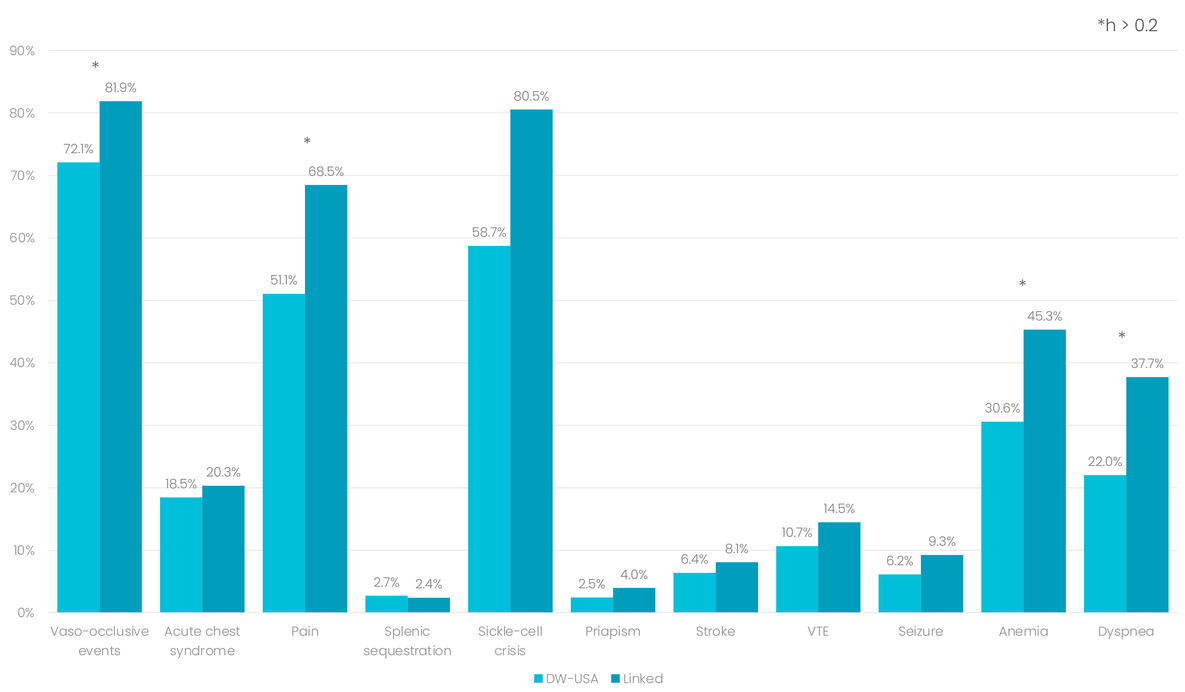


Figure 5. SCD-related diagnoses for acute and chronic complications in the variable follow-up period.

	Dataworks-USA (N=35,909)	Linked (EHR+Claims) (N=9,940)	Cohen's d/h
	n (%)	n (%)	
Emergency department visit, n (%)	21,905 (61.0)	7,745 (77.9)	0.37
Hospitalization, n (%)	19,000 (52.9)	5,770 (58.1)	0.10
Outpatient visit, n (%)	30,810 (85.8)	9,090 (91.45)	0.18

Table 2. Health system encounters in the variable follow-up period.

CONCLUSION

Using SCD as an example use case, the subset of Linked patients were similar to the source DW-USA population, particularly in regard to demographics and disease severity (based on blood counts and transfusions). While the Linked network includes less patients than the EHR-only network, linking closed claims improves longitudinal data capture and data continuity, including healthcare resource utilization and diagnosis capture. Linking closed claims also improves data granularity, particularly with regards to prescription dispensing information (e.g., days supplied).

Abbreviations

RWD=real-world data; EHR=electronic health record; SCD=sickle-cell disease; DW=Dataworks; SD=standard deviation; AA=African American; NH=Native Hawaiian; PI=Pacific Islander; AI=American Indian; AN=Alaska Native; VTE=venous thromboembolism

References

1 Schneeweiss S, Desai RJ, Ball R. A future of data-rich pharmacoepidemiology studies: transitioning to large-scale linked electronic health record + claims data. Am J Epidemiol. 2025; 194(2): 315-321.

2 Desai RJ, Marsolo K, Smith J, et al. The FDA Sentinel Real World Evidence Data Enterprise (RWE-DE). Pharmacoepidemiol Drug Saf. 2024; 33(10): e70028.

ISPOR Europe 2025 | November 9-12, 2025 | Glasgow, Scotland, UK

Design services were provided by Kelly Marsocci and were funded by TriNetX, LLC.

Download Poster PDF