

Evaluating Electronic Data Network (EDN) and Direct Facility Retrieval (DFR) Data Sources for Improved Patient Record Capture in Observational Studies

Priyanka Ramamurthy, Seth Copolnix, Emily Cibelli, Ruby Maa



Purpose

This study evaluated the contributions of healthcare encounter data obtained through Electronic Data Networks (EDNs) and Direct Facility Retrieval (DFR). Given the limited control over data retrieved via EDNs, we aimed to assess the potential for bias introduced through this method. Additionally, we investigated how combining retrieval approaches could improve the completeness of records for research.

Background

Retrieval through EDN is done by accessing an electronic network of medical record systems. Patient identifying information (e.g., full name, date of birth, address) is provided by the querying party requesting records to the network. The EDN matches the provided patient identifying information to medical records prior to delivering records to the querying party. Once retrieved, structured data elements are parsed with data pipelines and unstructured notes are processed by humans assisted by AI. Medical records are retrieved faster and provided mostly in an electronic format that eases processing. However, because the EDNs control the patient matching algorithm, there is less control and transparency in the retrieval process.

Conversely, DFR entails requesting records from individual facilities. Depending on the facility, the requests can be sent through eFax, email, portal or physical mail, and retrieved data may be sent through similar methods. Paper records are digitized. From there, similar to processing unstructured notes from EDN, humans assisted by AI process the records. DFR is slower and less scalable. Further, due to the format of the records received, more sophisticated technologies are required to digitize and process.

Methods

We collected records for 3,842 US-based patients with confirmed multiple sclerosis diagnosis who consented to record retrieval. EDN records were received in mostly XML format with a smaller percentage in PDF. DFR records were retrieved from individual records offices. DFR records were received in a wide variety of formats including paper, CD, and PDF. Records were processed in the manner described in the Background section, which produced *documents*, *section*, and *mentions* of different clinical entities.

To understand potential for bias introduced through EDN, we compared patient demographics across volume of documents retrieved from EDN. We examined the EDN document volume with the following patient demographics: sex, race, ethnicity, age, and geography.

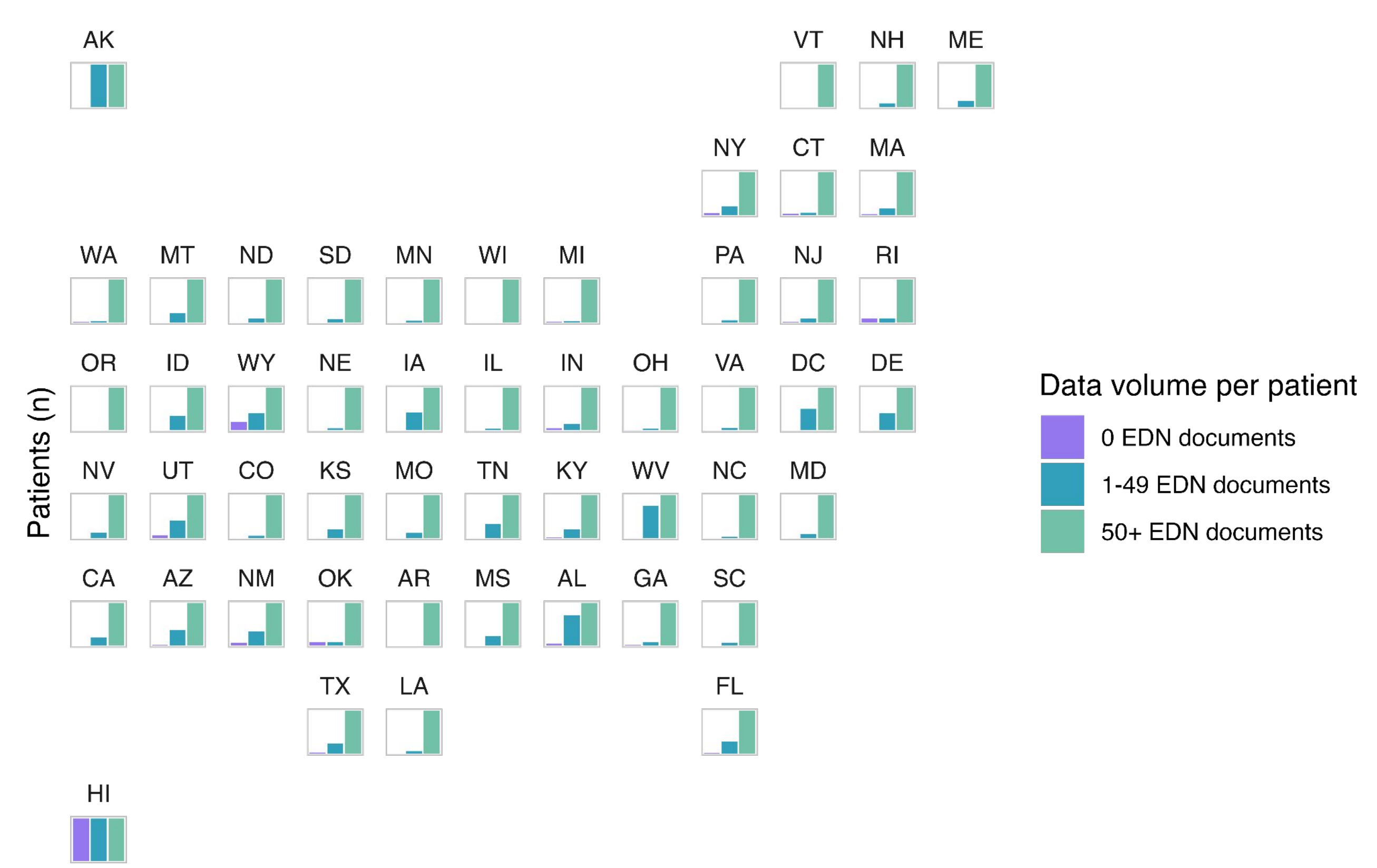
To understand how combining retrieval methods could improve completeness, we examined number of visits, number of specialists, and observation era across EDN only, DFR only, and combined. We did the same with lab measurements where we examined the total volume and annualized volume of labs. Individual lab mentions are deduplicated by collection date and lab concept.

RETRIEVED DOCUMENT VOLUME FROM EDN

Table 1: Patient demographics and retrieved EDN data volume

	0 EDN docs retrieved (N = 41)	1-50 EDN docs retrieved (N = 479)	50+ EDN docs retrieved (N = 3,322)
SEX (n, %)			
Female	30 (73%)	357 (75%)	2,631 (79%)
Male	11 (27%)	122 (25%)	689 (21%)
RACE (n, %)			
Black or African American	4 (10%)	69 (14%)	569 (17%)
Other	8 (19%)	63 (14%)	349 (11%)
White	29 (71%)	347 (72%)	2,404 (72%)
ETHNICITY (n, %)			
Hispanic or Latino	3 (7%)	59 (12%)	266 (8%)
Not Hispanic or Latino	38 (93%)	398 (83%)	2,882 (87%)
Unknown/Prefer not to say	0 (0%)	21 (4%)	164 (5%)
AGE, YEARS (median, IQR)			
	46 (40, 54)	47 (40, 55)	49 (40, 57)

Figure 1: Patient geography and retrieved EDN data volume



DATA OVERLAP AND DIFFERENCES

Figure 2: Visit overlap across DFR and EDN

TOTAL: 485,800

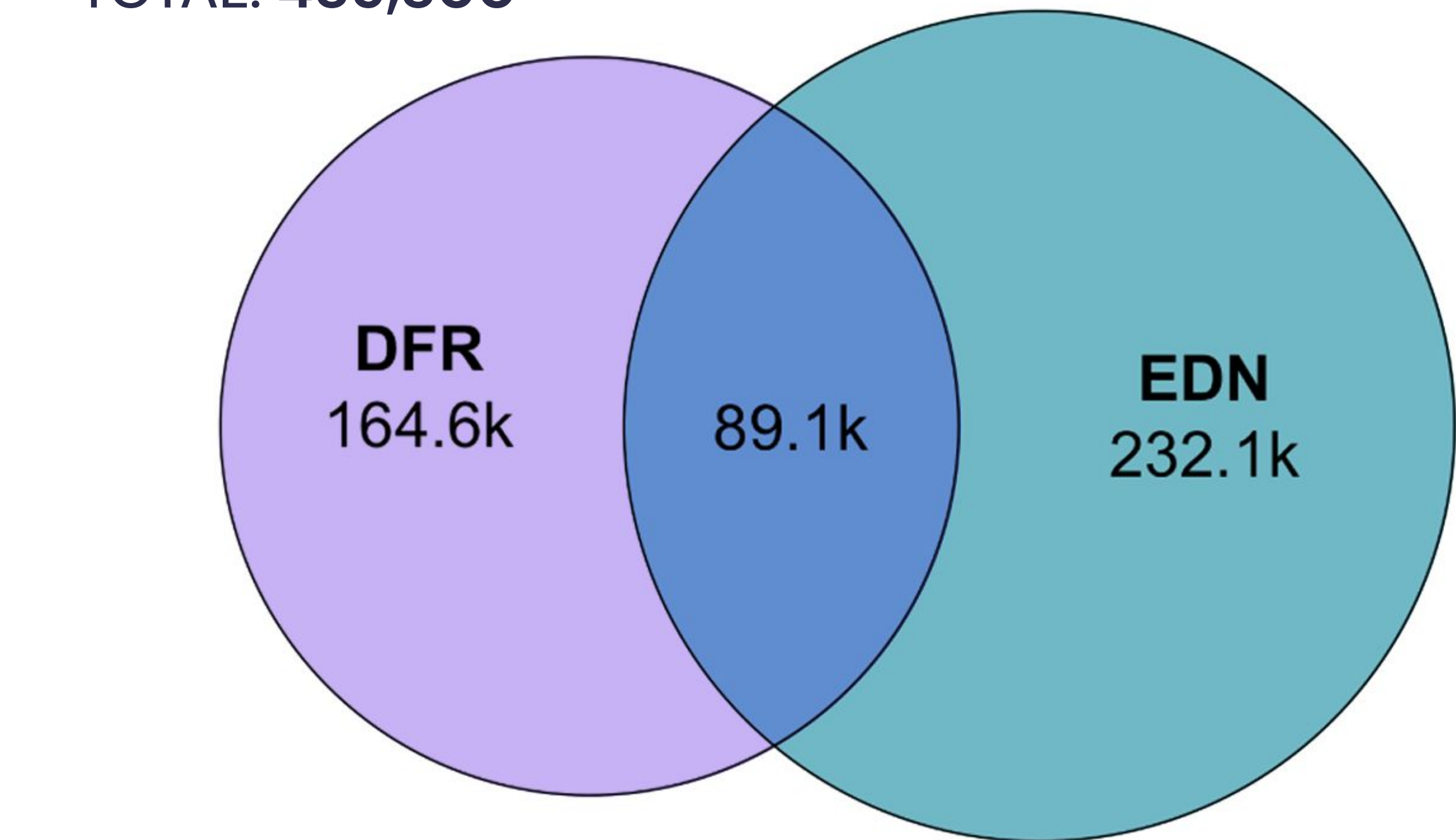


Table 2: Clinical data overlap and differences between EDN, DFR, and combined

	EDN	DFR	Combined
VISITS (N)			
Mean	83.6	66.1	126.5
Median	50	50	95
SPECIALISTS (N)			
Mean	16.5	16.3	22.9
Median	10	10	16
VISIT SPAN (YEARS)			
Mean	5.18	5.86	8.12
Median	3.91	5.63	7.9
LABS ANNUALIZED RATE (N/YEAR) ¹			
Mean	59.2	46.7	97.2
Median	35.4	31.8	71.2
LABS TOTAL (N) ²			
Mean	525.6	564.7	1005.3
Median	325	367.5	763.5

¹“Labs, annualized rate” is number of labs per year identified for the past 5 years
²“Labs, total” is the number of labs identified ever for the patient

Conclusion

In this study, we were able to retrieve 50+ documents from the majority of patients. Per Table 1, 3,322 (86%) of patients had 50+ EDN records available, 479 (12%) had 1–49 EDN records, and 41 (2%) had 0 EDN records.

We had a hypothesis that we might retrieve fewer documents from EDN for female patients since they commonly change their last name, introducing potential challenges to patient matching. We had a separate hypothesis that we might retrieve fewer documents for certain states who have different regulation on the participation in EDNs. When examining the patient demographics distribution across the different retrieved EDN data volume, there were no trends in age, sex, race, ethnicity, or geography distinguishing patients (Table 1, Figure 1).

In conclusion, we did not observe any individual level biases introduced from EDN retrieval.

Across EDN and DFR sources, we retrieved records from approximately 485,800 healthcare visits for the 3,842 patients. 66% (321.2k) of visits appeared in EDN and 52% (253.7k) in DFR records; with only 18% (89.1k) of visits having records from both sources (Figure 2).

The two retrieval methods enhanced granularity within existing observation periods, and also expanded the temporal scope of available data. The full dataset with both EDN and DFR data demonstrated a median visit span of 7.9 years, compared to the subset of records from DFR (5.63 years) or EDN (3.91 years) alone. Both the number of visits and specialist providers identified increased when combining both retrieval methods. When examining lab measurements extracted from the retrieved records, we observed a similar increase when combining both methods (Table 2).

Despite minimal demographic variations in EDN availability, the absence of EDN data for a minority of patients highlights potential systematic gaps in single-source datasets. The integration of EDN and DFR data is crucial for maximizing record capture for observational research. Anecdotally, we noticed that EDN provided a higher percentage of records that do not contain relevant medical information when compared to DFR. One potential future research direction is to systematically quantify and compare the relevant information density of both sources.

Ultimately, neither source alone provided comprehensive coverage: completeness as measured by visit count, specialist count, follow-up time, and measurement count were improved when both EDN and DFR retrieval were used.

Disclosures

Authors are employees of PicnicHealth.