Addition of Narrative Text Abstraction to ICD-Based Abstraction Significantly Improves Identification of Lupus Nephritis in Real-World Data

Meghan Tierney, PhD, RN[1]; Chris Rowe, PhD[1]
[1]PicnicHealth, San Francisco, CA

Background / Purpose

Lupus nephritis (LN) is often under-recognized in clinical practice and difficult to identify retrospectively in electronic health records (EHR), presenting challenges for clinicians and researchers hoping to explore this condition using real-world data (RWD). Common abstraction techniques use ICD codes to pull diagnoses from EHR. However, documentation of LN as a specific diagnosis is an inconsistent practice across providers, leading to underrepresented LN prevalence. Adding abstraction of unstructured narrative text may improve RWD for LN research.

Methods

We retrieved medical records for patients with lupus who had been recruited and onboarded using the PicnicHealth platform. To be included in the study, patients were required to have a lupus diagnosis documented in their medical record. Natural language processing and human-reviewed machine learning were used to abstract and structure the data. We compared 2 approaches to identifying LN in EHR: 1) Query structured sections of the EHR using an ICD-based case definition (ICD-only, see box to the right); 2) Employ the first approach but also query all unstructured narrative text for LN-related terms (ICD + narrative text, Figure 1). All terms identified using either approach were manually reviewed for validity. For each approach, we calculated the proportion of the sample with evidence of LN (percent, 95% CI) and compared performance using McNemar's test.

Figure 1: Examples of lupus nephritis related terms identified in narrative text of medical records of two patients.

Patient A

Right renal biopsy [redacted] diffuse lupus nephritis class IV global glomerular sclerosis (2/25) there is severe activity in no evidence of significant chronicity identified on the biopsy.

Patient B

Patient ID: [redacted] is a [redacted] y.o. female who returns for follow up of her lupus nephritis (class V), Ehler Danlos Syndrome, Arnold-Chiari malformation and chronic back pain. She was last seen in clinic in August, [redacted]

Table 1: Demographic and clinical characteristics of patient sample (n=635).

Characteristic	Median (IQR)
Female, *n (%)*	602 (95)
Age	47 (40-56)
Total Years of Medical Records	5.4 (2.4-10.2)
Number of Providers	6 (3-13)
Number of Care Sites	3 (1-5)

Figure 2: Geographic distribution of patient sample (n=635).

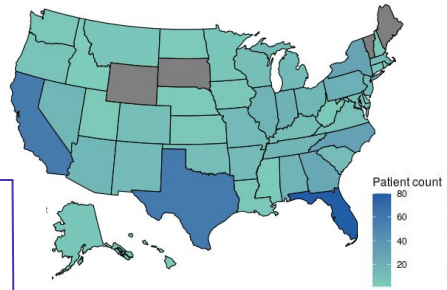
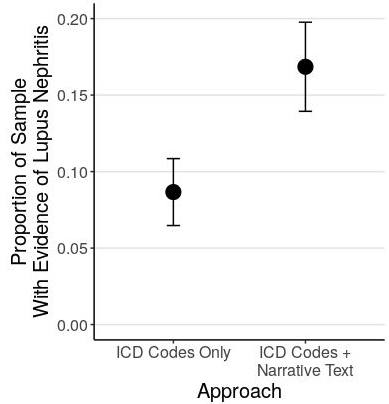


Figure 3: Comparison of the performance of ICD code-only abstraction to ICD code + narrative text abstraction of lupus nephritis from electronic health records.



ICD-based Lupus Nephritis Case Definition[1,2]: Patients were required to have either presence of the specific ICD-10 code for lupus nephritis (M32.14) *or* presence of ≥2 of the following ICD-9 codes for clinical indicators of nephritis: 583.81, 791.0, 581.9, 580.0, 582.*, 583.*, 584.*, 585.5, 585.9, 586.*, 585.6, or V56.*.

Results

Our sample of 635 patients with a diagnosis of lupus in their medical record was 95% female, had a median age of 47 years (IQR=39-56), and resided in 46 U.S. states and Washington D.C (Table 1; Figure 2). Records spanned a median of 5.4 years (IQR=2.4-10.2), including a median of 6 (IQR=3-13) providers and 3 (IQR=1-5) care sites.

The ICD-only approach identified evidence of LN in 55 (8.7%, 95%CI = 6.5-10.8%) patients and the ICD + narrative text approach identified 107 (16.9%, 13.9-19.8%) patients ($p < 0.001$, Figure 3). Leveraging unstructured narrative text identified 95% more patients with LN than only using ICD codes in structured sections of the EHR.

Conclusions

Narrative text abstraction significantly increased capture of LN in this sample of patients with lupus, demonstrating that use of ICD codes alone underrepresents prevalence. An ICD + narrative text has the potential to improve the quality of RWD and better facilitate the generation of RWD for lupus research. Future research should investigate the remaining gap between the rates of LN identified in this manner and the prevalence expected from clinical practice.

References

1.Li T, Lee, I Jayakumar D, et al. Development and validation of lupus nephritis case definitions using United States veterans affairs electronic health records. *Lupus* 2021; 30: 518-526.
2.Chibnik LB, Massarotti EM, Costenbader KH. Identification and validation of lupus nephritis cases using administrative data. *Lupus* 2010; 19:741-743

DISCLOSURE: Authors are employees of PicnicHealth.