Electronic medical records (EMR) are increasingly utilized in clinical practice and research, driving for more efficient availability of rich patient records. However, most use of EMR is limited to coded, structured data, while the vast majority of patient information (e.g., disease subtype, severity, medical device usage, etc.) is tied up in narrative clinical notes. Historically this has been done via timely and costly manual chart review, but as the amount of EMR data increases exponentially, manual chart review become impractical and impossible.

Advancements in Natural Language Processing (NLP) have demonstrated promising results in combining the capture of additional clinical note information with the efficiency of modern informatics.

A comparison of traditional abstraction of EMR data and NLP-driven data utilizing one of the five projects analyzed in this abstract can be found in Figure 1.

CONCLUSION

- The objective of this study is to demonstrate the relevancy and utility of NLP to extract health data from EMR in real-world observational studies.

METHODS

- We conducted a systematic review and meta analysis of performance metrics for five (5) NLP-driven projects involving oncology, inflammation, central nervous system disorders and medical devices, which had similar protocols and objectives.

- Figure 2 is a visual representation of the disease area, the project and the elements of the study that required NLP. The ANCA-Associated Vasculitides (AAV) project encompasses Granulomatosis with Polyangiitis (GPA), Microscopic Polyangiitis (MPA) and Churg-Strauss Syndrome (CSS), as described in Figure 3.

- We assessed and validated the accuracy of NLP algorithms, as well as heterogeneity of accuracy between studies using random effects meta-analysis (represented by I² value).

RESULTS

- A total of 382,523 patients were identified using NLP among the 5 studies.

- Accuracy was reported among 4 of the studies ranging from 95.2% to 100% (95% CI: 95.1%, 100%), with an I² value of 95.9% (95% CI: 92.9%, 97.7%).

NLP provides a unique opportunity to extract meaningful information from patient-level narrative clinical notes in EMR data sources with high degree of accuracy. This provides additional rich sources of data tied up in narrative clinical notes, that are otherwise not easily available, to support epidemiology and other real-world observational studies.

REFERENCES


