

IDENTIFICATION AND CHARACTERIZATION OF IMMUNE CHECKPOINT INHIBITOR INDUCED IMMUNE-RELATED ADVERSE EVENTS FROM ELECTRONIC HEALTH RECORDS USING NATURAL LANGUAGE PROCESSING TECHNIQUES

¹Hannah Barman, ²Sriram Venkateswaran, ²Antonio Del Santo, ¹Krishna Rao, ¹Bharathwaj Raghunatha, ¹Unice Yoo, ³Lisa Kottschade, ³Matthew Block, ²G Scott Chandler, ¹Tyler Wagner, ²Rajat Mohindra*. ¹*Inference, Cambridge, MA, USA*; ²*F. Hoffmann-La Roche, Basel, Switzerland*; ³*Mayo Clinic, Rochester, MN, USA*

Background Immune checkpoint inhibitors (ICIs) have revolutionized cancer treatment, yet their use is associated with immune-related Adverse Events (irAEs). Estimating prevalence and patient impact of these irAEs in the Real-World Data (RWD) setting is critical for further characterizing the benefit/risk profile of ICI therapies beyond the clinical trial population. Studies using International Classification of Diseases (ICD) codes have aimed to understand the safety and effectiveness of drugs. However, this approach does not comprehensively illustrate a patient's care journey, sub-optimally captures patients' concurrent medical conditions, and offers no insight into drug-AE causality. The present study aims to more accurately capture the relationship between ICIs and irAEs by using Augmented Curation (AC), a natural language processing (NLP) based innovation, on unstructured data in Electronic Health Records (EHRs).¹⁻⁴

Methods In a cohort of approximately 6,000 patients treated with ICIs at Mayo Clinic, we compared the prevalence of irAEs using ICD codes and AC which leverages SciBERT to perform downstream tasks including entity extraction, classification, and relationship extraction. Each task was performed by a bespoke AC model that was trained using clinical scientist annotated datasets from Mayo's EHRs.¹⁻⁴ These models were orchestrated together to create an ensemble workflow that detected mentions of drug-AE pairs in clinical notes with implied textual causality. Finally, select irAEs with high patient impact- myocarditis, encephalitis, pneumonitis and SCAR (MEPS) were analyzed further. AC-extracted corticosteroid/immunosuppressive administration and therapy discontinuations were used as proxies of severity.

Results For all irAEs, only 28.5% of the patients found by AC were also identified with structured codes. For MEPS, only 30.5% of patients found by AC were also identified by ICD codes. Using pneumonitis as an example, AC achieved a sensitivity of 0.94, while ICD codes achieved only 0.26. MEPS patients were found to receive corticosteroids for their respective irAE 69% of the time and subsequently discontinued the ICI due to the irAE 48% of the time.

Conclusions Overall, the AC model identified additional irAEs not detected by ICD codes, and the positive sentiment-based approach helped assess the drug-irAE relationships in unstructured clinical notes thus supporting assessment of causal association with use of ICI therapies. The use of AC to accurately detect key irAEs allows physicians to leverage EHRs to discover risk factors for specific irAEs in RWD settings, as well as to review clinical outcomes to identify best practices in treating irAEs.

REFERENCES

1. Murugadoss K, Rajasekharan A, Malin B, Agarwal V, Bade S, Anderson JR, Ross JL, Faubion WA Jr, Halamka JD, Soundararajan V, Ardhanari S. Building a best-in-class automated de-identification tool for electronic health records through ensemble learning. *Patterns (NY)*. 2021;**2**(6):00255. PMID: 34179842
2. Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. arXiv preprint arXiv. 2019;1903.10676

3. Pawlowski C, Venkatakrishnan AJ, Ramudu E, Kirkup C, Puranik A, Kayal N, Berner G, Anand A, Barve R, O'Horo JC, Badley AD, Soundararajan V. Pre-existing conditions are associated with COVID-19 patients' hospitalization, despite confirmed clearance of SARS-CoV-2 virus. *Eclinical Medicine*. 2021;**34**:100793. PMID: 33778434
4. Wagner T, Shweta F, Murugadoss K, Awasthi S, Venkatakrishnan AJ, Bade S, Puranik A, Kang M, Pickering BW, O'Horo JC, Bauer PR, Reasonable RR, Vergidis P, Temesgen Z, Rizza S, Mahmood M, Wilson WR, Challener D, Anand P, Liebers M, Doctor Z, Silvert E, Solomon H, Anand A, Barve R, Gores G, Williams AW, Morice WG 2nd, Halamka J, Badley A, Soundararajan V. Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. *Elife*. 2020;**9**:e58227. PMID: 32633720

Ethics Approval This study was approved by the Mayo Clinic Institutional Review Board; approval number 22-002906.

<http://dx.doi.org/10.1136/jitc-2022-SITC2022.1274>