HHS Public Access

Author manuscript

JCO Clin Cancer Inform. Author manuscript; available in PMC 2024 October 10.

Published in final edited form as:

JCO Clin Cancer Inform. 2024 May ; 8: e2400051. doi:10.1200/CCI.24.00051.

Navigating the complexities of AI-enabled real-world data collection for oncology pharmacovigilance

Jack Gallifant, MBBS[1,2], **Leo Anthony Celi, MD**[1,3,4], **Elad Sharon, MD**[5], **Danielle S. Bitterman, MD**[6,7,*]

[1]Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, MA 02139.

[2]Department of Critical Care, Guy's & St Thomas' NHS Trust, London, United Kingdom, SE1 7EH.

[3]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215.

[4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115.

[5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

[6]Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, MA, USA

[7]Department of Radiation Oncology, Brigham and Women's Hospital/Dana-Farber Cancer Institute, Boston, MA, USA

The ubiquitous uptake of electronic health records (EHRs) in the United States, combined with advances in artificial intelligence (AI), presents new opportunities to leverage real-world data (RWD) - collected as a part of routine clinical care - to complement clinical trials for oncology pharmacovigilance. Clinical trials are an essential component of drug safety evaluation but have inherent limitations that make them a necessary but insufficient data source for cancer drug safety. [1–3] Trials generally include selective populations that are not representative of many patients seen in cancer clinics who may be more susceptible to adverse events (AEs), and do not capture the full range of clinical practice.[4, 5] Further, trials in the United States predominantly include white patients and individuals with a higher socioeconomic status, which may affect the generalizability of results.[6–8] Clinical trials may also underestimate the actual long-term burden of AEs due to incomplete reporting[3] and limited or delayed reporting timelines. RWD in EHRs can provide more comprehensive evidence for drug safety and support a learning health system with real-time AE monitoring, thereby improving cancer treatment outcomes.[9] Here, we discuss how AI, especially natural language processing (NLP), can help realize this potential and discuss the complexities of appropriate development of these AI-enabled methods. We propose the "Three Ps" framework - Processing, Pipelines, and Patient Outcomes - to inform effective

[*]**Corresponding author:** Dr. Danielle S. Bitterman, Department of Radiation Oncology, Dana-Farber Cancer Institute/Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, dbitterman@bwh.harvard.edu, Fax: (617) 975-0985, Twitter handle: @dbittermanmd.

pharmacovigilance drug safety measures incorporating AE information from EHR text (Table 1).

The systematic and timely identification of real-world AE rates in EHRs are limited by their documentation in unstructured EHR text, requiring labor- and expertise-intensive manual review and abstraction of events. Relying only on structured data that can be directly abstracted, such as billing codes and laboratory data, often underestimates AEs, especially mild-moderate AEs, that can impact patients' quality-of-life and ability to stay on an otherwise effective cancer treatment.[10–13] Furthermore, structured data are often under-specified and cannot provide the level of detail about severity, causality, and temporal trends necessary to inform clinical decision-making processes and carry out quality RWD studies.

To overcome the bottleneck of manual EHR curation for AE pharmacovigilance, Barman et al. report the results of NLP methods to automatically extract immune-related AE (irAE) occurrences from unstructured text. NLP is a field of AI that enables computational processing of human language for various downstream tasks, including automated information extraction. In recent years, a class of deep learning models called Transformer language models [14, 15] has gained traction across the field and now forms the backbone for many NLP methods. Barman et al. fine-tune a Transformer-based language model for the automated curation of clinical notes for irAE detection and compare the rates of irAEs detected by the language model versus a set of pre-defined billing codes. There was low agreement between myocarditis, encephalitis, pneumonitis, and severe cutaneous AEs identified by language model and structured data methods.

This study falls within a broader literature showing the promise of NLP for detecting AEs, including irAEs, in EHRs [12, 13, 16–24] and highlights the challenges and complexities of AI-augmented EHR pharmacovigilance. In this study, irAE rates were much lower than reported elsewhere; for example, their overall irAE rate of 20.8% is lower than the aforementioned rates in RWD studies and clinical trials, where they have been reported to impact up to 80% of patients.[24–26] Similarly, most studies report immune-related pneumonitis rates of 10–20%,[27–30] compared to 2.9% in this study. These discrepancies underscore the challenges of shifting from easy-to-measure and available structured data that likely suffer in sensitivity and specificity, to incorporating information in unstructured text that requires manually labeled ground truths to guide learning strategies. One potential explanation for the lower rates of irAE is the data labeling method for fine-tuning the NLP model. Here, automated methods were used to identify irAE-containing text, which was subsequently used for model development. The irAE rates highlight the potential ramifications of using non-expert verified labels for model development. Training and evaluating NLP-based methods on unreliable ground truths is an ongoing challenge in the field, and best practices still rely on significant manual annotation.

This study also touches on the challenge of attributing an AE to its inciting agent, i.e., causality extraction. This is essential in oncology, where most patients are exposed to varying combinations of therapies, and determining the causative agent is often a diagnostic challenge. While some AEs can be definitively attributed via biopsy, this is not frequently

done in practice, and many others have no pathologic gold standard to determine causality. This leads to a reliance on supporting clinical data and judgment, both of which will be documented to varying degrees. Consequently, the accuracy and utility of NLP models in this context are inherently limited by the quality of the underlying documentation for direct attribution. As an alternative to relying on clinician judgment, statistical causal inference can be used to establish attribution from RWD,[31, 32] which requires detailed, granular extraction of AEs, treatments, relative timings, outcomes, and other potentially contributing clinical and demographic factors – and an accounting for biases and inequities in healthcare delivery, as described below.

The Three Ps Framework for NLP-Enhanced Pharmacovigilance

Leveraging NLP to mine unstructured EHR text for AE information holds potential to improve and potentially automate pharmacovigilance for cancer care. The “Three Ps” framework may guide considerations when developing methods that incorporate information extracted from text (Table 1):

- **Processing:** Development and evaluation of NLP methods for EHR text processing that ensures consistency, reproducibility, and verifiability of AE findings.
- **Pipelines:** Considerations for combining language model-extracted AE information with other EHR sources for comprehensive data coverage, quality, standardization, interoperability, and systematic reporting.
- **Patient Outcomes:** Considerations for developing methods that align with the ultimate goal of pharmacovigilance: improving patient care.

Future directions for AI-enabled pharmacovigilance

Recent advancements in language models have led to the most current generation of large language models (LLMs), which may be able to make predictions without task-specific training examples or only a few examples to guide the model - diverging from traditional methods that require larger labeled datasets. If successful, these models may reduce time and resource constraints associated with data annotation. At present, specialized fine-tuned language models still outperform generalist LLMs for most specialized tasks, including causality extraction,[33–35] although, with continued advances, this new paradigm might catalyze advances in EHR-based pharmacovigilance if evaluated and implemented appropriately.

In the future, AI models that incorporate multi-modal EHR data, such as text, labs, vitals, imaging, and pathology, may improve AE diagnosis and attribution. This may strengthen clinical evidence, drive translational research, and provide real-time diagnostic decision support. Similarly, by taking full advantage of all data within a patient’s EHR, such models could provide more consistent and standardized severity grading, a challenge with manual abstraction[36].

Once EHR extraction processes and pipelines are developed, new methods and data-sharing approaches are needed to take full advantage of RWD while addressing its biases and limitations for inference. Information about AEs and other clinical factors that may contribute to them, including demographics, cancer diagnosis, comorbidities, social determinants of health, treatment details (e.g., dosage, timing, and route), and AE-directed treatment, is often documented over long time periods, by multiple providers across different institutions. Data silos, closed EHR systems, and variability in what healthcare providers choose to document can lead to incomplete documentation within a single healthcare system.[37] Efforts to generate findable, accessible, interoperable, and reproducible (FAIR) data,[38] including the adoption of consensus data standards,[39–42] will also be imperative to overcoming these obstacles. Some data models are beginning to include AE elements,[40] but more work is needed to expand them to comprehensively capture AE information about severity, causality, and timing. In parallel, validated measures of RWD quality, including uncertainty, chart completeness, and documentation bias, are urgently needed for successful and safe implementation.

Finally, fairness and equity will need to be considered in the design of any pharmacovigilance system, especially those that will be AI-assisted. AEs may not be distributed equally, and the only way to understand the risks and benefits of treatments in diverse populations is to prioritize the representativeness of the collected data. This includes evaluation and monitoring of performance across different patient groups and the design of systems that can be widely implemented at institutions with varying resource capacities. Relatedly, any pharmacovigilance system design must have a causal framework backbone that considers left and right censoring. For example, patients with worse outcomes due to social determinants of health or social determinants of care are given less opportunity to report AEs.[43] Further, they are more at risk for competing events such as death,[44–46] and patients who are not offered treatment are not included in the denominator population for AE reporting.

In conclusion, integrating AI, particularly NLP methods, with EHRs is a significant opportunity to advance oncology pharmacovigilance, which could improve treatment outcomes and cancer patients' quality of life. Appropriate development, evaluation, and reporting will be essential to ensure that automated methods accurately identify and estimate AE rates so that we realize the full benefit and avoid the harms of AI-augmented pharmacovigilance.

Funding

JG is funded by the National Institute of Health through DS-I Africa U54 TW012043–01 and Bridge2AI OT2OD032701.

LAC is funded by the National Institute of Health through R01 EB017205, DS-I Africa U54 TW012043–01 and Bridge2AI OT2OD032701, and the National Science Foundation through ITEST #2148451.

DSB receives financial support from the Woods Foundation and the National Institute of Health through U54CA274516–01A1.

Disclaimers:

DSB: Associate Editor of Radiation Oncology, HemOnc.org (no financial compensation, unrelated to this work); Research funding unrelated to this work: AACR, NIH/NCI. Advisory Board: Mercurial AI.

References

1. Ioannidis JP, Lau J: Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA 285:437–443, 2001 [PubMed: 11242428]
2. Ioannidis JPA, Lau J: Improving Safety Reporting from Randomised Trials. Drug Saf 25:77–84, 2002 [PubMed: 11888350]
3. Miller TP, Getz KD, Li Y, et al. : Rates of Laboratory Adverse Events by Course in Pediatric Leukemia Ascertained Using Automated Electronic Health Record Extraction: A Retrospective Cohort Study Report from the Children's Oncology Group. Lancet Haematol 9:e678–e688, 2022 [PubMed: 35870472]
4. Zhao S, Miao M, Wang Q, et al. : The current status of clinical trials on cancer and age disparities among the most common cancer trial participants. BMC Cancer 24:30, 2024 [PubMed: 38166779]
5. Meyer S, Woldu HG, Sheets LR: Sociodemographic diversity in cancer clinical trials: New findings on the effect of race and ethnicity. Contemp Clin Trials Commun 21:100718, 2021 [PubMed: 33604484]
6. Bitterman DS, Bona K, Laurie F, et al. : Race Disparities in Proton Radiotherapy Use for Cancer Treatment in Patients Enrolled in Children's Oncology Group Trials. JAMA Oncol 6:1465–1468, 2020 [PubMed: 32910158]
7. Nazer L, Abusara A, Aloran B, et al. : Patient diversity and author representation in clinical studies supporting the Surviving Sepsis Campaign guidelines for management of sepsis and septic shock 2021: a systematic review of citations. BMC Infect Dis 23:751, 2023 [PubMed: 37915042]
8. Randomised controlled trial participants: Diversity data report [Internet][cited 2024 Feb 20] Available from: https://www.nihr.ac.uk/documents/randomised-controlled-trial-participants-diversity-data-report/31969
9. Celi LA, Moseley E, Moses C, et al. : From Pharmacovigilance to Clinical Care Optimization. Big Data 2:134–141, 2014 [PubMed: 26576325]
10. Kalinich M, Murphy W, Wongvibulsin S, et al. : Prediction of severe immune-related adverse events requiring hospital admission in patients on immune checkpoint inhibitors: study of a population level insurance claims database from the USA. J Immunother Cancer 9:e001935, 2021 [PubMed: 33789879]
11. Mandelblatt JS, Huang K, Makgoeng SB, et al. : Preliminary Development and Evaluation of an Algorithm to Identify Breast Cancer Chemotherapy Toxicities Using Electronic Medical Records and Administrative Data. J Oncol Pract 11:e1–8, 2015 [PubMed: 25161127]
12. Chen S, Guevara M, Ramirez N, et al. : Natural Language Processing to Automatically Extract the Presence and Severity of Esophagitis in Notes of Patients Undergoing Radiotherapy. JCO Clin Cancer Inform 7:e2300048, 2023 [PubMed: 37506330]
13. Geva A, Abman SH, Manzi SF, et al. : Adverse drug event rates in pediatric pulmonary hypertension: a comparison of real-world data sources. J Am Med Inform Assoc JAMIA 27:294–300, 2020 [PubMed: 31769835]
14. Vaswani A, Shazeer N, Parmar N, et al.: Attention is All you Need [Internet], in Advances in Neural Information Processing Systems. Curran Associates, Inc., 2017[cited 2024 Feb 20] Available from: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
15. Devlin J, Chang M-W, Lee K, et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Internet], 2019[cited 2023 Nov 13] Available from: http://arxiv.org/abs/1810.04805
16. Gupta S, Belouali A, Shah NJ, et al. : Automated Identification of Patients With Immune-Related Adverse Events From Clinical Notes Using Word Embedding and Machine Learning. JCO Clin Cancer Inform 5:541–549, 2021 [PubMed: 33989017]

17. Jagannatha A, Liu F, Liu W, et al. : Overview of the First Natural Language Processing Challenge for Extracting Medication, Indication, and Adverse Drug Events from Electronic Health Record Notes (MADE 1.0). Drug Saf 42:99–111, 2019 [PubMed: 30649735]
18. Henry S, Buchan K, Filannino M, et al. : 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. J Am Med Inform Assoc JAMIA 27:3–12, 2020 [PubMed: 31584655]
19. Wong A, Plasek JM, Montecalvo SP, et al. : Natural Language Processing and Its Implications for the Future of Medication Safety: A Narrative Review of Recent Advances and Challenges. Pharmacotherapy 38:822–841, 2018 [PubMed: 29884988]
20. Luo Y, Thompson WK, Herr TM, et al. : Natural Language Processing for EHR-Based Pharmacovigilance: A Structured Review. Drug Saf 40:1075–1089, 2017 [PubMed: 28643174]
21. Murphy RM, Klopotowska JE, de Keizer NF, et al. : Adverse drug event detection using natural language processing: A scoping review of supervised learning methods. PloS One 18:e0279842, 2023 [PubMed: 36595517]
22. Martins F, Sofiya L, Sykiotis GP, et al. : Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. Nat Rev Clin Oncol 16:563–580, 2019 [PubMed: 31092901]
23. Reynolds KL, Arora S, Elayavilli RK, et al. : Immune-related adverse events associated with immune checkpoint inhibitors: a call to action for collecting and sharing clinical trial and real-world data. J Immunother Cancer 9:e002896, 2021 [PubMed: 34215691]
24. Johnson DB, Nebhan CA, Moslehi JJ, et al. : Immune-checkpoint inhibitors: long-term implications of toxicity. Nat Rev Clin Oncol 19:254–267, 2022 [PubMed: 35082367]
25. Calabrese C, Kirchner E, Kontzias A, et al. : Rheumatic immune-related adverse events of checkpoint therapy for cancer: case series of a new nosological entity. RMD Open 3:e000412, 2017 [PubMed: 28405474]
26. Xu C, Chen Y-P, Du X-J, et al. : Comparative safety of immune checkpoint inhibitors in cancer: systematic review and network meta-analysis. BMJ 363:k4226, 2018 [PubMed: 30409774]
27. Suresh K, Voong KR, Shankar B, et al. : Pneumonitis in Non-Small Cell Lung Cancer Patients Receiving Immune Checkpoint Immunotherapy: Incidence and Risk Factors. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer 13:1930–1939, 2018
28. So AC, Board RE: Real-world experience with pembrolizumab toxicities in advanced melanoma patients: a single-center experience in the UK. Melanoma Manag 5:MMT05, 2018 [PubMed: 30190931]
29. Sun M, Ji H, Xu N, et al. : Real-world data analysis of immune checkpoint inhibitors in stage III-IV adenocarcinoma and squamous cell carcinoma. BMC Cancer 22:762, 2022 [PubMed: 35831785]
30. Naidoo J, Vansteenkiste JF, Faivre-Finn C, et al. : Characterizing immune-mediated adverse events with durvalumab in patients with unresectable stage III NSCLC: A post-hoc analysis of the PACIFIC trial. Lung Cancer Amst Neth 166:84–93, 2022
31. Pearl J: Interpretation and identification of causal mediation. Psychol Methods 19:459–481, 2014 [PubMed: 24885338]
32. Pearl J: An Introduction to Causal Inference. Int J Biostat 6:7, 2010
33. Guevara M, Chen S, Thomas S, et al. : Large language models to identify social determinants of health in electronic health records. Npj Digit Med 7:1–14, 2024 [PubMed: 38172429]
34. Lehman E, Hernandez E, Mahajan D, et al.: Do We Still Need Clinical Language Models? [Internet], 2023[cited 2024 Feb 20] Available from: http://arxiv.org/abs/2302.08091
35. Chen S, Li Y, Lu S, et al. : Evaluating the ChatGPT family of models for biomedical reasoning and classification. J Am Med Inform Assoc ocad256, 2024
36. Hsiehchen D, Watters MK, Lu R, et al. : Variation in the Assessment of Immune-Related Adverse Event Occurrence, Grade, and Timing in Patients Receiving Immune Checkpoint Inhibitors. JAMA Netw Open 2:e1911519, 2019 [PubMed: 31532516]
37. Zhang J, Morley J, Gallifant J, et al. : Mapping and evaluating national data flows: transparency, privacy, and guiding infrastructural transformation. Lancet Digit Health 5:e737–e748, 2023 [PubMed: 37775190]

38. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. : The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018, 2016 [PubMed: 26978244]

39. Osterman TJ, Terry M, Miller RS: Improving Cancer Data Interoperability: The Promise of the Minimal Common Oncology Data Elements (mCODE) Initiative. JCO Clin Cancer Inform 993–1001, 2020 [PubMed: 33136433]

40. Mayo CS, Feng MU, Brock KK, et al. : Operational Ontology for Oncology (O3): A Professional Society-Based, Multistakeholder, Consensus-Driven Informatics Standard Supporting Clinical and Research Use of Real-World Data From Patients Treated for Cancer. Int J Radiat Oncol Biol Phys 117:533–550, 2023 [PubMed: 37244628]

41. Data Standardization – OHDSI [Internet][cited 2024 Feb 20] Available from: https://www.ohdsi.org/data-standardization/

42. United States Core Data for Interoperability (USCDI) [Internet][cited 2024 Feb 20] Available from: http://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi

43. Halvorson EE, Thurtle DP, Easter A, et al. : Disparities in Adverse Event Reporting for Hospitalized Children. J Patient Saf 18:e928–e933, 2022 [PubMed: 35797590]

44. Rudolph JE, Lesko CR, Naimi AI: Causal inference in the face of competing events. Curr Epidemiol Rep 7:125–131, 2020 [PubMed: 33747714]

45. Siegel RL, Giaquinto AN, Jemal A: Cancer statistics, 2024. CA Cancer J Clin 74:12–49, 2024 [PubMed: 38230766]

46. Cancer Disparities - Cancer Stat Facts [Internet]. SEER [cited 2024 Feb 27] Available from: https://seer.cancer.gov/statfacts/html/.html

Table 1.

Overview of the Three Ps Framework for Language Model-Based Pharmacovigilance

Category	Element	Description
Processing	Definitions	Clear definition of AE information being extracted, and whether and how attribution to a treatment is defined.
	Accuracy	Precision and recall (sensitivity) of the model in identifying and extracting target AE information.
	Adaptability	Capacity of the model to incorporate new information and adjust its analysis accordingly.
	Annotation Guidelines	Development and release of clear annotation guidelines describing how AE information is labeled for model development and evaluation; inter-annotator agreement to report data labeling quality and d reproducibility.
	Code and Data Release	Sharing of code and data for verifiability and interrogation of findings.
Pipelines	Data Coverage	Comprehensive coverage of diverse data types and formats in EHRs to avoid gaps in data representation.
	Data Quality	Mechanisms and metrics to identify, quantify, and correct errors, inconsistencies, and biases in data to enable accurate, equitable causal inferencing.
	Data Standardization	Consistent harmonization and standardization strategies for comparing and merging unstructured and structured data.
	Systematic Reporting	Clear reporting of results at each stage of the pipeline and for the overall pipeline.
	Training and Validation Datasets	Reporting the number of notes for each split and breakdown according to AE type.
	Data Labeling	Documentation of labeling methods, inter-annotator agreement (for human labeling), and validation of automated labeling methods.
	Dataset Characteristics	Details about the demographics and clinical characteristics of the training and validation datasets.
Patient Impact	Early Detection	Capacity to enable early AE identification for timely intervention and management.
	Accurate Diagnosis	Capacity to enable correct AE diagnosis for correct specialist referral and management.
	Treatment Optimization	Potential of insights from methods to inform personalized treatment plans to reduce AE risk and enhance effectiveness.
	Quality of Life	Potential of insights from methods to minimize the impact of AEs on patients' quality of life.
	Health Equity	Potential of insights to provide an accurate understanding of AE risk across patient groups, especially vulnerable populations.

Abbreviations: AE, adverse events.