Apples and Oranges? Considerations for EHR-Based Analyses Aggregating Data From Interventional Clinical Trials and Point-of-Care Encounters in Oncology

Jessica A. Lavery, MS[1]; Margaret K. Callahan, MD, PhD[2]; and Katherine S. Panageas, DrPH[1]

The use of real-world evidence—defined by the US Food and Drug Administration (FDA) as "data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources"—is gaining momentum.[1-5] Information systems that were historically siloed, such as pathology, imaging, and genomic records, are becoming increasingly integrated into electronic health record (EHR) systems. This integration facilitates queries of the EHR for patients who meet specified inclusion criteria and allows for data to be aggregated into a form that is suitable for analysis, providing an opportunity to generate real-world evidence.

EHR-based cohorts may include patients who ever enrolled on an interventional clinical trial in conjunction with patients who never enrolled on an interventional clinical trial, and this is not often acknowledged or accounted for in analyses. Stipulations in clinical trial agreements, which may restrict access to data or portions of data (in arbitrary ways), can impact the completeness of the real-world data in ways that would be difficult to understand without specific knowledge of trial participation. Despite the increased integration of many clinical data elements, clinical trial enrollment status is not available by default in many EHR systems. A variable that indicates a patient's clinical trial enrollment status along with details on restrictions for data access specified by the clinical trial agreement would provide important utility in determining the appropriate analysis of these aggregated data.

Oncology patients enrolled on interventional clinical trials differ systematically from nontrial patients with respect to a number of important factors, such as demographic characteristics and health status.[6-8] For example, oncology trial inclusion criteria often specify a minimum ECOG performance status score,[9] which measures how a disease impacts a patient's ability to carry out daily living activities. Alternatively, some trials may enroll patients who are less healthy but have exhausted other treatment options. With respect to demographic characteristics, Black and Latino patients and older patients are underrepresented in oncology trials.[10-16]

There is also variation between clinical trial and point-of-care encounters with respect to the reason, frequency, and extent of medical assessments. Patients enrolled on a clinical trial follow a prespecified visit schedule, whereas point-of-care encounters typically indicate a need for medical attention.[17,18] Conversely, in mobile health studies, constant monitoring by mobile health devices may trigger a health care encounter. With respect to the frequency of assessments, in studies estimating progression-free survival, patients following a more frequent scanning schedule will have progression detected earlier compared with patients on a less frequent scanning schedule. Additionally, patients enrolled on a trial are more connected with the healthcare system due to intensive monitoring, with assessments beyond the standard of care.

These differences in patient case mix and in the reason for, frequency of, and extent of medical assessments between patients enrolled versus not enrolled in a clinical trial can lead to violation of modeling assumptions, biased estimates, and inflated type I error rates. Furthermore, because of these differences, specific analytic methods for analyzing EHR data, such as conditioning on the number of visits,[19,20] joint modeling,[21] reweighting estimators,[22] misclassification adjustment,[23] and pseudolikelihood estimation,[24] may not be appropriate when aggregating clinical trial and point-of-care encounter data. For example, the number of point-of-care encounters has been used as a proxy for a patient's health status in EHR-based analyses[19,20]; patients with a higher number of visits are likely to be sicker than patients with fewer visits. However, given that clinical trial patients follow a prespecified visit schedule, this relationship no longer holds. If relatively few clinical trial patients are included in an EHR-based cohort, using the number of visits as a proxy for health status may have few implications for statistical inference, but if the cohort includes a high proportion of patients who were on a clinical trial,

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on October 22, 2020 and published at ascopubs.org/journal/cci on January 7, 2021: DOI https://doi.org/10.1200/CCI.20.00096

inference will likely be inaccurate. It is not currently possible to easily assess the likelihood of such violations without a structured data field for clinical trial enrollment status in the EHR.

Capturing clinical trial enrollment status also has utility in addressing FDA real-world evidence guidelines. Current FDA draft guidance on submitting real-world evidence for drugs and biologics indicates that the real-world data sources that were used to derive real-world evidence should be described.[25] While clinical trial and point-of-care encounter data are all derived from the EHR, the distinctions between the two types of data outlined above highlight the importance of distinguishing the specific source of data (clinical trial *v* point-of-care encounter) within the EHR system. Attention to this detail by informaticists and statisticians is one component that can improve the transparency and reproducibility of real-world data analyses.

Apart from evaluating the potential implications of an analysis that aggregates two sources of EHR data, the secondary analysis of clinical trial data requires careful consideration in and of itself.[26,27] Issues of bias and multiplicity arise, highlighting the importance of an analyst being aware of whether an EHR-based cohort includes patients who were enrolled on a clinical trial. For patients who were enrolled on a clinical trial, the analyst may need further information, such as the randomization schema, stratification factors, and the study schedule, to properly conduct and interpret the analysis.[28]

Without the routine capture of clinical trial enrollment status, formal evaluation of the appropriateness of the established methods for analyzing EHR data, adherence to current FDA draft guidance, and appropriate consideration of the implications of a secondary analysis of clinical trial data are not possible. A deeper investigation into the implications of aggregating clinical trial and point-of-care encounter data with respect to their effect on statistical inference is beyond the scope of this paper; it would depend on the specific research question being addressed using the EHR-based cohort, the proportion of patients in the real-world EHR cohort who were ever enrolled on a clinical trial, the timing of their clinical trial enrollment with respect to the time period of interest, as well as specific characteristics of the clinical trial itself (eg, randomization schema, stratification factors, and visit schedule). Our intention is to highlight the importance of capturing clinical trial enrollment status as a key first step in being able to begin to fully quantify the implications of aggregating clinical trial and point-of-care encounter data.

In conclusion, EHRs are a complex but rich source of real-world data. With the development of EHR data standards and data governance frameworks to generate reproducible processes for data-intensive research, there is an opportunity to leverage custom EHR fields to ensure uniform documentation of clinical trial enrollment status in the EHR.[29,30] Being able to easily access this information will allow investigators to consider the analytic implications of aggregating data from clinical trial and point-of-care encounters, ensure compliance with clinical trial agreements, and sufficiently assess the quality and appropriateness of real-world evidence generated from real-world data.

AFFILIATIONS

[1]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY
[2]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY

CORRESPONDING AUTHOR

Jessica A. Lavery, MS, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 485 Lexington Avenue, 2nd Floor, New York, NY 10017; Twitter: @jessicalavs; email: laveryj@mskcc.org.

SUPPORT

Supported by Memorial Sloan Kettering Cancer Center by a core grant from the National Cancer Institute (P30 CA008748).

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.
Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Jessica A. Lavery
Research Funding: American Association for Cancer Research (AACR) Project Genomics Evidence Neoplasia Information Exchange Biopharma Collaborative (GENIE BPC)

Margaret Callahan
Consulting or Advisory Role: AstraZeneca, Moderna Therapeutics, Merck, Immunocore
Research Funding: Bristol-Myers Squibb
Other Relationships: Clinical Care Options, Potomac Center for Medical Education, Bristol-Myers Squibb

Katherine S. Panageas
Research Funding: American Association for Cancer Research (AACR) Project Genomics Evidence Neoplasia Information Exchange Biopharma Collaborative (GENIE BPC)
Stock and Other Ownership Interests: Catalyst Biotech, Dynavax Tech, Sunesis Pharmaceuticals, Viking Therapeutics

No other potential conflicts of interest were reported.

AUTHOR CONTRIBUTIONS

Conception and design: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. U.S. Food and Drug Administration: 21 USC §355g Utilizing Real World Evidence. Washington D.C., U.S. Food and Drug Administration, 2018
2. U.S. Food and Drug Administration: Center for Devices and Radiological Health, Center for Biologics Evaluation and Research: Use of real-world evidence to support regulatory decision-making for medical devices. https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf
3. Office of the Commissioner: Real-World Evidence. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence, 2020
4. Corrigan-Curay J, Sacks L, Woodcock J: Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA 320:867-868, 2018
5. Casey JA, Schwartz BS, Stewart WF, et al: Using electronic health records for population health research: A review of methods and applications. Annu Rev Public Health 37:61-81, 2016
6. Hutchins LF, Unger JM, Crowley JJ, et al: Underrepresentation of patients 65 years of age or older in cancer-treatment trials. N Engl J Med 341:2061-2067, 1999
7. Elting LS, Cooksley C, Bekele BN, et al: Generalizability of cancer clinical trial results: Prognostic differences between participants and nonparticipants. Cancer 106:2452-2458, 2006
8. Hersh WR, Weiner MG, Embi PJ, et al: Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 51:S30-S37, 2013
9. Oken MM, Creech RH, Tormey DC, et al: Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 5:649-655, 1982
10. Taha B, Winston G, Tosi U, et al: Missing diversity in brain tumor trials. Neurooncol Adv 2:vdaa059, 2020
11. Rencsok EM, Bazzi LA, McKay RR, et al: Diversity of enrollment in prostate cancer clinical trials: Current status and future directions. Cancer Epidemiol Biomarkers Prev 29:1374-1380, 2020
12. Nazha B, Mishra M, Pentz R, et al: Enrollment of racial minorities in clinical trials: Old problem assumes new urgency in the age of immunotherapy. Am Soc Clin Oncol Educ Book 39:3-10, 2019
13. Ludmir EB, Mainwaring W, Lin TA, et al: Factors associated with age disparities among cancer clinical trial participants. JAMA Oncol 5:1769-1773, 2019
14. Loree JM, Anand S, Dasari A, et al: Disparity of race reporting and representation in clinical trials leading to cancer drug approvals from 2008 to 2018. JAMA Oncol 5:e191870, 2019
15. Coakley M, Fadiran EO, Parrish LJ, et al: Dialogues on diversifying clinical trials: Successful strategies for engaging women and minorities in clinical trials. J Womens Health 21:713-716, 2012
16. Unger JM, Hershman DL, Osarogiagbon RU, et al: Representativeness of black patients in cancer clinical trials sponsored by the National Cancer Institute compared with pharmaceutical companies. JNCI Cancer Spectr 4:pkaa034, 2020
17. Phelan M, Bhavsar NA, Goldstein BA: Illustrating informed presence bias in electronic health records data: How patient interactions with a health system can impact inference. EGEMS (Wash DC) 5:22, 2017
18. Weiskopf NG, Rusanov A, Weng C: Sick patients have more data: The non-random completeness of electronic health records. AMIA Annu Symp Proc 2013:1472-1477, 2013
19. Goldstein BA, Bhavsar NA, Phelan M, et al: Controlling for informed presence bias due to the number of health encounters in an electronic health record. Am J Epidemiol 184:847-855, 2016
20. Goldstein BA, Phelan M, Pagidipati NJ, et al: How and when informative visit processes can bias inference when using electronic health records data for clinical research. J Am Med Inform Assoc 26:1609-1617, 2019
21. Song X, Mu X, Sun L: Regression analysis of longitudinal data with time-dependent covariates and informative observation times. Scand J Stat 39:248-258, 2012
22. Lin H, Scharfstein DO, Rosenheck RA: Analysis of longitudinal data with irregular, outcome-dependent follow-up. J R Stat Soc Ser B (Stat Methodol) 66:791-813, 2004
23. Chen Y, Wang J, Chubak J, et al: Inflation of type I error rates due to differential misclassification in EHR-derived outcomes: Empirical illustration using breast cancer recurrence. Pharmacoepidemiol Drug Saf 28:264-268, 2019
24. Fitzmaurice GM, Lipsitz SR, Ibrahim JG, et al: Estimation in regression models for longitudinal binary data with outcome-dependent follow-up. Biostatistics 7:469-485, 2006
25. U.S. Food and Drug Administration: Submitting documents using real-world data and real-world evidence to FDA for drugs and biologics: guidance for industry. Docket no. 2019-09529. https://www.fda.gov/regulatory-information/search-fdaguidance-documents/submitting-documents-using-real-world-dataand-real-world-evidence-fda-drugs-and-biologics-guidance
26. Clarke SP, Cossette S: Secondary analysis: Theoretical, methodological, and practical considerations. Pharmacoepidemiol Drug Saf 32:109-129, 2000
27. Marler JR: Secondary analysis of clinical trials: A cautionary note. Prog Cardiovasc Dis 54:335-337, 2012
28. Hollis S, Fletcher C, Lynn F, et al: Best practice for analysis of shared clinical trial data. BMC Med Res Methodol 16:76, 2016 (suppl 1)
29. Bertagnolli MM, Anderson B, Norsworthy K, et al: Status update on data required to build a learning health system. J Clin Oncol 38:1602-1607, 2020
30. Bertagnolli MM, Anderson B, Quina A, et al: The electronic health record as a clinical trials tool: Opportunities and challenges. Clin Trials 17:237-242, 2020