

CASE AND COMMENTARY: PEER-REVIEWED ARTICLE

Whom Should We Regard as Responsible for Health Record Inaccuracies That Hinder Population-Based Fact Finding?

Kathleen M. Akgün, MD, MS and Shelli L. Feder, PhD, APRN

Abstract

Electronic health records (EHRs) have revolutionized the scale, speed, and granularity at which health data can be collated and summarized for epidemiologic purposes. However, population-level analyses of patient-level data are only as reliable as the accuracy or completeness of patient reporting, clinician data entry, and how systems are programmed. This commentary on a case argues that responsibility for the validity of EHR data should be shared among key stakeholders, including patients. This commentary also proposes models for EHR data inquiry, data entry, and review processes that incorporate roles of community partners, frontline clinicians, and health science experts.

Case

T is a 43-year-old woman who has well-controlled asthma and visits Dr A for an annual checkup from a rural part of the state. Dr A reviews T's electronic health record (EHR), noting no documentation of COVID-19 vaccination. Dr A remembers results from a national study that rurality was associated with decreased odds for vaccination and asks T about her reasons for not getting any shots. "I did get 2 shots last year, but I didn't get them here. That's probably why you're not seeing them. I plan to get a booster as soon it's available."

Dr A wonders why many other patients' EHRs contain incomplete or inaccurate information and why. "Not only does this affect how I plan my time with my patients, but poor-quality data hinders epidemiological surveillance and tracking of population-level vaccine uptake. Wasn't the harrowing transition we've all just made to EHRs supposed to eliminate problems like this?"

Commentary

EHRs have become ubiquitous in clinical care in economically developed settings, including much of the United States. We review the evolution of these tools from their use in clinical care and billing to population-level health studies and pragmatic clinical trials. We then identify sources of biases and inaccuracies in EHR data and consider the ethics and consequences of using EHR-based data in research. Finally, we discuss the responsibilities of maintaining EHR data accuracy and propose ways to promote

engagement among key stakeholders (eg, health care systems and payers, EHR developers, patients, clinicians, and researchers) in building an accurate, representative EHR. We illustrate these issues in a study of vaccination outcomes for patients enrolled in rural and urban health systems.

Brief History of the EHR

Clinical information systems were the predecessors of the modern-day EHR and were first utilized in single clinical sites as early as the 1960s.¹ Efforts to transform health record keeping with EHR technology were promulgated with the development of the Department of Veteran Affairs’ VistA and Computerized Patient Record System in the 1970s and 1980s.¹ Since then, health care organizations have been incentivized—and eventually mandated—to transition from paper charts to EHRs: first with the passage of the Health Information Technology for Economic and Clinical Health Act (HITECH) as part of the 2009 American Recovery and Reinvestment Act. Through incentive payments, HITECH sought to maximize EHRs’ potential to improve patient safety (including by minimizing illegible handwriting and standardizing health data collection, entry, and reporting) and build EHRs into the scaffolding of health care delivery.² With the passage of the 21st Century Cures Act in 2016, the HITECH regulations were expanded to require the use of EHRs.³

In addition to improving patient safety, the EHR has been a boon for researchers. The proliferation of the EHR has enhanced the analyzability of clinical encounters through typed clinical note documentation accompanied by structured billing codes. Patient demographics are routinely collected and grouped by researchers to estimate the prevalence of health behaviors, determine at-risk patient populations, identify health disparities, and screen potential participants for clinical trial enrollment.⁴ Epidemiologic studies based on EHR data are critical for measuring population-level outcomes, including hospitalization or death. EHRs also allow for standardized data collection with the use and dissemination of templates for clinical notes, transforming unstructured text into structured data elements that can be easily extracted from the EHR.⁵

Sources of Error in EHR Data

Although the EHR has many benefits, limitations of data collection can affect data quality and bias research findings. Common domains of EHR data quality for research purposes include accuracy, completeness, consistency, credibility, and timeliness (see Table 1).⁶

| Table 1. Quality Domains for Data in the Electronic Health Record | | |
|---|---|---|
| Quality domain | Definition | Example from vaccination scenario |
| Accuracy | Extent to which data in EHR is valid “representation of the real-world value” | Vaccine receipt would indicate patient was vaccinated |
| Completeness | Frequency of missing data and patterns | Vaccine status entered across rural, suburban settings independent of race or ethnicity or payer |
| Consistency | Predictability of data collected in different systems or databases | Vaccine data across systems (eg, Medicare, VHA) uses comparable variable definitions |
| Credibility | Plausibility, believability of data | Dates of vaccination begin after vaccine available (12/14/20) and do not exceed present day |
| Timeliness | Lapse of time between data entry and ability to measure variable of interest | Vaccine data collated by week following launch of vaccination campaign to report vaccination trends |

Domains and definitions adapted from Feder.⁶
Abbreviations: EHR, electronic health record; VA, Veterans Health Administration.

Like any system data, EHR data are only as strong as their inputs.⁴ As the case illustrates, barriers to accurate vaccination documentation begin with the patient-clinician interaction. Inaccurate data inputs may result from poor patient-clinician communication and a lack of patient understanding or opportunity to ask for clarification about the questions being asked.¹ When reviewing their EHR, patients not uncommonly perceive mistakes.⁷ Among 22 889 US participants of the OpenNotes study who read their notes and completed error questions, 4830 (21%) identified an EHR mistake, 2043 (42%) of whom reported that the mistake was serious.⁸ In addition to mistakes, time constraints could lead to inaccurate EHR data. Additional sources of inaccurate data, which are intrinsic to clinical care and not unique to the EHR, include patient preferences regarding disclosure of sensitive information, the receipt of out-of-network care or at care centers utilizing different EHR systems, asynchronous data entry, or clinician omission (see Table 2).^{9,10,11,12,13}

Table 2. Sources of Electronic Health Record Errors and Examples of Mechanisms of Missingness

| Errors in EHR data input | Examples in clinical EHR data |
|---|--|
| Data entry errors | Incorrect medications, laboratories, or vital signs |
| Cut and paste errors | Dated health information; lack of updated conditions, medications, or procedures |
| Chart management errors | Charting information in wrong EHR |
| Chart completion errors | Delayed or incomplete chart documentation |
| Incorrect order entry | “Sound-alike” medication prescribing entered in error but could be folded into other clinicians’ documentation for the patient |
| Submechanisms of missingness | Examples in clinical EHR data |
| Data elements | Exposures, confounding variables, outcomes, relevant variables |
| Time points | Baseline, varying time for follow-up |
| Likelihood of measurement during clinical encounter | Blood draws not done at all primary care visits; telehealth visits |
| Outside care | Out-of-network subspecialty care |
| Changing clinical practice standards | Screening leads to more incidental findings |

Text adapted from Haneuse et al⁹ and Bowman.¹⁰
Abbreviation: EHR, electronic health record.

Data entry tools like drop-down menus, copy-paste features, and automatic laboratory value entries can enhance efficiency but can also contribute to system-level errors and omissions, perpetuating biases and inequities.^{14,15,16} In the case example, and as reported in a recent study,¹⁷ rurality was associated with decreased COVID-19 vaccination. However, data entry was incomplete, confounded by human factors that might be exacerbated in rural settings. Inaccurate or incomplete data entry can contribute to sweeping but biased generalizations about treatment disparities, which very well could exist, but are incompletely ascertained due to missing data.^{4,5,10,18,19}

Effects of Regulations, Missingness, and Representativeness on Research
Individual patient privacy and agency could be at odds with the need for high-quality population-level health data. Governing bodies overseeing research activities provide one layer of protection for patients by minimizing privacy risks and ensuring data security and investigator integrity and compliance with established rules of behavior in

research. Deidentification of EHR data (in compliance with the Health Insurance Portability and Accountability Act of 1996 Privacy Rule) is standard practice when collating system-wide EHR data for research purposes.²⁰ However, as EHR data is increasingly stripped of identifiers for subsequent public sharing and analyses, designating research based on such data as “not human subjects research” could jeopardize this oversight, with incompletely understood consequences for population-level studies and inferences. In addition, patient opt-out features can perpetuate biases in the final data based on who is or is not choosing this option.^{11,12,13}

Moreover, when epidemiologists analyze EHR data for population health impacts, they often encounter missingness, wherein data for variables of interest are unavailable for each included observation for various reasons. Relying solely on quantitative data inputted by clinical teams thus could limit conclusions, telling incomplete stories. Yet solely focusing on solutions like outreach does not improve rural vaccination rates in the setting of incomplete measurement. Qualitative and mixed methods studies could provide important context for EHR data capture and assist researchers in confirming and contrasting findings derived from the EHR. For example, patient interviews could identify barriers to and mechanisms of vaccine uptake and how and where patients are getting vaccinated. Studies assessing clinician perspectives of EHR data entry options and workflows could also uncover reasons for missing or erroneous patient vaccination history data.

Finally, a fundamental vulnerability of collated, population-level data is whether the included sample is indeed representative of the intended source population. Systemic biases, community engagement, and intersectionality across minoritized groups, sex or gender identity groups, and **race or ethnicity groups** could influence the visibility of specific populations in EHR data and the resultant output. Population health, health equity, and policy experts have begun to identify sources of and strategies for dealing with bias in the EHR.²¹ Patient-reported outcomes and health and general literacy are key areas that can be targeted to reduce bias in EHR-based studies.²²

Stakeholders’ Responsibilities

Just as the causes of EHR inaccuracies are multifaceted, so are responsibilities for ensuring EHR data fidelity, which are shared among key stakeholders: health care systems, vendors, clinicians, patients, and researchers using the data.²³ Vendors and health care systems remain accountable to the general public for EHR functionality, usability, and accuracy. Clinicians must remain engaged with health care systems to ensure their data entry maximizes efficient use of EHR data for clinical documentation, review, and research purposes.¹⁰ Reframing the roles of patients as partners in data generation rather than as study subjects or participants could motivate patients to contribute to solutions to high-quality health data collection.⁴ Specifically, inviting patients to identify strategies for EHR data entry that most align with their preferences could enhance the completeness, credibility, and timeliness of their data. For example, because patient consent is not routinely obtained before using EHR data, patients may be unaware of how third parties could use their data, and reidentification of patients might be easier in certain types of studies, including genetics or rare disease studies or those using diagnostic imaging or clinical text notes.²⁴ “Opt-out” features could return some agency to patients over their EHR data use in research but is not routinely done across health systems. However, this patient-centered approach could, itself, contribute to biased data, as patients who participate in such efforts might not reflect the overall patient population of interest.

Finally, data scientists and researchers are responsible for ensuring that high-quality EHR data are appropriately analyzed in population-level analyses. They should also remain vigilant in **recognizing biased results** of the analyses performed²⁵ by explicitly addressing the 5 domains of data quality—accuracy, completeness, consistency, credibility, and timeliness—throughout the research process. Research teams also must defend against data breaches and must routinely address limitations and biases of measured data in study manuscripts and other output.⁶

Accountable EHR Data Use

How can we improve and innovate EHRs to enhance the accuracy of vaccine documentation and other data? An integrated US health system is an attractive answer and could serve as the platform for clinical data integration, but it is unlikely to gain favor in the current politically polarized environment. The Affordable Care Act (ACA) of 2010 was the first reform to the US health care system in a generation and offered a mechanism to build towards integration by leveraging Medicaid infrastructure for beneficiary data collection,²⁶ but state-level single-payer policies have failed to gain political traction.²⁷ The ACA remains a lightning rod for fiscally conservative policy makers, and, although public support appears to have grown over time, an overhaul to the US health system is unlikely to be successful in the coming decade.

Given this landscape, individuals and organizations should undertake to improve EHR data quality.

- Health information vendors should continually reassess EHR data collection tools and interfaces with patients, clinicians, and data scientists to optimize their product's usability.
- Clinicians must advocate for meaningful approaches to EHR use to support clinical care, rather than simply plodding through required data fields for billing purposes.
- Organizations should consider expanding, standardizing, and integrating the data collected. Although patient-entered EHR data can be an attractive option to increase accuracy of patient data while also empowering patients to control their health care narrative firsthand, it could still result in data skewed towards patients who are computer literate and have access to broadband internet services. Vaccination and medication use data collection could be standardized by automating linkages between pharmacy manufacturing lots or similar measures and EHRs, thereby improving care tracking. Government- and private, nonprofit-supported applications, such as the Immunization Information System and health information exchanges, also hold promise for integrating and harmonizing health information—from vaccinations to medications, subspecialist evaluations, and clinical testing results—across participating regions.^{6,7,8,9}

Limitations

Despite attention to data quality, there exist no standard methods for assessing EHR data quality.²⁸ We acknowledge that the aforementioned proposed solutions for improving the accuracy of EHR data are practical only for measurable processes and outcomes of care. Other important aspects of care, such as patient-clinician demographic concordance or general communication styles, are not currently captured in structured data available in large health care system EHRs. Alternate care delivery models (minute clinics, telehealth, concierge medicine) might also have uncertain impacts on EHR completeness going forward. As conventions shift in health care delivery

and its documentation, key stakeholders are necessary to determine how best to tackle pressing health priorities of the population.

Conclusion

EHR use has revolutionized health information collection and analysis. This growth has led to opportunities to generate important reports about the health of hundreds of millions of people practically in real time. Steadfast commitment to high-quality data collection and reporting is necessary for all parties along the pathway of data generation: from EHR developers, programmers, and vendors to patients, clinicians, and epidemiologists. Pulling back the curtain on how each of these groups generate and interact with EHR data is imperative to assure measurement of accurate population-level health outcomes.

References

1. Atherton J. Development of the electronic health record. *Virtual Mentor*. 2011;13(3):186-189.
2. Trout KE, Chen LW, Wilson FA, Tak HJ, Palm D. The impact of meaningful use and electronic health records on hospital patient safety. *Int J Environ Res Public Health*. 2022;19(19):12525.
3. Phelan D, Gottlieb D, Mandel JC, et al. Beyond compliance with the 21st Century Cures Act Rule: a patient controlled electronic health information export application programming interface. *J Am Med Inform Assoc*. 2024;31(4):901-909.
4. Hollister B, Bonham VL. Should electronic health record-derived social and behavioral data be used in precision medicine research? *AMA J Ethics*. 2018;20(9):E873-E880.
5. Vuong K, Ivers R, Hall Dykgraaf S, Nixon M, Roberts G, Liaw ST. Ethical considerations regarding the use of pooled data from electronic health records in general practice. *Aust J Gen Pract*. 2022;51(7):537-540.
6. Feder SL. Data quality in electronic health records research: quality domains and assessment methods. *West J Nurs Res*. 2018;40(5):753-766.
7. Klinger EV, Carlini SV, Gonzalez I, et al. Accuracy of race, ethnicity, and language preference in an electronic health record. *J Gen Intern Med*. 2015;30(6):719-723.
8. Bell SK, Delbanco T, Elmore JG, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. *JAMA Netw Open*. 2020;3(6):e205867.
9. Haneuse S, Arterburn D, Daniels MJ. Assessing missing data assumptions in EHR-based studies: a complex and underappreciated task. *JAMA Netw Open*. 2021;4(2):e210184.
10. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manag*. 2013;10(fall):1c.
11. El Emam K, Jonker E, Moher E, Arbuckle L. A review of evidence on consent bias in research. *Am J Bioeth*. 2013;13(4):42-44.
12. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. *BMJ*. 2009;338:b866.
13. de Man Y, Wieland-Jorna Y, Torensma B, et al. Opt-in and opt-out consent procedures for the reuse of routinely recorded health data in scientific research and their consequences for consent rate and consent bias: systematic review. *J Med Internet Res*. 2023;25:e42131.

14. Boyd AD, Gonzalez-Guarda R, Lawrence K, et al. Equity and bias in electronic health records data. *Contemp Clin Trials*. 2023;130:107238.
15. Rozier MD, Patel KK, Cross DA. Electronic health records as biased tools or tools against bias: a conceptual model. *Milbank Q*. 2022;100(1):134-150.
16. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible sources of bias in primary care electronic health record data use and reuse. *J Med Internet Res*. 2018;20(5):e185.
17. Bernstein E, DeRycke EC, Han L, et al. Racial, ethnic, and rural disparities in US veteran COVID-19 vaccine rates. *AJPM Focus*. 2023;2(3):100094.
18. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018;178(11):1544-1547.
19. Getzen E, Ungar L, Mowery D, Jiang X, Long Q. Mining for equitable health: assessing the impact of missing data in electronic health records. *J Biomed Inform*. 2023;139:104269.
20. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. US Department of Health and Human Services. Reviewed October 25, 2022. Accessed February 22, 2024. <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>
21. Sun M, Oliwa T, Peek ME, Tung EL. Negative patient descriptors: documenting racial bias in the electronic health record. *Health Aff (Millwood)*. 2022;41(2):203-211.
22. Boyd AD, Gonzalez-Guarda R, Lawrence K, et al. Potential bias and lack of generalizability in electronic health record data: reflections on health equity from the National Institutes of Health Pragmatic Trials Collaboratory. *J Am Med Inform Assoc*. 2023;30(9):1561-1566.
23. Sittig DF, Singh H. Rights and responsibilities of users of electronic health records. *CMAJ*. 2012;184(13):1479-1483.
24. Holmes JH, Beinlich J, Boland MR, et al. Why is the electronic health record so challenging for research and clinical care? *Methods Inf Med*. 2021;60(1/2):32-48.
25. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol*. 2021;21(1):234.
26. Stulberg D. The Patient Protection and Affordable Care Act and reproductive health: harnessing data to improve care. *J Health Polit Policy Law*. 2013;38(2):441-456.
27. Sparer MS. States as policy laboratories: the politics of state-based single-payer proposals. *Am J Public Health*. 2019;109(11):1511-1514.
28. Lewis AE, Weiskopf N, Abrams ZB, et al. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc*. 2023;30(10):1730-1740.

Kathleen M. Akgün, MD, MS is an associate professor of medicine at VA-Connecticut Health Care System (VACHS) and the Yale School of Medicine in New Haven, Connecticut. A health services researcher focused on health care delivery and outcomes and advocacy, she also serves as co-director of the West Haven Network of Dedicated Enrollment Sites and as the director of the medical intensive care unit and co-chair of the Clinical Ethics Committee at VACHS.

Shelli L. Feder, PhD, APRN is an assistant professor at the Yale School of Nursing in New Haven, Connecticut. She is also an associate program director for the Yale National Clinician Scholars Program, an affiliate investigator at the PRIME Center, and an organizational health services researcher. She has over a decade of clinical experience as an advanced practice nurse in hospice, palliative care, and cardiovascular settings. Dr Feder's research program aims to create innovative models of care that improve access to high-quality, timely palliative care for people with cardiopulmonary conditions.

Editor's Note

The case to which this commentary is a response was developed by the editorial staff.

Citation

AMA J Ethics. 2025;27(1):E6-13.

DOI

10.1001/amajethics.2025.6.

Acknowledgements

Dr Feder and Dr Akgün acknowledge support from grant 1R56HL16652301A1 from the National Heart, Lung, and Blood Institute (primary investigator: Dr Feder) and from grant 13407 from the Center of Innovation of the Veterans Health Administration.

Conflict of Interest Disclosure

Dr Akgün has a contract with the Department of Veterans Affairs Central Office. Dr Feder disclosed no conflicts of interest.

The people and events in this case are fictional. Resemblance to real events or to names of people, living or dead, is entirely coincidental. The viewpoints expressed in this article are those of the author(s) and do not necessarily reflect the views and policies of the AMA.