

AN INTERDISCIPLINARY APPROACH TO REDUCING ERRORS IN EXTRACTED ELECTRONIC HEALTH RECORD DATA FOR RESEARCH

Posted on March 15, 2021 by Matthew

Category: [Spring 2021](#)

By Neelkamal Soares, MD; Sorabh Singhal; Casey Kloosterman, PharmD; and Teresa Bailey, PharmD.

ABSTRACT

Erroneous electronic health record (EHR) data capture is a barrier to preserving data integrity. We assessed the impact of an interdisciplinary process in minimizing EHR data loss from prescription orders. We implemented a three-step approach to reduce data loss due to missing medication doses:

- Step 1—A data analyst updated the request code to optimize data capture;
- Step 2—A pharmacist and physician identified variations in EHR prescription workflows; and
- Step 3—The clinician team determined daily doses for patients with multiple prescriptions in the same encounter.

The initial report contained 1421 prescriptions, with 377 (26.5 percent) missing dosages. Missing dosages reduced to 361 (26.3 percent) prescriptions following Step 1, and twenty-three (1.7 percent) records after Step 2. After Step 3, 1210 prescriptions remained, including 16 (1.3 percent) prescriptions missing doses. Prescription data is susceptible to missing values due to multiple data capture workflows. Our approach minimized data loss, improving its validity in retrospective research.

Keywords: Data integrity, electronic health record, data abstraction, prescription.

Background

By 2015 almost 96 percent of all non-federal acute-care hospitals possessed certified health information technology (HIT), and 86 percent of office-based practices had adopted an electronic health record (EHR).^{1,2} In addition to its primary function as a documentation tool used to store the clinical narrative, the EHR is a rich source of large amounts of information, including pharmacy, laboratory, financial, and patient demographic data. This gives the EHR utility for public health practitioners and researchers.

This “secondary use” of EHR data³ is especially powerful when combined with data from other sources, e.g., census tracts, surveys, morbidity and mortality data.⁴

From the technical perspective, data quality in EHR-based studies remains a challenge for investigators with respect to data completeness, standardization and the lack of use of controlled vocabularies.³⁻⁶ The process can be time-intensive, as it involves sifting through raw EHR data that is often disorganized and replete with uncodified variables. Furthermore, data point duplication errors often require specialized personnel with knowledge of the EHR data structure to run meaningful queries.⁷ Additionally, limitations include incomplete, inconsistent, or inaccurate data.⁵ Other barriers

include the presence of multiple disconnected data capture systems, the lack of timely access to EHR data, interoperability issues between EHR vendors,⁶⁻⁸ and changing data standards and practice guidelines over time.⁹

We had experienced some of these limitations first-hand while evaluating prescribing patterns of second-generation antipsychotic (SGA) medications in children and adolescents. Here we assess the impact of an interdisciplinary, stepwise approach to improve the quality of data by reducing missingness of data during an EHR data abstraction.

Methods

The Institutional Review Board of the primary author's institution approved the study with exempt status. We conducted a retrospective study using EHR data to evaluate SGA prescribing patterns in children and adolescents between June 1, 2017, and December 31, 2018. This included data collected from the Epic EHR (Epic Corp, Verona, WI). We utilized data from three health systems in and around a small Midwestern city in the United States, encompassing a Federally-Qualified Health Center, an academic teaching center, and a private community-based healthcare system. The interdisciplinary team involved a physician, a pharmacist, a biostatistician, and a data analyst who deidentified the data extracted from the EHR.

The requested data included all encounters during the specified time period for all patients under the age of 18 at time of visit who received an SGA prescription at any time during the specified period. The data analyst created a Structured Query Language (SQL) encounter-level report utilizing an abstraction from Epic's Clarity data warehouse. The Clarity database stores hospital-specific data, which analysts query to create complex, data-intensive reports. The analyst masked the patient medical record numbers (MRN) utilizing the community identification number (CID) within the Epic software, which deidentified the data while allowing the research team to group hospital encounters by individual patients. The query design utilized the medication list to pull medication names, frequencies, and doses.

Upon discovering a significant number of encounters with missing medication doses, we were unable to determine if the SGA encounter records without doses were different from the SGA encounter records with doses; we therefore could not rule out the potential of a confounding difference between the two subsets. This was our impetus to design a stepwise approach to minimize missing SGA medication doses:

Step 1: The data analyst updated the data request code to optimize data capture. This was necessary because the first query included all medication orders, including the cancelled and discontinued orders. As researchers, we had incorrectly assumed these orders would not be added. In order to be able to exclude these records from analysis, the analyst edited the query design criteria to only include active and completed medication orders.

Step 2: the pharmacist and physician on the team reviewed the EHR prescription workflows to identify and account for variations in prescribing methods.

Step 3: if there were multiple prescription records for the same patient on the same day, the clinician team determined daily medication doses in one of three ways: (1) duplicative orders were deleted; (2) multiple prescriptions with directions for different times of the day were added; or (3) records of multiple SGA prescriptions were left on multiple rows.

Results

A total of 346 individual patients met the inclusion criteria, representing 1,421 individual SGA prescriptions. The initial query yielded 377 (26.5 percent) entries with missing data. In discussions with the data analyst, we discovered “abandoned” orders in the data set, where a clinician initiated and then discarded a refill or new-prescription order. The data analyst thus required the database to select against “order status = cancelled”. This Step 1 intervention reduced the total SGA prescriptions to 1374, with 361 missing dosages (26.3 percent).

Given the continued missing data, we implemented Step 2, with review of clinician orders using the computerized physician order entry (CPOE) system. We found that clinicians were able to use “free text” fields to bypass the discrete dose data field. We then populated medication dose fields utilizing the free-text instructions. The challenges encountered are listed in [Table 1](#). This reduced the total number of prescription orders with missing dosage values to 23 (1.7 percent). The remaining unreconciled orders after Step 2 were “blank” free-text fields without dose instructions.

The clinician team then further analyzed data for patients who received multiple SGA prescriptions on the same day, and 373 prescription records met the criteria for Step 3 analysis. The data represented 175 unique encounters (of which 34 encounters had two SGAs prescribed), that contained prescriptions that were interpretable. Once the doses were interpretable, the final dataset contained 1,210 prescriptions, and only 16 (1.3 percent) had a missing dose. The results are summarized in [Figure 1](#).

Discussion

Despite utilizing standardized approaches to request specific data points to create an encounter-level report, we discovered a high level of missing data that would have impacted the validity of our intended analysis. Through a systematic, dialectic approach including all members of the team, we initially addressed EHR data request issues of which we were unaware. Even after optimizing the data request code, the persistent missing data required a “manual” review via specific expertise of investigators who were familiar with clinician prescribing workflows to reduce errors to an acceptable, though non-zero, level. We particularly emphasize the amount of variability in prescription ordering workflows within the sample; clinicians often tailor established EHR workflows to their preferences and personalize their approaches to using EHR features.¹⁰ Due to this

widespread practice, it is critical to recognize and capture this variation when evaluating data sets for accuracy and validity.

Medication prescribing orders are particularly vulnerable to errors related to the common structure of these orders. CPOE medication fields consist of three main regions: the prescription or "Rx line", which contains medication name, strength amount and units, and drug form; the patient instructions or "Sig line", which includes dose units, route of administration, frequency and duration in days; and a "Special Instructions line", which includes detailed instructions and clarifications. It is the latter that generates the free-text field.¹¹ In the clinical setting, prescriptions are used essentially as a communication tool between clinicians and pharmacists. Thus, discrepant or unclear instructions are resolved using either the pharmacists' knowledge or follow-up communication between the two providers, a system that leads to inefficiencies.¹¹

We found variable free-text data that led to analysis discrepancy after the first intervention in up to 24.7 percent of the data points, which was less than Palchuk's study,¹¹ but still required Step 2 of the intervention. Multiple strategies have been proposed to draw meaningful information from indiscrete data, including using established databases and rule-based systems¹² and machine learning analyses.¹³ Instead of retrospective approaches for data extraction and "cleaning", we stress the importance of proactive approaches that educate clinicians on the importance of following standard CPOE prescription workflows in the EHR. One strategy is to optimize prescribing workflows through a user-centric CPOE system, which includes using drop-down menus for drug dosing options. A system that prioritizes selecting commonly found medication doses and routes could prevent clinicians from bypassing discrete entry in favor of free-text options.¹⁴ Another way is to implement clinical decision support (CDS) processes that can perform checks on free-text orders¹¹ though there are limits to both the deployment and functioning of CDS.¹⁵ Until such time that a singular or combination of strategies are used to automate mitigation of errors in prescribing input, we believe that research teams should utilize an interdisciplinary approach with content-expert team members to contribute to the process. As evident from our method, although data analysts are integral to the data pull and serve as the "honest broker" for research protocols,¹⁶ they are often unable to account for the nuances in clinical workflows that impact data quality. This underscores the importance of having a clinician on the investigator team who is intimately familiar with the clinical process where the data are generated and collected. Similarly, clinical investigators often have limited knowledge of how data is structured within the EHR, and the analyst is needed to implement changes, as was the case in our Step 1 with the cancelled orders. An interdisciplinary approach has grown in healthcare research¹⁷ and has particular value in EHR-based research involving informaticians and clinicians. However, not every institution has access to a clinical informatician, hence it is critical to

build teams with interdisciplinary content expertise.

The limitations of our study included data from a single EHR covering three organizations in a single geographic area, and hence results may not be generalizable to other regions or EHRs. However, replication with larger sets of data points and collating data from multiple sources can extend our model and test its value. This was also a cross sectional approach, looking at a defined set of prescriptions as part of larger project. It would be important to look how pervasive the errors are when expanded over a longer period of time. Furthermore, system workflows change over time, so any follow-up studies must account for practice or guidelines changes.¹⁸ Finally, we did not compare a human review to machine learning or other automated methods.

Conclusions

An interdisciplinary, stepwise approach to reducing EHR-based prescription data loss can be effective. Prescription data is particularly susceptible to missing values given the multiple methods of data entry, including the use of free-text fields, sometimes replete with medical abbreviations and typographical errors. We advocate that EHR developers explore ways to simplify interfaces and options with an eye to reduce human error opportunities, so that clinicians can document in the EHR using the method that best matches their workflow. Multiple methods of data entry can complicate the secondary use of EHR-data for research. On research teams involving EHR data, input from clinicians and other content specialists is critical to optimizing the completeness, accuracy and validity of an EHR-data abstraction, since quality of studies that use EHR data are only as good as the data that is available.

Disclosures and Acknowledgements

The authors have no conflicts of interest to disclose.

The authors attest that we have not submitted this manuscript for publication elsewhere.

The authors would like to acknowledge Dr. Phillip Kroth for his input and guidance, Joseph Billian for statistical support, and Melissa Sherfield and Melissa Medina for their database and analytical support.

Author Biographies

Neelkamal Soares, MD (Neelkamal.soares@med.wmich.edu) is Professor of Pediatric & Adolescent Medicine, Western Michigan University Homer Stryker M.D. School of Medicine, Kalamazoo, Michigan.

Sorabh Singhal (Sorabh.singhal@med.wmich.edu) is a Medical Student, Western Michigan University Homer Stryker M.D. School of Medicine, Kalamazoo, Michigan.

Casey Kloosterman, PharmD, (kloostc2@ferris.edu) is a doctoral graduate, Ferris State University College of Pharmacy, Big Rapids, Michigan.

Teresa Bailey, PharmD, (teresabailey@ferris.edu) is Professor of Pharmacy Practice, Ferris State University College of Pharmacy, Big Rapids, Michigan.

REFERENCES

1. Adler-Milstein J, Holmgren AJ, Kralovec P, et al. "Electronic Health Record Adoption in US Hospitals: The Emergence of a Digital 'Advanced Use' Divide." *J Am Med Inform Assoc* 2017; 24 (6): 1142–48.
2. The Office of the National Coordinator for Health Information Technology. "Health IT Dashboard." Accessed March 2020.
3. Safran C, Bloomrosen M, Hammond WE, et al. "Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper." *J Am Med Inform Assoc* 2007. 14 (1): 1–9.
4. Casey JA, Schwartz BS, Steward WF, et al. "Using Electronic Health Records for Population Health Research: A Review of Methods and Applications." *Annu Rev Public Health* 2016. 37 (1): 61–81.
5. Botsis T, Hartvigsen G, Chen F, et al. "Secondary Use of EHR: Data Quality Issues and Informatics Opportunities." *Summit Transl Bioinform* 2010: 1–5.
6. Cowie MR, Blomster JJ, Curtis LH et al. "Electronic Health Records to Facilitate Clinical Research." *Clin Res Cardiol* 2017. 106 (1): 1–9.
7. Milinovich A, Kattan MW. "Extracting and Utilizing Electronic Health Data from Epic for Research." *Ann Transl Med* 2018. 6 (3): 42–42.
8. Goldstein BA, Navar AM, Pencina MJ, et al. "Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review." *J Am Med Inform Assoc* 2017. 24 (1): 198–208.
9. Evans RS. "Electronic Health Records: Then, Now, and in the Future." *Yearb Med Inform* 2016. 25 (S 01): S48–61.
10. Ancker JS, Kern LM, Edwards A, et al. "How Is the Electronic Health Record Being Used? Use of EHR Data to Assess Physician-Level Variability in Technology Use." *J Am Med Inform Assoc* 2014. 21 (6): 1001–8.
11. Palchuk MB, Fang EA, Cygielnik JM, et al. "An Unintended Consequence of Electronic Prescriptions: Prevalence and Impact of Internal Discrepancies." *J Am Med Inform Assoc* 2010. 17 (4): 472–76.
12. Karystianis G, Sheppard T, Dixon WG, et al. "Modelling and Extraction of Variability in Free-Text Medication Prescriptions from an Anonymised Primary Care Electronic Medical Record Research Database." *BMC Med Inform Decis Mak* 2015. 16 (1): 18.

13. Tao C, Filannino M, Uzuner Ö. "Prescription Extraction Using CRFs and Word Embeddings." *J Biomed Inform* 2017. 72 (August): 60–66.
14. Brown CL, Mulcaster HL, Triffitt KL, et al. "A Systematic Review of the Types and Causes of Prescribing Errors Generated from Using Computerized Provider Order Entry Systems in Primary and Secondary Care." *J Am Med Inform Assoc* 2017. 24 (2): 432-440.
15. Wright A, Hickman TT, McEvoy D, et al. "Analysis of Clinical Decision Support System Malfunctions: A Case Series and Survey." *J Am Med Inform Assoc* 2016. 23 (6): 1068–76.
16. Rajiv D, Patel AA, Winters S, et al. "A Multidisciplinary Approach to Honest Broker Services for Tissue Banks and Clinical Data." *Cancer* 2008. 113 (7): 1705–15.
17. Van Noorden R. "Interdisciplinary Research by the Numbers." *Nature* 2015. 525 (7569): 306–7.
18. Agniel D, Kohane IS, Weber GM. "Biases in Electronic Health Record Data Due to Processes within the Healthcare System: Retrospective Observational Study." *BMJ* 2018. 363: k4416.

There are no comments yet.