

Electronic Health Records in Mental Health Research: A Framework for Developing Valid Research Methods

Enrico G. Castillo, M.D., Mark Olfson, M.D., M.P.H., Harold Alan Pincus, M.D., David Vawdrey, Ph.D., T. Scott Stroup, M.D., M.P.H.

This Open Forum describes research methods to ensure high-quality health services research using electronic health records (EHRs) as a data source. The authors describe unique characteristics of EHRs that could introduce data inaccuracies into research, and they outline a framework for methods to evaluate the validity of EHR-derived data, including the

development of data extraction rules and ways to validate those rules. The authors argue that the use of such methods will help ensure the validity of research using EHRs as a data source.

Psychiatric Services 2015; 66:193–196; doi: 10.1176/appi.ps.201400200

Electronic health records (EHRs) have been widely adopted in the United States, hastened by the Health Information Technology for Economic and Clinical Health Act, which appropriated billions of dollars to create incentives for the “meaningful use” of electronic health systems (1). The National Institute of Mental Health’s strategic plan calls for “real world” research studies using existing infrastructures, including electronic medical records (2). EHRs hold immense promise for mental health research. Investigators are increasingly using EHRs in nearly every aspect of mental health research—to recruit research participants, gather data, implement and evaluate evidence-based practices, and conduct retrospective observational research and prospective clinical trials (3–5). EHRs are emerging as a key data source for mental health research alongside administrative data and national registries.

Informatics researchers have long recognized the research potential of EHRs as well as the inherent challenges, such as threats to patient confidentiality, lack of interoperability between different health information systems, incomplete or inaccurate data, and complexities of EHR operating environments (6–8). Efforts are under way to address these challenges. Projects such as the Health Information Technology Evaluation Collaborative are linking multiple electronic records into health information exchanges to facilitate coordination of care, quality improvement projects, and large-scale research (9). National and state governments are also partnering with EHR developers to formulate quality metrics and clinical vocabulary that can be standardized and studied across EHR platforms (10). Progress is also being made on developing sophisticated technologies, including natural language processing programs and predictive models, to extract and analyze

information in EHR systems (11). Advances in health information technology will continue to have an impact on clinicians and investigators who use EHRs.

Mental health services researchers who use all sources of secondary data encounter challenges in regard to accuracy and appropriateness of analytic methods. In health services research that uses administrative data or national registries, long-standing questions about data validity led the field to develop research methods to help ensure accuracy. Methods that demonstrate data validity provide assurance that research conclusions rest on a solid foundation. The goal of this Open Forum is to help investigators produce high-quality mental health services research using EHRs as a secondary data source by providing a framework for evaluating the validity of EHR-derived data.

EHRs AND DATA ACCURACY

Accuracy of information is critical to biomedical research, and all data sources have limitations. Paper medical records, which often serve as a standard to which other data sources are compared, do not necessarily capture what transpires in patient-physician encounters (12). An internal medicine study had physicians blindly evaluate actor-patients. Then both the physician and actor documented what transpired during the exam. The paper medical records agreed with the actors’ reports on 70% of the quality-of-care items measured; but for the actor-patients’ diagnoses, the extent of agreement was only 48% (12). EHRs have similar potential for inaccuracy. Like paper records and other sources of secondary data, EHRs are not designed primarily for research; instead they are designed to facilitate the delivery of clinical care and to

support administrative (for example, billing) functions (13). These motivations strongly influence what is recorded and the thoroughness with which it is documented in EHRs.

Three key features of EHRs distinguish them from other sources of secondary data and point to the importance of using research methods that help increase data accuracy. First, EHRs often have complex designs that can affect data accuracy (13). EHRs commonly include structured elements, such as checklists or templates, for the mental status examination, physical examination, and other aspects of routine assessment. These elements are intended to aid clinical assessments and ensure documentation of elements required for billing. Clinicians enter data by selecting from a menu of predetermined options that may, in some cases, improve data completeness (14). In other cases, however, structured data may lead to inaccuracies. For example, an EHR may include templates with certain fields prepopulated as “normal” or may automatically insert information from prior notes or from other areas of the EHR to speed documentation. Structured templates require vigilance to edit incorrect information that would otherwise automatically be entered into the EHR. In analyzing structured data, it is important to consider that providers’ lapses in vigilance could lead to inaccurately entered or omitted data.

The second feature is unstructured or narrative textual elements, which may introduce opportunities for inaccuracies. An advantage of unstructured EHR data over the text in paper records is that large amounts of electronic text across clinical sites and patient populations can be searched and analyzed in an automated way, using natural language processing or similar methods (11). Narrative text, however, may also introduce opportunities for data coding errors. For example, a pediatric study of 465 EHRs for patients presenting for otitis media demonstrated 278 subtle variations on the documentation of a fever (for example, fever, $T > 101^{\circ}\text{F}$, and temperature $> 102^{\circ}\text{F}$) (13). In evaluating variable construction, it is important to consider all pertinent phrases that could refer to the variable of interest. Without an accurate data extraction rule—that is, a search strategy for collecting pertinent data in an EHR—inadvertently missed or misinterpreted pertinent data can lead to substantial inaccuracies.

The context in which a given EHR system is used may influence the description of clinical phenomena. For example, if providers use EHRs while interviewing and examining patients, then this distraction can lead to inaccurate data entry. In addition, EHR systems are routinely used by various personnel in a number of clinical settings, each with local customs influencing providers’ personal preferences in how they interface with the EHR (6). Inconsistent documentation practices across clinical settings and provider groups can lead to poor data comparability and inaccuracies (6). For example, depending on local preferences, diabetes associated with antipsychotic medications may be described with a host of terms (for example, diabetes, DM2, metabolic syndrome, and insulin resistance) and documented in numerous locations in an EHR—in the unstructured text of a progress note, in notes

specifically dedicated to adverse medication effects, or in a section devoted to medical history.

This calls attention to a third unique feature of EHRs that affects secondary data use for clinical research. EHRs are optimized to present clinicians with a great deal of information about an individual patient. For data to be analyzed across many patients, however, data must typically be extracted from the EHR and transformed into a format that can be indexed, searched, and analyzed for research (13). Through this process, the fidelity of data can diminish, such as when specific, locally defined terms are cross-mapped to less specific categories. As a result of the transformation, pertinent data for variables of interest (such as a specific diagnosis or medication or a clinical event or outcome) may become lost or fragmented. A detailed working knowledge of the EHR’s content and the context of its use will help investigators identify and compile all of the pertinent information for their research.

In summary, EHRs have unique features that make them attractive sources for research data, combining the breadth of variables available in paper records with the automated search functions of administrative billing data or national registries. Alongside these advantages, however, EHRs have the potential to introduce inaccurate information. The full extent and degree of inaccuracies introduced into the research literature in this way are not currently quantified. This Open Forum addresses this issue by outlining a methodological framework that will allow researchers to reduce inaccuracies and enable peer reviewers to independently quantify the extent and degree of inaccuracies in future research that uses EHRs as a data source.

RECOMMENDATIONS

Some variables of interest are easily and accurately accessed from EHRs. Frequently, however, variables must be elicited by culling information from multiple areas in the EHR. Biomedical informatics researchers have described a variety of techniques for identifying patient cohorts with specific variables of interests (15). Data extraction rules locate variables of interest within the larger landscape of EHR information. This process is analogous to the case definition process that is a foundation of research using administrative data. Methods to validate these case definition algorithms are an important step in research using administrative data but are not yet standard in EHR research (16). An evaluation of 126 unique EHR-based studies in health outcomes research found that only 24% included validation methodologies (17).

The following recommendations are intended as a guide for mental health services researchers who are considering use of EHR data in their research. We describe the development of data extraction rules in detail and ways to validate and publish those methods, adapting lessons learned from research on administrative data sources and informatics in other fields (16,18). We use clozapine-associated agranulocytosis to illustrate the key steps.

Identify Relevant Data

A detailed working knowledge of the EHR's design and the ways in which it is used—the EHR's content and context—is essential to locate all of the pertinent EHR data (6). Data relevant to the variable of interest may be located in structured and unstructured data fields. For example, to identify clozapine-prescribing practices, structured medication lists and unstructured text of clinical notes might prove useful. To identify agranulocytosis, one could examine structured data (problem or diagnosis lists, laboratory data, and lists of adverse medication effects) and unstructured clinical notes, taking care to identify all potential phrases that pertain to agranulocytosis.

Rule Development as an Iterative Process

After all of the relevant sources of data are identified, initial data extraction rules are then developed. For example, a rule defining clozapine prescriptions could require documentation of a prescription in a structured medication list or a textual reference to a prescription in clinical notes. A rule defining clozapine-induced agranulocytosis would need a different approach, requiring, for example, documentation in a structured data element or a textual reference (including a list of all potential phrases that pertain to agranulocytosis, such as neutropenia, granulocytopenia, and low WBC) that temporally coincides with a clozapine prescription.

Researchers can refine these rules to more accurately capture variables of interest. This might involve specifying the types of clinical notes and health care settings included in the search, varying the requirement for documentation in multiple patient visits, or including or excluding certain free-text phrases that refer to the variables of interest.

Preliminary data extraction rules can then be tested on a sample of patients' records to determine the rules' accuracy and refine them accordingly. The provisional rules might then be compared with data collected by an alternative method. Some alternative methods, listed in approximate order of rigor, include independent assessments of patients by using structured diagnostic interviews, patient self-assessment tools, manual chart review, administrative billing data, and clinician surveys (19). By comparing the EHR data collection to another standard method for the same patient sample, the performance characteristics of the preliminary rules (that is, sensitivity, specificity, and positive and negative predictive values) can be determined. Rules can then be refined and selected by examining how subtle adjustments affect performance statistics. The variable of interest may guide which performance characteristic to optimize.

On the basis of guidelines for validation of administrative data, we recommend using several performance characteristics and reporting 95% confidence intervals to indicate the precision of the estimates (16,18). The prevalence of variables of interest in the validation sample and larger research population helps to provide a context for interpreting positive and negative predictive values.

We are aware of no mental health informatics literature to guide when and how EHR data extraction rules should be validated. In general, data extraction rules that are more complex warrant more rigorous validation methods. Another situation that favors the use of rigorous validation methods includes instances when research conclusions are of great concern, such as when the results may have a direct impact on practices and policies in a large system of care.

Publishing the Rule Development and Validation Process

Many studies using EHR data do not report methods used to ensure accuracy. This may indicate that this process was withheld from publication or omitted altogether. Publication of these methods helps reviewers and readers assess the rigor of the research and the validity of the findings. We urge investigators to report EHR data extraction rule development and validation efforts in peer-reviewed publications, echoing similar recommendations in mental health research using administrative data and EHR research in other medical fields (16,20). Rule development and validation can be reported in a stand-alone publication, as is common for studies that use administrative data or data from patient registries, or in the methods section of research reports. We further urge that researchers describe the rule development process, the rationale for the choice of validation method if applicable, and the performance characteristics of the final rule or rules. In addition, especially when complex rules or rigorous validation methods are used, we recommend reporting the qualifications of the raters and the characteristics (for example, demographic characteristics and prevalence of the variable of interest) of the validation sample compared with the overall research sample (16).

FUTURE DIRECTIONS AND CONCLUSIONS

EHRs offer key opportunities for mental health services research. By raising awareness of methodological issues and the importance of data accuracy in research using EHRs as a secondary data source, we highlight the need to conduct this work with the same rigor as that used in research on other data sources. We urge funders such as the Patient-Centered Outcomes Research Institute and the Agency for Healthcare Research and Quality to support work that promotes accuracy in EHR data and research methods. High-priority topics include development and evaluation of optimal research methods for identifying mental health EHR variables of interest that can be generalized across EHR platforms. This is especially, but not exclusively, important for variables such as depression and completed and attempted suicide, which were designated by the U.S. Food and Drug Administration's Mini-Sentinel project as moderate and high research priorities, respectively, for future validation studies because of inconsistently or poorly performing algorithms to identify these variables in administrative claims data (18). Other high-priority topics are drafting expert consensus guidelines on

types of EHR variables and the range of methods that can be used to validate them and drafting expert consensus standards on the reporting of validation methods in mental health EHR research.

AUTHOR AND ARTICLE INFORMATION

Dr. Castillo, Dr. Olfson, Dr. Pincus, and Dr. Stroup are with the Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York State Psychiatric Institute, New York City. Dr. Vawdrey is with the Department of Biomedical Informatics, Columbia University. Send correspondence to Dr. Stroup (e-mail: stroups@nyspi.columbia.edu).

Dr. Pincus has received travel funds from Value Options and funds for presentations from Johnson & Johnson. Dr. Stroup has participated in a CME program funded by Genentech and has received research funding from Auspex Pharmaceuticals.

The other authors report no financial relationships with commercial interests.

REFERENCES

1. An Introduction to the Medicaid EHR Incentive Program for Eligible Professionals. Baltimore, Centers for Medicare and Medicaid Services, EHR Incentive Programs, 2012. Available at www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EHR_Medicaid_Guide_Remediated_2012.pdf. Accessed April 4, 2014
2. Strategic Research Priorities: Strategic Objective 3.4: Identify and Systematically Study Elements of Personalized Mental Health Care. Bethesda, National Institute of Mental Health, 2008. Available at www.nimh.nih.gov/research-priorities/strategic-objectives/strategy-3.4.shtml. Accessed April 4, 2014
3. Gur RC, Irani F, Seligman S, et al: Challenges and opportunities for genomic developmental neuropsychology: examples from the Penn-Drexel collaborative battery. *Clinical Neuropsychologist* 25:1029–1041, 2011
4. Jensen PB, Jensen LJ, Brunak S: Mining electronic health records: towards better research applications and clinical care. *Nature Reviews. Genetics* 13:395–405, 2012
5. Perlis RH, Iosifescu DV, Castro VM, et al: Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychological Medicine* 42:41–50, 2012
6. Berg M, Goorman E: The contextual nature of medical information. *International Journal of Medical Informatics* 56:51–60, 1999
7. Edwards A, Hollin I, Barry J, et al: Barriers to cross-institutional health information exchange: a literature review. *Journal of Healthcare Information Management* 24:22–34, 2010
8. Shoenbill K, Fost N, Tachinardi U, et al: Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations. *Journal of the American Medical Informatics Association* 21:171–180, 2014
9. Health Information Technology Evaluation Collaborative (HITEC). Albany, New York State Department of Health, 2008. Available at www.health.ny.gov/technology/goals_and_activities/evaluation.htm. Accessed April 4, 2014
10. Kern LM, Dhopeswarkar R, Barrón Y, et al: Measuring the effects of health information technology on quality of care: a novel set of proposed metrics for electronic quality reporting. *Joint Commission Journal on Quality and Patient Safety* 35:359–369, 2009
11. Meystre SM, Savova GK, Kipper-Schuler KC, et al: Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook of Medical Informatics* 2008:28–144, 2008
12. Luck J, Peabody JW, Dresselhaus TR, et al: How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *American Journal of Medicine* 108:642–649, 2000
13. Wasserman RC: Electronic medical records (EMRs), epidemiology, and epistemology: reflections on EMRs and future pediatric clinical research. *Academic Pediatrics* 11:280–287, 2011
14. Ho LM, McGhee SM, Hedley AJ, et al: The application of a computerized problem-oriented medical record system and its impact on patient care. *International Journal of Medical Informatics* 55:47–59, 1999
15. Shivade C, Raghavan P, Fosler-Lussier E, et al: A review of approaches to identifying patient phenotype cohorts using electronic health records. *Journal of the American Medical Informatics Association* 21:221–230, 2014
16. Benchimol EI, Manuel DG, To T, et al: Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *Journal of Clinical Epidemiology* 64:821–829, 2011
17. Dean BB, Lam J, Natoli JL, et al: Review: use of electronic medical records for health outcomes research: a literature review. *Medical Care Research and Review* 66:611–638, 2009
18. Carnahan RM: Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiology and Drug Safety* 21(suppl 1):90–99, 2012
19. Townsend L, Walkup JT, Crystal S, et al: A systematic review of validated methods for identifying depression using administrative data. *Pharmacoepidemiology and Drug Safety* 21(suppl 1):163–173, 2012
20. Weiskopf NG, Weng C: Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association* 20:144–151, 2013