Title: Recommendations for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research

---

Authors: William R. Hersh[i], James Cimino[ii], Philip R.O. Payne[iii], Peter Embi[iii], Judith Logan[i], Mark Weiner[iv], Elmer V. Bernstam[v], Harold Lehmann[vi], George Hripcsak[vii], Timothy Hartzog[ix], Joel Saltz[vii]

i: Oregon Health & Science University
ii: National Institutes of Health Clinical Center
iii: The Ohio State University Wexner Medical Center
iv: AstraZeneca
v: The University of Texas Health Science Center at Houston
vi: Johns Hopkins University
vii: Columbia University
viii: Emory University
ix: Medical University of South Carolina

Introduction

The last several years have seen substantial investment in the adoption of electronic health records (EHRs) in the US and elsewhere, offering great potential to improve the quality, safety, and cost of healthcare.[1] EHR adoption is also likely to improve our ability to advance biomedical and healthcare science and research through the re-use of clinical data.[2-4]

At the same time, there has been substantial US investment in various aspects of clinical and translational research, including comparative effectiveness research (CER) that aims to study populations and clinical outcomes pertinent to real-world clinical practice.[5] Additional federal investment in research infrastructure includes the Clinical and Translational Research Award (CTSA) program of the US National Institutes of Health, leading many institutions funded by CTSA awards to develop research data warehouses derived from operational EHRs and other systems. [6, 7] Another source of federal investment has come from the Office of the National Coordinator for Health Information Technology (ONC) through its Strategic Health IT Advanced Research Projects (SHARP) Program, with one of the four major funded research areas focusing on re-use of clinical data.[8]

Our previous paper reviewed some of the successes of efforts to use operational EHR data for research.[9] One prominent success has come from the Electronic Medical Records and Genomics (eMERGE) Network, which has demonstrated the ability to validate existing research results and generate new findings mainly

in the area of genome-wide association studies (GWAS) that associate specific findings from the EHR (the "phenotype") with the growing amount of genomic and related data (the "genotype").[10, 11] Another successful effort has come from analyses derived from the Health Maintenance Organization Research Network's Virtual Data Warehouse (VDW) Project, where, for example, researchers were able to use data to demonstrate a link between hyperglycemia in pregnancy and childhood obesity.[12, 13] Using similar methods, other researchers have been able to replicate the findings of randomized controlled trials using EHR data and appropriate statistical methods.[14-18]

As we noted, however, routine clinical data are collected for clinical and billing uses, not research [9]. We described and detailed several caveats for the use of such data for CER, including inaccuracy, incompleteness, transformation in ways that undermine their meaning, inaccessibility for research, unknown provenance, insufficient granularity, and incompatibility with research protocols. Informed by these caveats, and motivated by the potential benefits of re-using operational clinical and administrative data for research, quality measurement and improvement, and other analytical purposes, we believe that there have been significant informatics advances in support of such reuse. In this paper, we adopt the critical appraisal approach of evidence-based medicine (EBM) to the problem, discuss some of the informatics challenges to the use of operational clinical data for CER, and then develop an informatics roadmap for moving forward.

Recommendations for Using EHR Data for CER

The following sections outline nine major recommendations for advancing the use of operational EHR for CER (see Figure 1). Table 1 provides a summary and description of each recommendation, which correspond with the sections and facets in the rest of the paper.

Table 1 – Summary of recommendations for advancing the use of operational EHR data for CER.

Recommendation	Description
Apply an Evidence-Based Approach	Ask an answerable question, find the best EHR data ("evidence"), appraise the data, apply data to the question
Evaluate and Manage Data	Assess availability, completeness, quality (validity), and transformability of data
Create Tools for Data Management	Create software (especially pipelines) for data aggregation, validation and transformation
Determine Metrics for	Determine whether a particular site's data are "research

Data Assessment	grade”
Develop Methods for Comparative Validation	Develop tools that support analysis of multi-site data collections
Develop a Methodology Knowledge Base	Develop a data catalogue that relates data elements to recommended transformations
Standardize Reporting Methods	Provide details of data sources, provenance and manipulation, to support comparison of data
Engage Informatics Expertise	Ensure validity of findings derived from data collected from disparate sources
Include an Informatics Research Agenda	Generate systematic studies of inherent biases in EHR and data collection methods, such as data entry user interfaces

Apply an Evidence-Based Approach

The EBM process offers many analogies for that can guide the use of operational clinical data for CER and other forms of research. Some may consider EBM to be antagonistic to EHR data re-use, as EBM gives the most value to evidence from controlled experiments, especially randomized controlled trials (RCTs), while viewing the use of real-world observational data as a lower quality form of evidence, due to its being incomplete, incorrect, and/or inconsistent. In reality, observational data analysis holds a role toward CER (or any type of clinical research) that is different from and complementary to that of traditional clinical trials. RCTs provide information on efficacy in well-defined populations under tightly regulated circumstances. Observational studies, by their nature, provide a more real-world view, cover broader populations than can be analyzed in a clinical trial setting, and can explore the associations with and effects of different levels of exposure to the intervention being studied. Therefore, both types of studies are necessary to understand how and why an intervention is working compared with expectations.

In contrast to clinical trials data, EHR data are voluminous and much more reflective of the real clinical world. In addition, performing RCTs on all clinical questions of interest would be prohibitively expensive. But we can look to the process of EBM to guide us in how to best consider EHR data as part of our body of evidence. EBM uses a principled approach to find and apply the best available information to make clinical decisions. In particular, EBM uses four steps that we can apply analogously to research use of EHR data, whether in single institutions

or across multiple institutions and settings, analogous to the use of systematic reviews and meta-analysis:[19]

1. Ask an answerable question – In applying EBM, the first step of asking an answerable question may be the most important. For example, knowing to ask whether a test or treatment is efficacious is not enough. Rather, we need to know at a minimum the effectiveness of the intervention relative to some alternative approach in a particular patient population or setting. This same approach is obviously necessary for using operational EHR data for clinical research. In selecting EHR data for use in a study, the researcher should adopt the established population, intervention, comparison, outcomes, timing, and setting (PICOTS) framework for formulating the research question.[20]
2. Find the best evidence – For the second step, the principle from EBM is very much the same, even if the techniques of obtaining evidence are very different. The "evidence" in this case is the data in EHRs and other systems that, as noted previously, may be incomplete, incorrect, and inconsistent. We therefore need to determine if we have the proper data and, if so, whether they can be applied to answer our question.
3. Critically appraise the evidence – In the third step, just as with EBM, the evidence must be critically appraised. Can we trust the inferences and conclusions from the data? Are there confounding variables in those data? Have the appropriate methods been used? These may be critical with EHR data where assignment of cause and effect could be difficult, if not impossible. The solution likely comes back to asking the right question, i.e., one for which we can have confidence that we have found the correct answer.
4. Apply it to the patient situation – Lastly, we need to ask, can the evidence be applied in our setting? Just as some RCTs answer questions in patient populations very different from those of the clinician making decisions, we must ascertain if the results obtained from this approach can be applied to a specific clinical context.

As EBM recognizes that the single studies are limited to reflecting a single setting and often with small sample sizes, the techniques of systematic review and meta-analysis have been developed to aggregate results from multiple studies and, where possible, aggregate their data in meta-analysis. The equivalent of this approach for clinical data is combining clinical data from multiple settings or institutions.

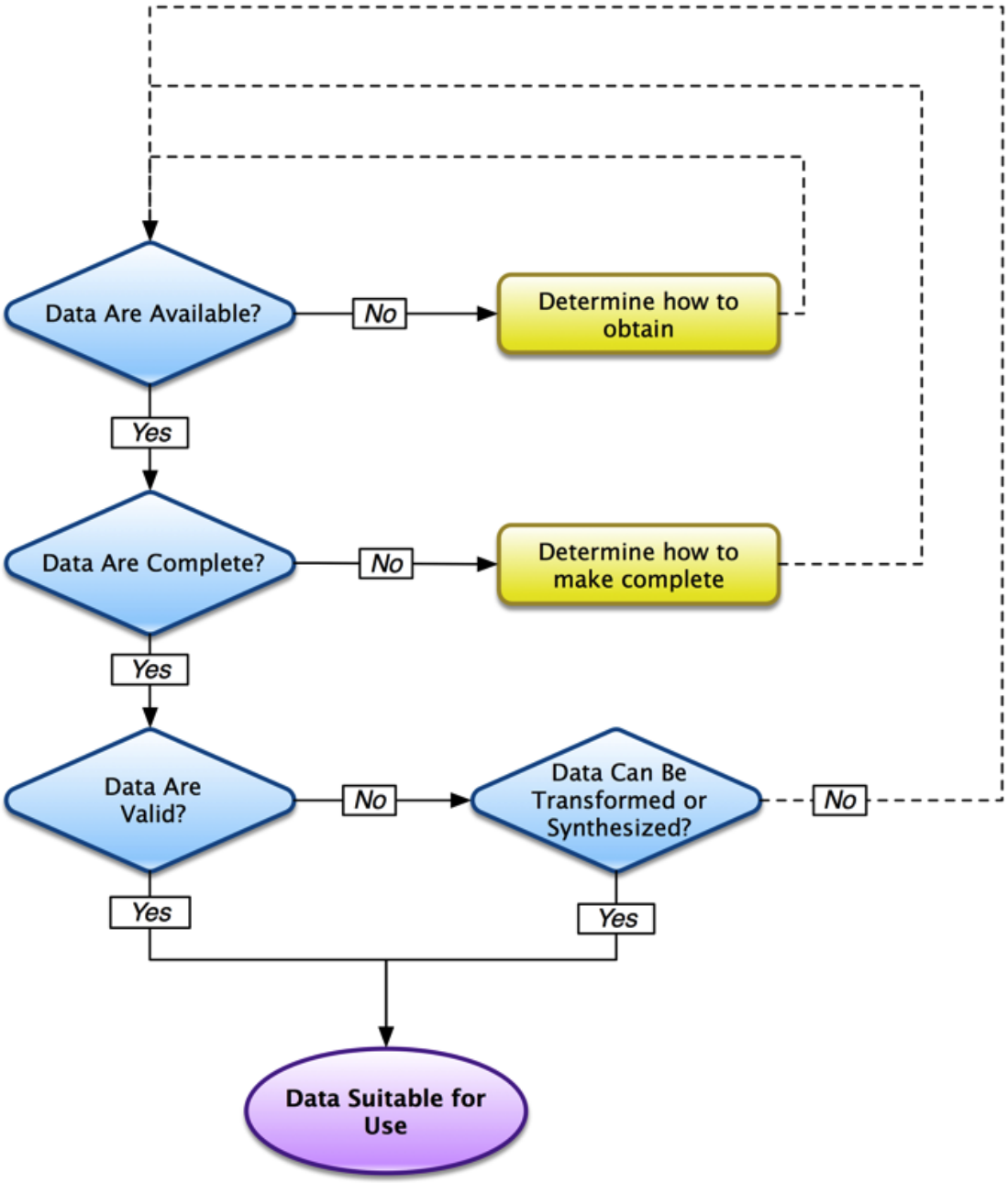
Evaluate and Manage EHR Data for CER

Any research study, whether CER or not, require the highest quality data available for observing the object of that study. For data coming from an operational EHR, this means we need to determine whether the data required are available, complete, and valid.

To illustrate the need for these types of data evaluation and management apparatus and tools, figure 1 shows a prototypical workflow consisting of the major steps needed to assess such factors for each datum in a study that uses EHR data. This process is inclusive of recommendations about data management, including evaluating data, tools for data management, metrics for data assessment, and methods for comparative validation.

Once the researcher defines the question, he or she must then identify the EHR data elements corresponding to the desired data and determine whether they are actually available. For each desired datum, he or she must next assess if that element is complete, i.e., available for all individuals whose data will be used in the study. Of course, determination of data completeness and data quality are challenging.[21,22]

Figure 1 – Assessing data from operational sources for clinical research.

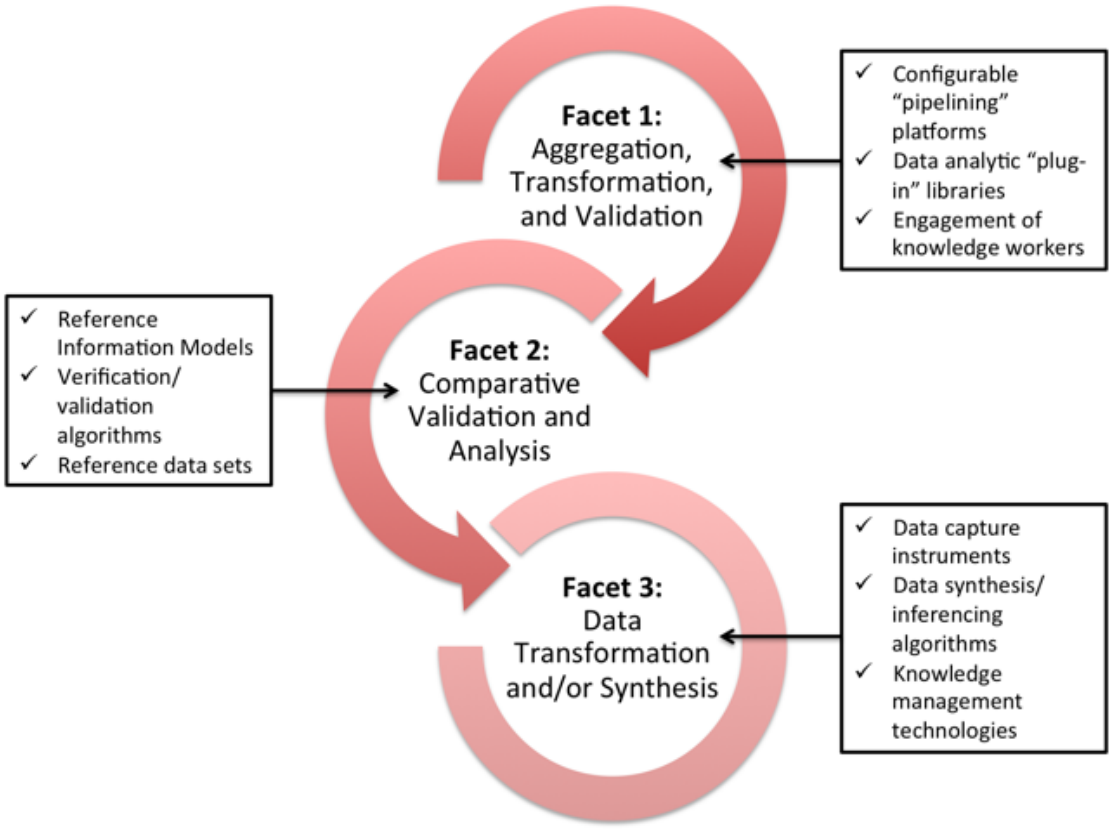


In addition to determining whether data are available and complete, the researcher must also ascertain whether the data are valid. As was seen in the caveats above, we know there are many reasons why data are not valid. A review by Weiskopf and Weng identified methods that have been used in the past to assess the validity

of data.[23] These included comparison to a gold standard, agreement of data elements among sources, comparison of data distribution across sources, validity checks, review of data capture logs, and element presence in multiple sources. If the data are not valid, then the researcher must determine if it is possible to transform them. There are a number of ways that data can be transformed, such as combining them with other data, e.g., linking claims and EHR data or adding data from other sources.[24, 25] Of course, such transformation requires further validity-checking.

If individual data are not available, complete, or transformable, then the researcher must either continue the study with noted limitations, or he or she must find alternative and/or surrogate sources of data. The researcher must also carefully consider the usefulness and relevance of each data element against a testable hypothesis and an analytic model or analytic plan in order to be efficient with the time of those maintaining and performing operations on the data. Often the scope of technical and methodological skills and tools required to address the preceding information needs is beyond the capabilities of an individual researcher, necessitating the engagement of a broader, multi-disciplinary team of investigators, informaticians, and research or technical support staff. Building upon this requirement and informed by the findings and caveats introduced previously, we consider several critical facets of our recommendations, with the objective of ultimately motivating a CER-relevant informatics research and development agenda aligned with them. An illustration of the interrelationships between our proposed recommendations, their constituent facets, and the data-analytic workflow introduced in Figure 1 is then provided by the elements in Figure 2. We now describe these facets.

Figure 2 – Relationships between core datum assessment operations and the facets of informatics research and development needed to support and enable comparative effectiveness research.



Facet 1: Tools and methods capable of supporting the aggregation, transformation, and validation of heterogeneous data sets

Investigators who carry out certain types of research, such as CER but also including health services research, epidemiology, pharmacogenomics, and genome-wide association studies, all need to aggregate information obtained from different health systems and their data sources. In all of these cases, investigators must also aim to verify the accuracy of information obtained from each source and use the appropriate tools to transform and optimize the data for research use. In order to satisfy such information needs in a scalable and contextually

appropriate manner, CER investigators must be provided with access to highly-usable, configurable, and shareable tools and methods that can be used to build, instrument, and apply data-analytic pipelines targeting the preceding requirements. Such pipelines would ideally be composed of multiple "plug ins" or modules that are made available for community-wide re-use via public libraries or equivalent means.

CER researchers using EHR and other operational data must also apply methods to ensure their completeness and quality. Emerging methods for data completeness include explicit and predictive analysis of data in the EHR.[21] Likewise, methods for data quality assessment include such techniques as checks for adherence to common data models across sites and comparing actual versus expected observations within sites.[22, 26] CER researchers must also work with clinicians, health care organizations, and others to improve the data that informs their "downstream" research.

It is also possible to gain insight into potential approaches to specific data aggregation and quality assurance best practices based on analogous efforts that have been undertaken in the context of methods used to evaluate the quality and reproducibility of clinical laboratory results. A variety of tests can be carried out to characterize internal consistency of results at each site. Results obtained by laboratories are typically compared; for example, laboratories analyze specimens produced by an accrediting body such as the College of American Pathologists (CAP), which then in turn generates reports comparing the performance of participating laboratories. A comparable approach to testing EHR data could be developed to ascertain a clinical organization's ability to produce "research-grade" data.

<u>Facet 2 – Tools and methods that can enable the comparative validation and analysis of multi-site data collections</u>

Comparisons between clinical sites are complicated by the fact that patient populations naturally differ between health systems. However, statistical properties associated with relationships between diagnostic entities, laboratory values and medication orders can be evaluated for each site and then compared from site to site, possibly applying normalization or other transformations of the data where appropriate. The key here is likely to lie in examining inter-relationships between co-morbidities with given patient demographics; one would for instance expect characteristic collections of co-morbidities related to each broad class of diseases. Thus we must develop and use a compendium of reference information models (RIMs), verification/validation algorithms, and

reference data sets (e.g., for use as comparative "gold standards") so that CER investigators can draw upon and rigorously apply a well-known, validated, and reproducible "tool box" of comparative analysis methods. For example, a rich set of "sanity check" characterizations can and should be carried out to assess internal consistency of each institution's clinical data characterizations. For instance, one can expect characteristic constellations of diagnostic codes, laboratory test values and medication orders for a patient classified as having diabetes mellitus (DM) with renal, ophthalmic or neurological manifestations; e.g., a patient classified as having DM with renal complications would be expected to have abnormal estimated glomerular filtration rate (eGFR) and/or proteinuria.

Facet 3 – Tools and methods for data transformation and/or synthesis in support of CER

As noted previously, CER researchers require validated point-by-point comparative data among the study populations. The use of EHR data presents unique challenges for obtaining sufficient data of the particular comparators whether the study compares, for example, different medications, medical devices, or types of surgical vs. non-surgical interventions. Many EHR systems are not designed to collect accurate and complete data sets that are sufficiently robust for such comparisons

Other unique challenges and needs in CER are the lack of nationally adopted data-related research methodologies that can be utilized in person-centered comparative outcomes studies and pragmatic studies that reflect real-world settings. EHR data that are collected, archived, and made accessible must be standardized across EHR vendors and users with ongoing procedures (as described in the previous section) to ensure consistency of results within sites and comparability across sites. For example, patient-reported outcomes, patient problem lists, and functional outcome measures are now collected using different tools, and may not be uniformly or completely collected in EHRs [27]. In addition, there are differences in services provided, data entered by health care entities, and information provided by patients that contribute to the inconsistency of completeness and accuracy of EHR data for CER.

The above challenges need to be addressed to advance CER through the development of widely applicable methodological and technical solutions.[28] One approach is the creation, maintenance, and dissemination of collections of platform-independent: 1) data capture instruments; 2) algorithms for the inference or synthesis of complex and potentially temporally dependent phenotypes from primary data; and 3) knowledge management technologies and best practices.

Such tools could support improved understanding of data provenance in EHRs and other systems such that researchers have a sufficient understanding of the sources and potential biases surrounding the data for a given research study.

Develop a Methodology Knowledge Base

In addition to devising methods and tools for improved data reconciliation and integration as introduced in the three preceding facets, solutions to the challenges that face those attempting to re-use data for research purposes require an informatics roadmap. Such an informatics roadmap must include resources to aid CER researchers in applying complex methods arising from the use of operational data. At the simplest level, a resource that lays out the concerns raised in this article is needed to which CER researchers can refer. This may include for example, a table relating data elements and types and transformations recommended or not. Such a resource – a methodology knowledge base – would need to be updated as novel issues are raised, as novel solutions are offered, and as instructive examples are shared. The next level would be a more interactive resource that supplies this knowledge in a more active mode than a passive knowledge base. CER researchers, especially novice ones, employing EHR data could benefit from the integration of methodological decision support into their research environment. This should also include new research methodologies as well as new ways to measure and encourage research-related activities during routine practice.[29, 30, 31]

Another informatics challenge and opportunity is making the efficient recording of high-quality and standardized data easier for clinicians. As noted above, clinical documentation is not usually a priority for busy clinicians.[32, 33] Research has shown clinicians already spend a substantial amount of time (20-103 minutes per day) entering data related to patient care.[34] Where possible, clinicians must be aided by the easy incorporation of other sources of reliable data.[35] We also must "bake in" data standards so that their use by clinicians and others in the health system is seamless.[36] Better adherence to standards upfront will be make data interoperability easier downstream, especially in its use for CER.

Whereas traditional observational studies usually enlist a human being to enter data into a research database, manual curation in large-scale EHR-based studies is not practical. Therefore, an even more detailed description of the exact data sources and methods to encode the data into research variables is needed to allow future researchers to repeat a study more reliably, to compare studies, and to better identify potential biases in the results. For example, when there are multiple database queries for the same intended search, e.g., "all patients with diabetes,"

this may be searched as all patients "with a diabetes-related billing code," "with a diabetes-related problem listed on their last problem list," "with an order for insulin, "with an elevated hemoglobin A1C level," or any number of other methods. The performance characteristics (e.g., sensitivity and positive predictive value) of these queries should be ascertained or at least alternative methods should be compared and reported.

Standardize Metadata Reporting

Likewise, the readers and consumers of research using these data require reporting methods that go beyond stating the source of the data or a cursory explanation of how they were transformed. To this end, much more detailed methodology of data sources, provenance, and manipulation must be provided in sufficient detail to provide an understanding and the limitations of such use. This will likely require a reporting standard comparable to the Consolidated Standards of Reporting Trials (CONSORT) and STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) tools. [37, 38] The STROBE checklist, for example, stipulates that all outcomes, exposures, predictors, potential confounders, and effect modifiers be clearly defined, giving diagnostic criteria (if applicable) and that for each variable of interest, sources of data and details of methods of assessment (measurement) be given, describing comparability of assessment methods if there is more than one group being studied.

Creating a common format for reporting such details would be beneficial, as researchers would be encouraged to provide complete information that make comparison of studies easier. The format can be seen as way of presenting an explicit description of a study's data assumptions. Multicenter observational studies may require separate descriptions for each institution if the algorithm used to generate the study variables differs among institutions. The description may require not just the researcher's procedures, but also a description of relevant aspects of how the EHR is used in the institution. For example, if only a subset of clinicians use order entry, then using orders as a source for coding the use of a medication could bias the results.

Studies using EHR data must also embrace the diversity of disease definitions in ways that set different expectations for presentation of results than for clinical trials. Results of clinical trials provide a single point estimate and a confidence interval. Observational studies, on the other hand, should present multiple results under different assumptions. Sensitivity analyses should be the norm in reports of these types of analyses. Readers should see how results compare under different

assumptions regarding disease definitions. If results are robust to assumptions, the findings are strengthened. If results differ, then the trends in the difference related to the characteristics of the underlying cohorts should be clinically plausible. Adhering to a "clinical trial" standard of making an a priori decision about cohort definition, and presenting findings based on that one analysis, can be unintentionally misleading (or purposefully misleading if the investigators did test different cohorts and reported on only one result that supports their hypotheses).

Engage Informatics Expertise

Of course, as the sections above illustrate, there are numerous informatics issues that must be addressed to enable the successful re-use of clinical data for CER. To that end, another key recommendation must be for data-based research projects to include experts in biomedical informatics, a field closely related to but distinct from computer science, IT, and biomedicine.[39] Just as the inclusion of those with particular methodological expertise in key research activities (e.g., biostatisticians, subject matter experts, etc.), the inclusion of informatics experts can help address the validity of findings generated during CER studies that rely on the re-use of clinical data often collected initially for different and varied purposes and deriving from disparate systems .[40]

Generate an Informatics Roadmap for CER

Achieving the goals of the recommendations as outlined will require an informatics research agenda. There has been very little study, for example, of how the EHR (or really any medical record) works and what biases the health care process creates in its data. [41, 42] This is an especially challenging area because whereas completeness and accuracy are somewhat apparent and measurable, bias is insidious. The data may look about right, but due to influences such as the health care process [41], the effects on the recorded data can influence the outcome .[42] We must study workflows, develop and evaluate user interfaces that allow the entry of high-quality data in time-efficient ways, and characterize the limitations of all data to better assess how they can be improved. Such research is not simply a part of the methodology of a clinical study but a research endeavor in its own right.

While there are many caveats related to the reuse of operational clinical data, there are also tremendous opportunities. Some successes have already been achieved, and with attention to the techniques proposed in this paper, we believe that additional ones will likely be forthcoming.

References

1. Blumenthal D and Tavenner M, *The "meaningful use" regulation for electronic health records.* New England Journal of Medicine, 2010. 363: 501-504.
2. Safran C, Bloomrosen M, Hammond WE, Labkoff SE, Markel-Fox S, Tang P, et al., *Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper.* Journal of the American Medical Informatics Association, 2007. 14: 1-9.
3. Weiner MG and Embi PJ, *Toward reuse of clinical data for research and quality improvement: the end of the beginning.* Annals of Internal Medicine, 2009. 151: 359-360.
4. Hripcsak G and Albers DJ, *Next-generation phenotyping of electronic health records.* Journal of the American Medical Informatics Association, 2012. 20: 117-121.
5. Sox HC and Goodman SN, *The methods of comparative effectiveness research.* Annual Review of Public Health, 2012. 33: 425-445.
6. Collins FS, *Reengineering translational science: the time is right.* Science Translational Medicine, 2011. 3: 90cm17. http://stm.sciencemag.org/content/3/90/90cm17.full.
7. MacKenzie SL, Wyatt MC, Schuff R, Tenenbaum JD, and Anderson N, *Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey.* Journal of the American Medical Informatics Association, 2012. 19(e1): e119-e124.
8. Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, et al., *Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project.* Journal of Biomedical Informatics, 2012. 45: 763-771.
9. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al., *Caveats for the use of operational electronic health record data in comparative effectiveness research.* Medical Care, 2013. 51(Suppl 3): S30-S37.
10. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al., *The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies.* BMC Genomics, 2010. 4(1): 13. http://www.biomedcentral.com/1755-8794/4/13.
11. Denny JC, *Mining Electronic Health Records in the Genomics Era*, in *PLOS Computational Biology: Translational Bioinformatics*, Kann M and Lewitter F, Editors. 2012.

12. Hornbrook MC, Hart G, Ellis JL, Bachman DJ, Ansell G, Greene SM, et al., *Building a virtual cancer research organization.* Journal of the National Cancer Institute Monographs, 2005. 35: 12-25.
13. Hillier TA, Pedula KL, Schmidt MM, Mullen JA, Charles MA, and Pettitt DJ, *Childhood obesity and metabolic imprinting: the ongoing effects of maternal hyperglycemia.* Diabetes Care, 2007. 30: 2287-2292.
14. Tannen RL, Weiner MG, Xie D, and Barnhart K, *A simulation using data from a primary care practice database closely replicated the Women's Health Initiative trial.* Journal of Clinical Epidemiology, 2007. 60: 686-695.
15. Weiner MG, Barnhart K, Xie D, and Tannen RL, *Hormone therapy and coronary heart disease in young women.* Menopause, 2008. 15: 86-93.
16. Tannen RL, Weiner MG, and Xie D, *Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication.* Pharmacoepidemiology and Drug Safety, 2008. 17: 671-685.
17. Tannen RL, Weiner MG, and Xie D, *Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings.* British Medical Journal, 2009. 338: b81. http://www.bmj.com/cgi/content/full/338/jan27_1/b81.
18. Danaei G, Rodríguez LA, Cantero OF, Logan R, and Hernán MA, *Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease.* Statistical Methods in Medical Research, 2011. 22: 70-96.
19. Straus SE, Glasziou P, Richardson WS, and Haynes RB, *Evidence-Based Medicine: How to Practice and Teach it, 4e*. 2010, New York, NY: Churchill Livingstone.
20. Buckley DI, Ansari M, Butler M, Williams C, and Chang C, *The Refinement of Topics for Systematic Reviews: Lessons and Recommendations From the Effective Health Care Program*. 2013, Agency for Healthcare Quality & Research: Rockville, MD, http://www.ncbi.nlm.nih.gov/books/NBK121274/.
21. Weiskopf NG, Hripcsak G, Swaminathan S, and Weng C, *Defining and measuring completeness of electronic health records for secondary use.* Journal of Biomedical Informatics, 2013. 46(5):830-836.
22. Brown JS, Kahn M, and Toh D, *Data quality assessment for comparative effectiveness research in distributed data networks.* Medical Care, 2013. 51(Suppl 3): S22-S29.

23. Weiskopf NG and Weng C, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2012. 20: 144-151.
24. Dore DD, Liang C, Ziyadeh N, Norman H, Bayliss M, and Seeger JD, *Linkage of routinely collected oncology clinical data with health insurance claims data—an example with aromatase inhibitors, tamoxifen, and all-cause mortality.* Pharmacoepidemiology and Drug Safety, 2012. 21(Suppl 2): 29-36.
25. Lagu T, Rothberg MB, Nathanson BH, Steingrub JS, and Lindenauer PK, *Incorporating initial treatments improves performance of a mortality prediction model for patients with sepsis.* Pharmacoepidemiology and Drug Safety, 2012. 21(Suppl 2): 44-52.
26. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, and Steiner JF, *A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research.* Medical Care, 2012. 50(Suppl): S21-S29.
27. Iezzoni LI, *Multiple chronic conditions and disabilities: implications for health services research and data demands.* Health Services Research, 2010. 45: 1523-1540.
28. Hirsch BR, Giffin RB, Esmail LC, Tunis SR, Abernethy AP, and Murphy SB, *Informatics in action: lessons learned in comparative effectiveness research.* Cancer Journal, 2011. 17: 235-238.
29. Horn SD and Gassaway J, *Practice based evidence: incorporating clinical heterogeneity and patient-reported outcomes for comparative effectiveness research.* Medical Care, 2010. 48(6 Suppl): S17-S22.
30. Adamson J, vanStaa T, and Torgerson D, *Assuring the safety and effectiveness of new drugs: rigorous phase IV trials randomizing general practices to delayed access to new drugs.* Journal of Health Services Research & Policy, 2012. 17: 56-59.
31. Embi PJ and Payne PR, *Evidence generating medicine: redefining the research-practice relationship to complete the evidence cycle.* Medical Care, 2013. 51(Suppl 3): S87-S91.
32. deLusignan S and vanWeel C, *The use of routinely collected computer data for research in primary care: opportunities and challenges.* Family Practice, 2005. 23: 253-263.
33. Poissant L, Pereira J, Tamblyn R, and Kawasumi Y, *The impact of electronic health records on time efficiency of physicians and nurses: a systematic review.* Journal of the American Medical Informatics Association, 2005. 12: 505-516.

34. Hripcsak G, Vawdrey DK, Fred MR, and Bostwick SB, *Use of electronic clinical documentation: time spent and team interactions.* Journal of the American Medical Informatics Association, 2011. 18: 112-117.
35. Cimino JJ, *Improving the electronic health record: getting what we wished for.* Journal of the American Medical Association, 2013. 309: 991-992.
36. Dolin RH and Alschuler L, *Approaching semantic interoperability in Health Level Seven.* Journal of the American Medical Informatics Association, 2011. 18: 99-103.
37. Schulz KF, Altman DG, and Moher D, *CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials.* Annals of Internal Medicine, 2010. 152: 726-732.
38. vonElm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, and Vandenbroucke JP, *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.* Annals of Internal Medicine, 2007. 147: 573-577.
39. Bernstam EV, Hersh WR, Johnson SB, Chute CG, Nguyen H, Sim I, et al., *Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical and Translational Science Award programs.* Academic Medicine, 2009. 84: 964-970.
40. Hersh WR, Carey TS, Ricketts T, Helfand M, Floyd N, Shiffman RN, et al., *Comparative Effectiveness Workforce - Framework and Assessment*, in *Learning What Works: Infrastructure Required for Comparative Effectiveness Research*, Olsen LA, Grossman C, and McGinnis JM, Editors. 2011, National Academies Press: Washington, DC. 192-220.
41. Hripcsak G, Knirsch C, Zhou L, Wilcox A, and Melton G, *Bias associated with mining electronic health records.* Journal of Biomedical Discovery and Collaboration, 2011. 6: 48-52. http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/jbdc/article/view/3581/2997.
42. Hripcsak G and Albers DJ, *Correlating electronic health record concepts with health care process events.* Journal of the American Medical Informatics Association, 2013. 20: 117-121.