



Raman, S. R. et al. (2018) Leveraging electronic health records for clinical research. *American Heart Journal*, 202, pp. 13-19.
(doi: [10.1016/j.ahj.2018.04.015](https://doi.org/10.1016/j.ahj.2018.04.015)).

This is the author's final accepted version.

There may be differences between this version and the published version.
You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/164877/>

Deposited on: 05 July 2018

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Leveraging Electronic Health Records for Clinical Research

Short title: Raman et al.: Leveraging EHR*

Sudha R. Raman, PhD^a; Lesley H. Curtis, PhD^a; Robert Temple, MD^b; Tomas Andersson, MD, PhD^c; Justin Ezekowitz, MBBCH, MCs^d; Ian Ford, PhD^e; Stefan James, MD, PhD^f; Keith Marsolo, PhD^g; Parsa Mirhaji, MD, PhD^h; Mitra Rocca, PhD^b; Russell L. Rothman, MD, MPPⁱ; Barathi Sethuraman, PhD^j; Norman Stockbridge, MD, PhD^b; Sharon Terry, MA^k; Scott Wasserman, MD^l; Eric D. Peterson, MD, MPH^a; Adrian F. Hernandez, MD, MHS^a.

From the ^aDuke Clinical Research Institute, Durham, NC, USA; ^bFood and Drug Administration, Silver Spring, MD, USA; ^cAstraZeneca R&D, Molndal, Sweden; ^dCanadian VIGOUR Centre, University of Alberta, Edmonton, Alberta, Canada; ^eRobertson Centre for Biostatistics, University of Glasgow, Glasgow, UK; ^fUppsala Clinical Research Center, Uppsala University, Uppsala, Sweden; ^gCincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Cincinnati, OH, USA; ^hAlbert Einstein College of Medicine, Bronx, NY, USA; ⁱVanderbilt University Medical Center, Nashville, TN, USA; ^jSt. Jude Medical, Saint Paul, MN, USA; ^kGenetic Alliance, Washington, DC, USA; ^lAmgen, Inc., Thousand Oaks, CA, USA

*The contents of this paper report the findings of the Leveraging EHR for Clinical Research
Now! Think Tank (Washington, DC, February 18–19, 2016)

Address for correspondence: Sudha Raman, PhD; Department of Population Health Sciences,
Duke University School of Medicine, 2200 West Main Street East, Durham, NC 27705; Tel:
919-668-2690; Fax: Email: Sudha.raman@duke.edu

Word count: 4,650

DRAFT

Abstract

Electronic health records (EHRs) can be a major tool in the quest to decrease costs and timelines of clinical trial research, generate better evidence for clinical decision-making, and advance healthcare. Over the past decade, EHRs have increasingly offered opportunities to speed up, streamline, and enhance clinical research. EHRs offer a wide range of possible uses in clinical trials, including assisting with pre-study feasibility assessment, patient recruitment, and data capture in care delivery. In order to fully appreciate these opportunities, healthcare stakeholders must come together to face critical challenges in leveraging EHR data, including data quality/completeness, information security, stakeholder engagement, and increasing the scale of research infrastructure and related governance. Leaders from academia, government, industry, and professional societies representing patient, provider, researcher, industry, and regulator perspectives convened the Leveraging EHR for Clinical Research Now! Think Tank in Washington, DC (February 18–19, 2016) to identify barriers to using EHR in clinical research and to generate potential solutions. Think tank members identified a broad range of issues surrounding the use of EHRs in research and proposed a variety of solutions. Recognizing the challenges, the participants identified the urgent need to look more deeply at previous efforts to use these data, share lessons learned, and develop a multi-disciplinary agenda for best practices for using EHR in clinical research. We report the proceedings from this think tank meeting in the following paper.

Key words: electronic health records; clinical research; think tank meeting proceedings

Introduction

An electronic health record (EHR) can be defined as a digital longitudinal repository of electronic health information about an individual patient that provides medical personnel with the information necessary for patient care and healthcare delivery.^{1,2} There is a growing need for efficient and cost-effective evidence that can guide real-world clinical decision-making; EHRs offer breakthrough opportunities to meet this need in clinical research. Since the enactment of the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009,³ 100% of hospital-based and 90% office-based physicians in the United States are now using EHRs, which is a five-fold increase in EHR adoption over the past eight years. Although EHRs are only one type of real-world data,⁴ the major shift toward physician and hospital EHR adoption coupled with recent technological advances, offers an unprecedented opportunity to make use of data routinely collected in EHRs to help answer questions about patient health.

The widespread use of EHRs provides opportunities, but there are many challenges and considerations that need to be addressed before EHRs can be widely used in clinical trials. First, we must consider data quality. An ideal EHR would provide data that is relevant, accurate, complete, standardized, and in the correct format for clinical research, yet also accessible at a reasonable monetary and time-related cost; there currently remains a large gap between ideal and actual EHR data. Second, we must find appropriate methods for reducing bias and confounding. Third, in order to unlock EHRs' full potential, we must tackle the technical, ethical, and cultural challenges that limit the efficient use of EHRs for research.⁵ Fourth, as clinical trial initiatives using EHRs evolve to include increased participant engagement, more complex informed consent, and increased need for data sharing, researchers will have to find ways to ethically involve patients, share data, and ensure that new research endeavors return value to trial

participants, clinicians, and researchers alike. Finally, from a regulatory perspective, we must consider what standards will be acceptable and applicable for broader use of real-world data.

Given the myriad of EHR opportunities and challenges, leaders from academia, government, industry, and professional societies convened the Leveraging EHR for Clinical Research Now! Think Tank in Washington, DC (February 18–19, 2016) to identify the most pressing challenges facing the integration of EHR systems and data into clinical research; generate potential approaches for addressing these issues; and encourage efficient development of the relevant knowledge, personnel, and infrastructure necessary to maximize the integration of EHRs. This paper outlines the proceedings from this meeting. Although the primary focus of the paper concerns research within the United States (U.S.), many of the issues discussed are generic and will apply to healthcare systems internationally.

Major challenges in use of EHR data

The utility of EHRs in research will depend on many factors, including the specific type of research used, the kinds of questions that are asked, the clinical settings in which these questions are asked, the specific endpoints measured, the development phase of the drug or device being tested, whether the intervention is at the patient-, program-, or system-level, how well the types of required data are routinely captured in the EHR, and the study design. For example, depending on the study design, the extent of comorbidity data in the EHR may be suitable for identifying and recruiting patients for a clinical trial,^{6,7} but these data may not be accurate or complete enough to control for confounding in a comparative effectiveness medication study.⁸ Similarly, a randomized trial within an EHR system can assure more effective control of known and unknown confounders and can increase our confidence in any resulting causal inferences. In this

setting, problems of data quality should not lead to bias, but they could create variability that would mask treatment differences. The challenges outlined below address data quality and data access in terms of sharing data and patient consent, as well as questions about how best to engage clinicians in the research process.

Understanding the quality of EHR data for clinical research

EHR data are created to record a clinical interaction, serving as the longitudinal health record for a patient and fueling the administrative process of coding healthcare interactions for billing and system management purposes. There seems to be considerable overlap between the recorded patient outcomes and interventions, and the kind of events records in clinical trials, so efforts to use EHRs in clinical research are increasing and may facilitate the implementation of blinded randomized controlled trials. At the same time, concerns arise about whether the quality of the data is suitable for research.^{5,9} The quality of EHR data directly impacts the research that may follow, and can be illustrated by our ability to characterize missingness. That is, can we tell whose data are missing from the EHR or what types of data are missing? What is the best way to characterize the population, interventions, covariates, and outcomes of interest in a way that is complete and accurate?

Characterizing the study population

If we define a population as those people who interact with a given healthcare system, then a population observed in EHRs may actually be more representative of the general population than a typical clinical trial study group. On the other hand, there is some potential selection of a non-representative population, depending on who is approached for a clinical trial and who consents

to participate. Often, clinical trial participants are healthier than the broader population (i.e., non-trial participants who also have the condition of interest) and this could be true of EHR-identified patient populations. Alternatively, populations that are in contact with healthcare systems (and, therefore, included in EHR data) do not represent a random sample of the underlying population¹⁰; rather, people in more frequent contact with healthcare systems tend to be sicker and are more likely to be insured.¹¹ In addition, an EHR does not typically indicate accurate information about when a given person would be observable for a research study (e.g., an entry date may be recorded, but an exit date [when a person no longer would seek care at that health system] is not). As a result, unlike clinical trial data (where entry and exit dates correspond to the study start and stop dates) or insurance claims data (where entry and exit dates are tied to the dates of enrollment), EHR data in the U.S. do not include well-characterized entry and exit data for a given population. Consequently, a clinical trial that uses EHR data would need to consider the collection of additional individual data about follow-up and patient status during study time.¹²

Characterizing the interventions and outcomes

The types of EHR data that are available for individuals in a given population pose a challenge for clinical research. An EHR may record factors related to a clinical visit, but the types of details captured may be systematically influenced by the subsequent use of these records for billing purposes; therefore, an EHR may be less likely to reflect the determinants of care that are relevant to research goals. For example, problem lists may be sufficient for clinical decisions and administrative purposes, but not detailed enough for a researcher to determine the onset date of a given comorbidity. Similarly, an EHR text field that indicates a diagnosis may be sufficient for

clinical care, but would not show the criteria used to determine the diagnosis, nor would the text field reflect the absence of a diagnosis—both of which are necessary when considering data use in a clinical trial setting.

Additionally, patient-reported outcomes (PROs) may be used in studies to measure the impact of interventions, but PRO data may not be included in EHR. If PRO data are included, then they are typically not codified¹³ or routinely collected as part of healthcare delivery, which makes the prospect of using this information for research purposes quite difficult. PROs, or other study-specific tests (such as exercise tests), could not be a study endpoint unless special efforts were made to incorporate these study-specific tests either as part of the trial procedures or routinely as part of clinical practice requirements. Finally, documentation about clinical decisions is usually omitted from EHRs; therefore, clinical researchers for observational studies must infer a provider's rationale (e.g., indications, contraindications) and are not able to include or quantify the influence of family and patient preferences on healthcare decision-making and outcomes.

EHR data sources that exist in other countries (Table 1) often include unique personal identifiers that allow the linkage of healthcare data to national population, census, and education registries; however, the U.S. does not have this type of data, nor the capability to easily perform these types of linkages. Consequently, we are limited in our ability to account for confounding by non-healthcare determinants of health, such as employment status, contact with the judicial system, or a person's housing/location.

All types of health data are compromised by coding errors and inconsistencies across providers and time; however, the unique state of fragmentation within the U.S. healthcare system results in EHR data that lack vital information about the care a patient receives outside of a given

healthcare system. Furthermore, if such data are included, they may be incomplete or inaccessible. Without interoperability between different healthcare systems, the extent to which an individual receives care at other facilities is unknown for most EHR systems.

The Meaningful Use program was designed to encourage eligible professionals and hospitals to successfully demonstrate meaningful use of EHRs in order to qualify for incentive payments.¹⁸ Meaningful Use, and other programs of this kind, were created to increase interoperability between health systems, yet a large portion of the sophisticated, detailed patient information encouraged in these programs are unstructured text or images.¹⁹ As a result, patient information is not easily accessible or standardizable for clinical research. In particular, behavioral and mental health information are likely to be missing from the research data, especially if these data were not ascertained using systematized measures (Table 2).²⁰

Challenges to sharing EHR data

EHR data in the U.S. reflects at best a poorly coordinated or fragmented system. In order for us to leverage our health data into a useful tool for interventional and observational research, data sharing is vital to creating data sources that are complete and accurate. There are several ways to increase the utility of a single system's EHR data, including linking datasets to the birth and death index, to existing registries for specialized cohorts, to private and public insurance claims data, and (perhaps most importantly) to other EHRs that may have complementary data.²¹

However, there are challenges to developing a scalable infrastructure for data linking to conduct EHR research across multiple institutions. First, there is often a lack of standardized data elements, which require a common understanding of the meaning and format of the data, as well as logical, technical, and information security concerns. Second, EHR data can present complex

ethical questions and can cause difficulty reconciling the often competing interests of the various data stewards and researchers.^{22,23} Finally, although awareness and understanding of patient perspectives about the use of the EHR for clinical research is growing, there remains a lack of understanding from the patient perspective regarding linking to other data sources.²⁴ All of these challenges hinder the development of large-scale research projects that could fully leverage the diverse wealth of clinical data to generate new health-related knowledge.

Emerging dimensions of patient consent and involvement

Patient consent in EHR research

EHR data are collected primarily for use in the clinical setting. Accordingly, the consent obtained by the health system from the patient generally applies only to the use of these data for clinical care and other operational purposes. Consent is typically not required when patient data are used by the health system to optimize care delivery through quality improvement initiatives, by payers to optimize their business, or by clinical researchers. In traditional clinical research, including trials, the consent that is obtained at the start of the trial governs the collection of study data, which is generally gathered from the patient at standalone clinical visits via a dedicated electronic case report form (eCRF). Increasingly, as pragmatic clinical trial study designs emerge and data are collected in the course of regular healthcare, many questions arise about the type and extent of patient consent that should govern these activities.²⁵ There is growing concern that the current ethical standards around consent for using EHR data for clinical research are limiting the generalizability of research results (e.g., selection of research participants) and increasing research costs.^{26,27} Furthermore, when presented with new ways of using EHR data for prospective research, patients may have reservations about whether their health data are being

used for profit purposes or in ways they would not permit if they were fully informed. Rather than creating standalone research datasets, clinical research increasingly involves the development of registries and data repositories that can serve as platforms for a broad range of future research uses, including linkage to other data sources. Uncertainty exists about the best way to obtain patient consent for ongoing secondary use of such data, since future research projects cannot usually be identified at the time of initial consent. These concerns suggest that traditional approaches to obtaining patient consent data may not be appropriate for EHR research.

Other types of patient involvement in EHR research

Looking beyond the initial consent issues, patient involvement in EHR-based clinical research includes the ways in which patients are contacted for research purposes via the EHR, how data are collected from these patients, what access is permitted to the research data within the EHR, and how EHR systems can be used to share research results with participants. Since clinical research often uses information held within the EHR to define and construct study cohorts, requests to the patient for research involvement are increasingly common, but risk overwhelming potential participants. Once enrolled in a study, patients may provide data at routine medical care visits through web portals or their own wearable medical devices; both of these technologies introduce a new layer of contextual data about the patient between clinic visits. Many questions remain unanswered regarding patients' readiness to make these data available, the reliability of sources, and the data's ultimate purpose in clinical research. As our understanding of controlling EHR data shifts from an ownership model to one of data stewardship, we must delineate a patient's ability and preference to view, edit, and participate in the secondary uses of EHR.²⁸

Finally, most research results currently exist in the form of products directed toward clinicians and researchers, with the anticipation that trickle-down effects will result in practice change. If these research results are directed at patients, then they are often in the form of standalone web sites or educational materials, rather than integrated into the EHR. For each of these domains, there is wide variability in how patients, health systems, and researchers interact. Only recently has evidence emerged about the best practices for eliciting EHR research-related patient preferences and priorities.

Engaging clinicians in EHR research

Clinical research is increasingly taking place in real-world settings. As a result, in pragmatic clinical trials embedded in learning health systems that rely upon EHR, the importance of the clinician's role increases. In the course of daily clinical care, non-researcher clinicians are often involved in many research activities that require interaction with EHR; these activities include obtaining access to and recruitment of patients, education and obtaining informed consent, implementation of the protocol, definition and collection of data elements from clinical practice, and generating the initial clinical research questions.

Clinician engagement is critical to all successful interventional research, but engagement may be particularly difficult in pragmatic clinical trials,²⁹ because of a clinician's focus on the complexities of clinical practice rather than research participation, as well as the financial implications of time not spent on clinical activities.³⁰ Clinician engagement can be influenced by the clinician (e.g., belief in the importance of the study, appropriate training, and communication), clinical site (e.g., utilizing effective study champions, staffing, flexibility to accommodate clinical processes, and priorities), and health system leadership (e.g., culture of

support for research, incentives, and allowances for clinicians, as well as effective communication channels).^{31,32} One of the barriers to using EHR in clinical research is the lack of a clear incentive structure to encourage clinician involvement in EHR research, as well as little understanding of the necessary requirements to encourage effective clinician participation.³³

Approaches and future directions

There are examples of successful approaches to address the challenges we have outlined so far, but few of these approaches have spanned institutions. Some of the suggested and existing approaches are outlined in the sections below (Table 3).

Building EHRs to capture research-quality data

Since EHR data are not created for purposes of clinical research, numerous challenges arise from the structure, content, form, and completeness of these data. A solution would involve producing EHR data in a way that is more suitable for clinical research. Specifically, if clinical research was considered a co-primary use of EHR systems and data, research purposes could be considered in upstream engineering decisions about EHR structure and function. The current movement towards evidence-generating medicine³⁴ embodies the concept of building research considerations into the planning, organization, and function of healthcare systems.³⁵ Point-of-care trials are a research method that adapts the existing structure and function of EHR to recruit and randomize patients, as well as to ascertain outcomes.³⁶ In a point-of-care trial, research activities are incorporated into the everyday clinical work flow, and the ongoing results are immediately able to be implemented into practice. Clinician collaborations with professional EHR specialists have resulted in sets of EHR-derived data elements in specific disease areas that

include endpoints that are useful to both clinicians and researchers,³⁷ and that do not unduly increase collection burden. This type of collaborative approach can help both clinicians and researchers appreciate EHR as a tool for generating and using data for evidence-based clinical practice.

Developing empirical research about EHR data quality

Empirical research about EHR data quality is critical. For example, EHR data have been validated for completeness against claims data for the provision of diabetes and general preventative services.^{38,39} Concordance between EHR and claims data was higher depending on patient characteristics (e.g., continuous insurance coverage) and type of service (e.g., vaccination or cancer screening that are both covered by insurance do not need additional referrals).

Although data quality assessment frameworks exist,⁴⁰⁻⁴² there needs to be further study of how data from clinical trials compare to EHR data from the same patients. An increasing number of researchers have compared adjudicated clinical trials events to EHR data-ascertained events, reporting important differences between event rates, as well as relatively consistent agreement between the intervention effect measured in both data sources⁴³⁻⁴⁵ As there may also be differences in the quality of data in different therapeutic areas (e.g., cardiovascular outcomes vs. psychiatric outcomes), these comparisons are needed in a variety of patient populations. Encouraging groups conducting studies to make EHR data quality assessment a part of research reporting will generate more information on how quality may impact findings.^{8,46}

Sharing data

With the hope of using the diverse wealth of clinical data to maximize new health knowledge, several initiatives have been started to address the technical issues inherent to large-scale data sharing, including the Sentinel Initiative, PCORnet,⁴⁷ the Observational Health Data Sciences and Informatics (OHDSI) program, and a European initiative called the Electronic Health Records for Clinical Research (EHR4CR) project.^{48,49} In the context of many healthcare systems, these initiatives have found ways to securely perform the following tasks: 1) distribute queries from authorized researchers through network software; 2) execute the queries against the local data; and 3) return aggregated results to the researcher.^{47,50} Concurrent with interoperability efforts in technology and clinical data standards,⁵¹ these distributed research networks use common data models to harmonize codified elements to be included in datasets. In this way, these networks are able to link disparate data to increase the scale of each research study.

On a more local level, linkages between diverse data sources have occurred in the past on an ad hoc basis; however, the time-intensive process of establishing appropriate governance for the numerous privacy and information security concerns prevents the scaling up of this kind of data sharing. One response to this dilemma has been to establish honest broker systems, which, at their simplest, are methods allowing for reliable linkage of health data among clinical or research entities. An honest broker facilitates the sharing of data for both clinical and research purposes, maintains a master file of participant identities, and creates de-identified datasets for research.⁵²

The traditional culture among health researchers is only to make the findings from their research available to the public in a timely manner, but not the raw data; however, as mandated by many public granting agencies, access to raw data is increasingly being granted. This level of access creates the potential for other researchers to verify results, inquire into new research

questions, or perform meta-analyses using raw data from multiple studies. The access must be accompanied by proper oversight, so that the researchers new to the data understand its strengths and limitations. Nevertheless, this new culture of open data sharing may be hindered by motivations to control access to data, either for reasons of research productivity or commercialization and/or intellectual property potential.⁵³ These perspectives can be countered by gradually changing how researchers measure data impact, so that the creation of datasets is valued and rewarded.⁵⁴

Finally, empirical research is needed to understand how patients feel about their personal data being shared for research purposes. Retrospective research is often exempt from patient consent, due to the practical difficulty of contacting patients months or years after data creation, as well as the potential for selection of a non-representative population because of the exclusion of those patients who did not consent. Academic research is also considered separate from both quality improvement and public health activities, which are generally exempt from requiring patient consent and ethical evaluation. Yet as the boundaries become increasingly blurred between activities that traditionally require consent and those that do not, we need to better understand the patient's perspective and be able to describe to stakeholders the intended use of medical records.^{55,56} A greater awareness of the potential uses of health data, along with the care of individuals, could lead to alternative models of consent for the collection, storing, and sharing of data.

Building a culture of trust

To enable EHR use in clinical research, efficiently obtaining patient consent is necessary. Nevertheless, the operational aspects of this process must be built upon a culture of trust;

specifically, there must be a strong relationship between clinicians, health systems, patients, researchers, and participants. One way of building trust may be to increase patients' awareness of how their EHR data serve the public and common good. An example of such an approach comes from the United Kingdom (UK), where EHR data has been collated on a national level for decades and the general public has an understanding that secondary use of the data, including linkages to other sources, is for societal benefit.²⁶ The West of Scotland Coronary Prevention Study (WOSCOPS) was a randomized cardiovascular outcome trial of pravastatin that extended follow-up through the use of administrative data. The UK study survived many regulatory and ethical oversight changes and was able to publish 20-year results.⁵⁷ More recently, UK researchers examined patients' understanding of EHR and consent preferences, describing the distribution of preferences found among lay people and patients, and recommending awareness activities for specific target groups about the value and use of EHR in clinical research.⁵⁵

Research and education that is similar to what the UK has undertaken could be adapted for domestic use in order to increase the public value of secondary research and decrease the barriers related to consent. Yet given the role of private and for-profit entities within U.S. healthcare, there are unique challenges to building an environment that supports trust among all stakeholders, patients, clinicians, healthcare systems, drug and device companies, and EHR data system vendors. Transparency about profits resulting from patients' health data would increase the potential for trust.⁵⁸ Trust can also be facilitated by implementing technological solutions that preserve patient privacy, such as distributed networks that use data only under the protection of institutional firewalls.⁴⁷

Operationally, eliciting the right consent for research projects at the correct time may mean building processes within the structure of EHR data that capture patient consent, ideally at

the time of registration with the health care system.⁵⁹ Currently, for some observational research using EHR data sources, such as the Sentinel Initiative, the need for individual patient consent has been waived as they were considered public health practice activities. However, other solutions would need to be established to elicit more specific consent for prospective or interventional research. Examples of EHR-integrated patient consent dashboards are demonstrating how patients can be active stewards of their own data, as well as controlling and managing consent and permissions.⁶⁰

Lastly, sharing with patients the clinical research results stemming from their involvement in EHR, face-to-face clinic visits that focus on engaging the patient in their care, and in-person contact with an individual who represents the research institution may all be ways of facilitating a culture of trust between patients and researchers.

Targeting clinician engagement

Several helpful observations emerge from research on the perceptions and preferences of clinicians with regard to their role in EHR-based research. Including clinicians in the research process from the point of conception through results dissemination may be a means to engage clinicians in EHR research. Learning from clinicians about their daily practice realities and preferences would allow for the development of mutually feasible study protocols. Providing the simultaneous support of site leadership is also important for clinicians to feel supported in their research roles.⁶¹

Unlike the traditional clinical trial model, where sites are reimbursed for their involvement, pragmatic trials using EHR and clinicians as the backbone of the data collection methods require mutually beneficial relationships.³¹ Non-financial incentives have been

recommended to engage clinicians in EHR studies including population management EHR interface tools that provide clinicians with a bigger picture of patient health and practice patterns,⁶² or certification programs for non-researcher clinicians who participate in research. Practical activities during a trial may include an introduction letter that outlines the benefits of participation, site visits to all facilities to talk with frontline workers, in-person training sessions of involved clinicians, a regular schedule of ongoing activities (such as regular coaching calls for site unit champions), and return visits to underperforming sites. Despite these recommendations, questions still exist about clinician engagement including what resources are needed to enable an ideal level of clinician engagement, what practices have been shown to work, and what methods can engage clinicians efficiently in larger-scale projects.

Conclusions

The challenges and approaches to integrating EHR data into medical research that are presented in this paper focus on the research endeavor, improving quality and accessibility to maximize the use of these data, and the experiences of the research participants (including patients and clinicians). Robust research results that originate from EHR data could be more generalizable to the types of patients that clinicians see every day. Nevertheless, for these results to actually impact clinical care, we will need to build upon the constantly evolving discussion about the conditions and requirements under which EHR research results would meet regulatory standards.^{63,64} When randomization can be embedded within the care system and EHR data is of sufficient quality with complete outcomes, then a number of stakeholders will be able to make decisions based on the results. To ensure successful use of EHR data, trial data would still need to be evaluated according to data integrity standards, accuracy, and provenance in order to meet

regulatory framework goals. Addressing the challenges outlined in this paper can help ensure that future clinical research is more efficient, affordable, and relevant to clinical decision-making. Ultimately, finding solutions to these challenges will secure our continued progress towards improving the health and care of the populations we serve.

DRAFT

Acknowledgments

Conflict of Interest Disclosures

S Raman: Dr. Raman reports grants from GSK, outside the submitted work.

LH Curtis: Dr. Curtis has nothing to disclose.

T Andersson: Dr. Andersson is a full-time employee of AstraZeneca.

M Rocca: Dr. Rocca has nothing to disclose.

ED Peterson: Dr. Peterson reports grants and personal fees from AstraZeneca, grants and personal fees from Bayer Corp. US, grants and personal fees from Janssen Pharm., grants and personal fees from Merck & Co., grants from Regeneron, and grants and personal fees from Sanofi, outside the submitted work.

N Stockbridge: Dr. Stockbridge has nothing to disclose.

R Temple: Dr. Temple has nothing to disclose.

S Wasserman: Dr. Wasserman is a full-time employee of Amgen, Inc.

J Ezekowitz: Dr. Ezekowitz has nothing to disclose.

S James: Dr. James reports grants and personal fees from Astra Zeneca, grants and personal fees from The Medicines Co, grants and personal fees from Abbot, grants and personal fees from Boston Sc, grants and personal fees from Jansen, outside the submitted work.

K Marsolo: Dr. Marsolo has nothing to disclose.

B Sethuraman: Dr. Sethuraman is a full-time employee of Abbott.

S Terry: Dr. Terry has nothing to disclose.

I Ford: Dr. Ford has nothing to disclose.

P Miraji: Dr. Miraji has nothing to disclose.

R Rothman: Dr. Rothman reports personal fees from Edlogics, and personal fees from Boehringer Ingelheim, outside the submitted work.

AF Hernandez: Dr. Hernandez reports grants and personal fees from AstraZeneca, personal fees from Bayer, personal fees from Boston Scientific, grants from Luitpold, grants and personal fees from Merck, grants and personal fees from Novartis, and grants from GlaxoSmithKline, outside the submitted work.

Sources of Funding

Funding support for the meeting was provided through registration fees from AstraZeneca, Amgen, Janssen Pharmaceuticals, J&J Medical Device Companies, Pfizer, St. Jude Medical, Inc., The Medicines Company.

References

1. Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform* 2008;77:291–304.
2. Institute of Medicine. *Key Capabilities of an Electronic Health Record System: Letter Report*. Washington, DC: The National Academies Press; 2003.
3. Henry J, Pylypchuk Y, Searcy T, et al. Adoption of Electronic Health Records Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015. The Office of the National Coordinator for Health Information Technology web site. <https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php>. Published May 2016. Accessed October 11, 2017.
4. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Using real world data for coverage and payment decisions; the ISPOR Real World Data Task Force Report. ISPOR web site. https://www.ispor.org/workpaper/RWD_TF/RWTFManuscript.pdf. Accessed October 10, 2017.
5. Bayley KB, Belnap T, Savitz L, et al. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. *Med Care* 2013;51:S80–6.
6. Ateya MB, Delaney BC, Speedie SM. The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials. *BMC Med Inform Decis Mak* 2016;16:1.

7. Miotto R, Weng C. Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. *J Am Med Inform Assoc* 2015;22:e141–50.
8. Haneuse S. Distinguishing selection bias and confounding bias in comparative effectiveness research. *Med Care* 2016;54:e23–9.
9. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* 2013;51:S30–7.
10. Dixon BE, Gibson PJ, Frederickson Comer K, et al. Measuring population health using electronic health records: exploring biases and representativeness in a community health information exchange. *Stud Health Technol Inform* 2015;216:1009.
11. Romo ML, Chan PY, Lurie-Moroni E, et al. Characterizing adults receiving primary medical care in New York City: implications for using electronic health records for chronic disease surveillance. *Prev Chronic Dis* 2016;13:E56.
12. Yu TC, Zhou H. Benefits of applying a proxy eligibility period when using electronic health records for outcomes research: a simulation study. *BMC Res Notes* 2015;8:229.
13. Snyder CF, Jensen RE, Segal JB, et al. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. *Med Care* 2013;51:S73–9.
14. Gaist D, Sørensen HT, Hallas J. The Danish prescription registries. *Dan Med Bull* 1997;44:445–8.
15. Kimura M, Croll P, Li B, et al. Survey on medical records and HER in Asia-Pacific region: languages, purposes, IDs and regulations. *Methods Inf Med* 2011;50:386–91.
16. James S, Rao SV, Granger CB. Registry-based randomized clinical trials--a new clinical trial paradigm. *Nat Rev Cardiol* 2015;12:312–6.

17. Jick SS, Kaye JA, Vasilakis-Scaramozza C, et al. Validity of the general practice research database. *Pharmacotherapy* 2003;23:686–9.
18. HealthIT.gov. How to attain meaningful use. HealthIT.gov web site. <https://www.healthit.gov/providers-professionals/how-attain-meaningful-use>. Updated January 15, 2013. Accessed October 10, 2017.
19. Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *N Engl J Med* 2010;363:501–4.
20. Madden JM, Lakoma MD, Rusinak D, et al. Missing clinical and behavioral health data in a large electronic health record (EHR) system. *J Am Med Inform Assoc* 2016;23:1143–49.
21. Dusetzina SB, Tyree S, Meyer AM, et al. *Linking Data for Health Services Research: A Framework and Instructional Guide*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014. National Center for Biotechnology Information (NCBI) web site. <https://www.ncbi.nlm.nih.gov/books/NBK253313/>. Accessed October 10, 2017.
22. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. *Am Heart J* 2009;157:971–82.
23. Hripcsak G, Bloomrosen M, Flatley, et al. Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. *J Am Med Inform Assoc* 2014;21:204–11.
24. Xafis V. The acceptability of conducting data linkage research without obtaining consent: lay people's views and justifications. *BMC Med Ethics* 2015;16:79.
25. Wendler D. “Targeted” consent for pragmatic clinical trials. *J Gen Intern Med* 2015;30:679–82.

26. Hill EM, Turner EL, Martin RM, et al. "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. *BMC Med Res Methodol* 2013;13:72.
27. Kho ME, Duffett M, Willison DJ, et al. Written informed consent and selection bias in observational studies using medical records: systematic review. *BMJ* 2009;338:b866.
28. Petersen C. The future of patient engagement in the governance of shared data. *EGEMS (Wash DC)* 2016;4:1214.
29. Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016;375:454–63.
30. Rendell JM, Merritt RD, Geddes JR. Incentives and disincentives to participation by clinicians in randomised controlled trials. *Cochrane Database Syst Rev* 2007:MR000021.
31. Johnson KE, Tachibana C, Coronado GD, et al. A guide to research partnerships for pragmatic clinical trials. *BMJ* 2014;349:g6826.
32. Larson EB, Tachibana C, Thompson E, et al. Trials without tribulations: Minimizing the burden of pragmatic research on healthcare systems. *Healthc (Amst)* 2016;4:138–41.
33. Embi PJ, Tsevat J. Commentary: the relative research unit: providing incentives for clinician participation in research activities. *Acad Med* 2012;87:11–4.
34. Embi PJ, Payne PR. Evidence generating medicine: redefining the research-practice relationship to complete the evidence cycle. *Med Care* 2013;51:S87–91.
35. Morrison Z, Fernando B, Kalra D, et al. National evaluation of the benefits and risks of greater structuring and coding of the electronic health record: exploratory qualitative investigation. *J Am Med Inform Assoc* 2014;21:492–500.

36. Fiore LD, Brophy M, Ferguson RE, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin Trials* 2011;8:183–95.
37. Yazdany J, Robbins M, Schmajuk G, et al. Development of the American College of Rheumatology's Rheumatoid Arthritis Electronic Clinical Quality Measures. *Arthritis Care Res (Hoboken)* 2016;68:1579–90.
38. Heintzman J, Bailey SR, Hoopes MJ, et al. Agreement of Medicaid claims and electronic health records for assessing preventive care quality among adults. *J Am Med Inform Assoc* 2014;21:720–4.
39. Devoe JE, Gold R, McIntire P, et al. Electronic health records vs Medicaid claims: completeness of diabetes preventive care data in community health centers. *Ann Fam Med* 2011;9:351–8.
40. Kahn MG, Raebel MA, Glanz JM, et al. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care* 2012;50:S21–9.
41. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013;20:144–51.
42. Weiskopf NG, Hripesak G, Swaminathan S, et al. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform* 2013;46:830–6.
43. Hlatky MA, Ray RM, Burwen DR, et al. Use of Medicare data to identify coronary heart disease outcomes in the Women's Health Initiative. *Circ Cardiovasc Qual Outcomes* 2014;7:157–62.

44. Kjølner E, Hilden J, Winkel P, et al. Agreement between public register and adjudication committee outcome in a cardiovascular randomized clinical trial. *Am Heart J* 2014;168:197–204.e1–4.
45. Barry SJ, Dinnett E, Kean S, et al. Are routinely collected NHS administrative records suitable for endpoint identification in clinical trials? Evidence from the West of Scotland Coronary Prevention Study. *PLoS One* 2013;8:e75379.
46. Benchimol EI, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement: methods for arriving at consensus and developing reporting guidelines. *PLoS Med* 2015;10:e0125620.
47. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff (Millwood)* 2014;33:1178–86.
48. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;216:574–8.
49. De Moor G, Sundgren M, Kalra D, et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform* 2015;53:162–73.
50. Fleurence RL, Curtis LH, Califf RM, et al. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014;21:578–82.
51. Hazlehurst BL, Kurtz SE, Masica A, et al. CER Hub: An informatics platform for conducting comparative effectiveness research using multi-institutional, heterogeneous, electronic clinical data. *Int J Med Inform* 2015;84:763–73.

52. Boyd AD, Saxman PR, Hunscher DA, et al. The University of Michigan Honest Broker: a Web-based service for clinical and translational research and practice. *J Am Med Inform Assoc* 2009;16:784–91.
53. Longo DL, Drazen JM. Data sharing. *N Engl J Med* 2016;374:276–7.
54. Kratz JE, Strasser C. Comment: making data count. *Sci Data* 2015;2:150039.
55. Riordan F, Papoutsis C, Reed JE, et al. Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. *Int J Med Inform* 2015;84:237–47.
56. Page SA, Manhas KP, Muruve DA. A survey of patient perspectives on the research use of health information and biospecimens. *BMC Med Ethics* 2016;17:48.
57. Ford I, Murray H, McCowan C, et al. Long-term safety and efficacy of lowering low-density lipoprotein cholesterol with statin therapy: 20-year follow-up of West of Scotland Coronary Prevention Study. *Circulation* 2016;133:1073–80.
58. Haynes K, Selvam N, Cziraky MJ. Bidirectional data collaborations in distributed research. *EGEMS (Wash DC)* 2016;4:1205.
59. Marsolo K, Corsmo J, Barnes MG, et al. Challenges in creating an opt-in biobank with a registrar-based consent process and a commercial EHR. *J Am Med Inform Assoc* 2012;19:1115–8.
60. Kim H, Bell E, Kim J, et al. iCONCUR: informed consent for clinical data and bio-sample use for research. *J Am Med Inform Assoc* 2016;24:380–7.
61. Moloney RM, Tambor ES, Tunis S. Clinician Engagement in the NIH Health Care Systems Collaboratory. *Value Health* 2016;19:A284.

62. Marsolo K, Margolis PA, Forrest CB, et al. A digital architecture for a network-based learning health system: integrating chronic care management, quality improvement, and research. EGEMS (Wash DC) 2015;3:1168.
63. Berger M, Daniel G, Frank K, et al. A framework for regulatory use of real-world evidence. Duke Margolis Center for Health Policy web site.
https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf. September 13, 2017. Accessed October 11, 2017.
64. Food and Drug Administration (FDA). Use of electronic health record data in clinical investigations: guidance for industry. FDA web site.
<https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM501068.pdf>. Published May 2016. Accessed October 11, 2017.

Table 1. International EHR examples

<p>Denmark: All Danish citizens have a unique personal identification number; therefore linkage at the individual level between a number of long-standing country-wide registries is possible. In addition to a cause of death registry, health care (hospital care, prescription (dispensed), and condition specific (heart, diabetes, cancer) registries are available. Danish registries on economic and social issues include registries containing data about education, employment, income, and housing. Datasets are stored at Statistics Denmark, from which individual level data are available for online access by Danish institution affiliated researchers.¹⁴</p>
<p>Singapore: The national electronic health record (NEHR) uses a unique national identifier number to link longitudinal data from registries and administrative and medical records at hospitals and clinics to create a single health care record for each person. Clinicians have real-time access to a view-only summary of the most recent interactions with the health care system, including clinical events, investigation reports, alerts, emergency and hospitalizations, medications and allergies. Access to additional detailed reports is possible using a record locator system. Integration of legacy systems into the NEHR started in 2011, and will grow to support services such as the call centers, telehealth, and personal health management. Research using the NEHR has begun.¹⁵</p>
<p>Sweden: Sweden's electronic health data collection system is facilitated by a unique identifier for each resident. This unique number allows linkage between a number of mandatory Swedish registers (population registry, census, education registry, prescription registry, in-patient and out-patient health care registry, cancer registry, cause of death registry). In addition to the mandatory patient registries there are a number of clinical registries that aim to measure and enhance quality care. For example, the SWEDEHEART registry is a national health care registry for heart disease, reporting data about baseline characteristics, procedures, and outcomes of patients hospitalized for coronary and valve care. Various features facilitate the online entry of quality data, including data checks, auto-populated fields and calculated variables. In addition, automatic linkage with the population registries allows data about name and sex, and the personal identification number allows for complete longitudinal follow-up. These EHR data are being used for traditional epidemiological research as well as registry based trials.¹⁶</p>
<p>UK: The National Health Service in the United Kingdom has had a long history of the use of</p>

electronic patient care records, especially at the primary care level. Most of the population is registered with a single general practice, which serves as the gatekeeper to the rest of the system. Most prescription and hospitalization data are gathered and added to the record. This has allowed the construction of research datasets from the data of participating general practices,¹⁷ and recently more integration of data from primary, secondary and tertiary care.

DRAFT

Table 2. Research implications of EHR data characteristics

	EHR data feature	Research implications	Applicable for randomized trials	Applicable to observational studies
Characterizing the correct population	<ul style="list-style-type: none"> • Population coverage/definition • Difficulty in defining entry and exit from population 	<ul style="list-style-type: none"> • Extent of external generalizability • Affects available study time and capture of baseline characteristics and outcomes • Affects measurement of long term health states 	✓	✓
Characterizing the interventions and outcomes - what types of data are missing?	<ul style="list-style-type: none"> • Data recorded for clinical purposes • Clinical decision making process not well reflected • Patient reported outcomes, social and psychological data are not included 	<ul style="list-style-type: none"> • Does not capture sufficient detail at the correct time to estimate exposure/outcome /confounders • Difficult to evaluate data errors, determine causal mechanisms or generate next research question or appropriate intervention • Research results may be systematically biased due to missing subjective, self-report or preference data 	✓	✓
Characterizing the interventions	<ul style="list-style-type: none"> • Fragmentation of EHR data across providers 	<ul style="list-style-type: none"> • Risk of ascertainment and misclassification bias • Multiple data sources may 	✓	✓

<p>and outcomes – are the data included complete, correct and accessible?</p>	<ul style="list-style-type: none"> • Lack of interoperability between systems and locked narrative text (missing data) • Completeness of data types and fields 	<p>be needed for validation</p> <ul style="list-style-type: none"> • Missingness can be at random or informative, methods to mitigate have not been established • Range of exposures, outcomes and confounders of interest may be limited. Effect estimates may be biased, and this bias may differ by condition and type of data used. 		
---	--	---	--	--

Table 3. Challenges and approaches to the use of EHR for clinical research

	Challenges	Approaches
Data quality	<ul style="list-style-type: none"> • For available data – concerns about incompleteness and missingness • Selection bias – who is represented in EHR data during what time period • Lack of meaningful patient-reported data, lack of documentation about decision making 	<ul style="list-style-type: none"> • Integrate research functions into the EHR to improve quality data – evidence generating medicine paradigm • Prioritize research about EHR data quality– treat process and regulations as empirical questions
Data sharing	<ul style="list-style-type: none"> • Clarity about elements that are needed for data sharing include meaning and usefulness of the data, and governance requirements • Lack of understanding about patient perspectives of data sharing 	<ul style="list-style-type: none"> • Use distributed research networks; PCORnet, Sentinel • Development of scalable methods to use identifiers to link disparate data while preserving privacy • Create incentives for data sharing • Research about patient consent preferences in data sharing/linking
Patient consent and involvement	<ul style="list-style-type: none"> • Uncertainty about obtaining patient consent for ongoing secondary use of data • Ownership and access to patient data • How to contact and engage patients about the use and 	<ul style="list-style-type: none"> • Increase value of research as a public good (international examples) • Build a trust environment • Demonstrate transparency about profit over time from patients’ data • Utilize technological solutions to

	results of their data	preserve privacy
Clinician engagement	<ul style="list-style-type: none"> • Clinician engagement is limited by time, money, and interest • Lack of an incentive structure to encourage clinician engagement 	<ul style="list-style-type: none"> • Build a common ground between the goals of clinician and researchers - non-financial incentives such as population management tools/clinical interface, or certification programs for participation in research • Share examples of sustainable methods to engage clinicians

DRAFT