

Examining Researcher Needs and Barriers for using Electronic Health Data for Translational Research

Kari A. Stephens, PhD^{1,2,3}, E. Sally Lee, PhD², Hossein Estiri, PhD², Hyunggu Jung, MS³

Department of Psychiatry & Behavioral Sciences¹; Institute of Translational Health Sciences²; Department of Biomedical Informatics & Medical Education³
University of Washington, Seattle, WA

Abstract

To achieve the Learning Health Care System, we must harness electronic health data (EHD) by providing effective tools for researchers to access data efficiently. EHD is proliferating and researchers are relying on these data to pioneer discovery. Tools must be user-centric to ensure their utility. To this end, we conducted a qualitative study to assess researcher needs and barriers to using EHD. Researchers expressed the need to be confident about the data and have easy access, a clear process for exploration and access, and adequate resources, while barriers included difficulties in finding datasets, usability of the data, cumbersome processes, and lack of resources. These needs and barriers can inform the design process for innovating tools to increase utility of EHD. Understanding researcher needs is key to building effective user-centered EHD tools to support translational research.

Introduction

The Institute of Medicine has advocated for the creation of a Learning Health Care System that integrates use of electronic health data seamlessly into practice and research¹. We are far from this achievement. However, electronic health record systems (EHRs) continue to grow in adoption^{2,3} and have become an integral part of translational research in part due to the Health Information Technology Act (HITECH) Act of 2009 and the focus on funding of large data sharing architectures (i.e., PCORNet, MiniSentinel, eMERGE)⁴⁻⁸. The field of data science is also advancing methods and tools to efficiently translate large complex datasets into actionable information (i.e., predictive analytics, data visualization), targeting innovations with EHR data⁹.

Existing tools for researchers have been developed with functionality for cohort discovery, data queries across federated data systems, and rudimentary data profiling¹⁰⁻¹³. However, many developed tools have not capitalized on user-centered design methods to ensure needs of researchers, the primary users of these tools, are met. User-centered design approaches hold great promise for building effective tools that are driven by the concrete needs of the users¹⁴⁻¹⁶. Utilizing user-centered design methods that include qualitative methods can help capture in-depth perspectives of the users and their needs, which can then be translated into tool design.

A qualitative study was conducted through semi structured interviews to understand the processes that different researcher often follow to access electronic health data (EHD). Researchers were recruited from two groups: 1) researchers with experience using an existing tool that enables cohort estimation across a de-identified institutional repository of EHD captured across systems of care in an academic healthcare setting, and 2) researchers with experience and / or interest in using EHD for research locally and nationally. This exploratory study targeted understanding experience with using EHD for research, namely search channels used to access the data and both positive and negative experiences associated with finding and using EHD resources.

Methods

We approached 33 participants, with 22 participants volunteering to be interviewed using semi-structured interview guides. Participants represented a diverse group of data users across gender (16 female), years of experience (range 2-30+ years), occupation (medical residents, graduate students, junior and senior faculty, and research scientists and statisticians), research areas of expertise (i.e., infectious disease, pharmacogenomics, mental health, health services), and location (local and nationally based researchers). Participants had experience conducting basic science, clinical trial based, and dissemination research across the T1 to T4 spectrum of translational science. Each interview was guided by structured questions addressing general needs and barriers of doing research using EHD and all analyses were conducted after the interviews were completed to avoid leading of the participants inadvertently. All interviews lasted between 20 and 45 minutes, were audio taped and transcribed for coding, and conducted in person or by phone. The interview transcripts were analyzed using qualitative analysis steps guided by grounded theory¹⁷. Emerging base themes were identified and iteratively refined individually by the first two of authors. The themes were then triangulated and grouped into higher categories by both researchers.

Results

Base themes were organized into two broad categories related to use of EHD for research: 1) needs researchers

express to successfully use EHD, and 2) barriers researchers face when using EHD for research. Sub themes were grouped under these two base themes and definitions and key quotes are provided in the tables below.

Table 1. The needs researchers expressed that will help them successfully use electronic health data.

Sub Theme	Definition	Key Representative Quotes
Confidence in the data		
Confidence that the data are good and worth using	Researchers need to feel confident that the data source they use is trustworthy and worth pursuing, they often rely on word-of-mouth or published works	"...And then once we identified the data sets we then go searching for more information on the web or other papers about them... what we do is we contact other investigators in that area and ask them for recommendations of what they've used..."
Easy access to data		
Resources to locate data easier	Resources needed to support easy access included mentors, contacts, literature searches - with quick and easy on-line access with options to speak quickly to individual experts	"...you hear about these things through peers, colleagues, and...going to national meetings where you hear about the use of the data set. Or you read about a study so you read a publication, of manuscripts that's been published that used this data set."
Usable data search tools	Data resources for searches need to be free of jargon, easy, and friendly to use, with intuitive information and user interfaces	"...I wish there was some type of ... inventory... some portal you could go to where you could kind of type in what key variables of interest or general key words about what you're trying to find and then it would pop up... These are the five data sets that... may be appropriate and then here's very clear information about how to find out more..."
Detailed descriptions of data	Access to data dictionaries are important to search available data before making a specific request	"I think those are incredibly helpful, definitely, to have a data dictionary of some sort, especially if you're working across systems and you wanna be sure that, you know, how somebody defines a diabetic is the same as..."
Easy ability to search for appropriate cohorts	Researchers need to be able to do rudimentary searches to determine if the data repository will be suitable for their research; often cohort finding	"I'd say it's very helpful to get a sense of just kind of raw numbers, like in terms of the feasibility of a research project... it's really a good launching point."
Clear data access process		
Access to clear process for data access	A clear process must be easy to access and understand for requesting and receiving the data	"And then people intend to go through the IRB process, but they're busy, it takes a while... So I think as simple as that process can be made, it's only going to make things easier and more eager to use the database."
Resources		
Help from data experts	Data experts easily accessible are needed to iterate questions, clarify data content, and help researchers formulate appropriate data requests	"There's nothing to take the place of just having a consultation... there's a lot of imprecision...because people don't code it appropriately, there's a lot of miscoding..."
Funding for data access or free data resource	Affordable ways are needed to access the data sets for initial searches and data extracts, particularly for junior investigators	"...if there was some kind of essential funding at the university level [that] would kind of support some of the data management or data sets or other sources of data... that would help defer the costs to specific units of accessing that data
Training on data tools	Need adequate training resources on-line to be able to use search tools	"I liked the training a lot and I liked evaluating, or seeing the training video before. I think it was very good."

Table 2. The barriers researchers face using electronic health data.

Sub Theme	Definition	Key Representative Quotes
Data are not usable for research		
Difficult to get data out	Systems to explore or extract data have technical barriers and it becomes difficult to get the datasets	"...you have to use Sequel to get it out. It's not friendly, not the slightest."
Lack of trust in data quality	The data quality is too poor or undiscoverable to allow use of the data	"But the data is very messed up. I mean...I can't tell you. Right now I'm cleaning like about 600,000 records from...diagnoses and it's just like...there's no quality in what they put in because everyone types into it...nurses, doctors, staff."
Data complexity	Researchers have a hard time understanding the complexity of the data, limiting the utility	"...how to look for the right term and find the right test without being confused sometimes. We don't know if we are really looking for the right test or not, even we have like the same terms."
Data content is unsuitable for research	The data have idiosyncrasies due to how they were collected that make them difficult to use for research or they are only in text format	".....that using...electronic clinical data secondarily for research use, it's basically not ready for prime time. So we have data that are collected for clinical purposes and recorded for clinical purposes in certain ways, and when we try to extract or abstract those data for use in data sets for research purposes..."
Lack of resources		
No experts available	No person is available to quickly answer critical path questions to explore if a dataset is worth using	"They don't have enough people with your type of training and your knowledge..."
Lack of funding / lack of time	Expense of getting the data extracted is too high and / or it takes too long	"I'm always thinking because our fellows and residents and faculty are so busy, they work long hours, so..." "That's great. That's huge for those of us like me who don't have a lot of extra research funding..."
Access to data is not available		
Data cannot be extracted	Researchers recognize the data are collected, but no system exists to be able to pull the data out to be used after they are collected	"I mean there's some very large databases out there that someone like me just doesn't have access to, so for me this is an accessible database to go back and try to when some of these questions come up, we always thought to ourselves, oh, if only we had a database for that, if only there was some way to look at that."
Researchers cannot find datasets		
No clear path to finding datasets	Researchers have a hard time easily finding existing and available secondary data sets suitable for their research	"So you know, there's information out there through papers and websites and some of the websites are very up-to-date and other ones are not and so it's hard to really know, kind of what's in there."
Cumbersome process		
Difficult data approval and access process	Often extracting and using secondary datasets involves a very cumbersome and long approval and access process	"...in the data set without having to go through getting in touch with the person and they email you back. Wait a week 'cuz they're on vacation and... you have to sign a data release form and send that in and... a month later the committee meets to approve your data release form and...by then you've moved on to another project, and you don't care, you know?"

Discussion

The distilled sub themes revealed various needs and barriers expressed by researchers that can be used as a base

to provide useful guidance for developing tools and methods for supporting the use of EHD in translational research. The needs spanned from concrete technical tools and resources to clear processes and subjective evaluations of confidence in datasets. These diverse needs provide guidance for specific tool functions like user search interfaces. But on a more complex level, they provide guiding principles creating actionable directions to improve use of electronic health data for translational research and developing functions and processes supported by the tools (see Table 3). For example, having processes clearly in place as well having the tool support walking researchers through exploring and accessing datasets led to improved ability to use EHD. However, building in functionality that supports easy access to these processes and inserts resources, like the ability to connect with an expert to iterate the data requests could also improve researchers' ability to use EHD. The barriers identified by researchers similarly spanned concrete technical issues of data usability and lack of search tools and resources to process issues. The barriers provide helpful guidance on pitfalls to avoid in developing new functionality, such as making researchers wade through too much jargon or assuming high technical skills, failing to track data quality to substantiate data utility, and failing to provide timely technical expertise to ensure the process continues to move forward for the researcher. The sub themes associated with the needs and barriers supported each other, specifically those related to resources and processes, emphasizing the risk of not meeting these needs as significant failure points in linking EHD with researchers. Many researchers stated the critical need to have data experts to iterate their data explorations and requests, and to answer critical path questions before committing to use of EHD for research. Lack of such resources often resulted in lack of traction in requesting a dataset and / or at times ending pursuit of a research idea itself. Providing an expert at the right time in the research process is critical and challenging both in terms of how best to connect the researcher to the expert as well as feasibility of maintaining this costly and rare resource for all researchers who make inquiries. Tools may be able to alleviate some of this need for human experts and also provide assistance with the processes to refine questions for a more efficient asynchronous interaction with an expert. The researchers also articulated the need for simple and clear processes, recounting experiences of bailing on use of EHD when processes became too cumbersome. Tools can be developed to support data access processes and improve efficiency of various steps involved in data access.

Table 3. Actionable directions for improving secondary use of electronic health data for translational research.
Clear processes must be defined, easily communicated, and efficient
Clear validity of the data and provenance must be available and easily accessed
Data expert availability must be easily accessible throughout the research project proposal development
Web based data profiling tools are needed to provide efficient, asynchronous exploration of available datasets

Participant comments also reflected that EHD sources need to have credibility and lack of confidence in the data was a key failure point. Participants emphasized discovering and developing confidence in EHD datasets most often occurred by word-of-mouth. This suggests that tools that emphasize successful use cases to inspire confidence and give instruction for best use can help offer the function that word-of-mouth currently provides. The research community may need to culturally shift to using online resources to find high quality EHD sources rather than relying mainly on word-of-mouth if the large data sharing infrastructures are going to succeed at attracting broad usage. Data profiling tools that address data complexity, data quality, and easy utility to do cohort searching will address many of the barriers and needs related to data usability and may enhance online data resources. Data visualization methods and iterative tracking of data harmonization issues as well as phenotyping algorithms can be added to data profiling tools to address data usability. Such methods help break complex information into consumable format without the reliance on scarce human resources or laborious explorations of the datasets.

Conclusion

The researcher needs and barriers identified in this study provide actionable user-centered data for improving tool design to support researchers' use of EHD for translational research. The distilled sub themes of needs and barriers spanned diverse topic areas and provide useful guidance for developers of EHD tools. User-centered design approaches can be used to generate scenarios, storyboards, and prototypes to support tool development. Tool functionality should target the needs while avoiding the pitfalls of the barriers, with emphases on promoting confidence in the data, providing and supporting clear efficient processes, integrating key experts into the process in a timely way, and incorporating useful data profiling to convey depth and breadth of data to support usability. Future tool development efforts that integrate researcher perspectives have potential to powerfully and effectively bridge EHD and researchers to support translational research innovation.

Acknowledgements

This publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000423. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. National Research Council. The Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine). Washington, DC; 2007. Available at: <http://www.iom.edu/Reports/2007/The-Learning-Healthcare-System-Workshop-Summary.aspx>. Accessed February 19, 2014.
2. Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *N. Engl. J. Med.* 2009;1-11. Available at: https://wp11.webpine.washington.edu/pub/getach.tcl/Jha,_DesRoches,_Campbell,_Donelan,_Rao,_Ferris,_Shields,_Rosenbaum,_and_Blumenthal,_2009.pdf?h=wY0HRqFppvyS2jgp5BpEgF5312vzsmnhtGnJP5z0RdestP25eBgVIL5U3sFXpUE4.
3. DesRoches CM, Charles D, Furukawa MF, et al. Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012. *Health Aff. (Millwood)*. 2013;32(8):1478-85. doi:10.1377/hlthaff.2013.0308.
4. Furukawa MF, Patel V, Charles D, Swain M, Mostashari F. Hospital electronic health information exchange grew substantially in 2008-12. *Health Aff. (Millwood)*. 2013;32(8):1346-54. doi:10.1377/hlthaff.2013.0010.
5. Pathak J, Wang J, Kashyap S, et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. *J. Am. Med. Inform. Assoc.* 18(4):376-86. doi:10.1136/amiajnl-2010-000061.
6. Califf RM. The Patient-Centered Outcomes Research Network: a national infrastructure for comparative effectiveness research. *N. C. Med. J.* 75(3):204-10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24830497>. Accessed August 24, 2014.
7. Psaty BM, Breckenridge AM. Mini-Sentinel and regulatory science--big data rendered fit and functional. *N. Engl. J. Med.* 2014;370(23):2165-7. doi:10.1056/NEJMp1401664.
8. Ross MK, Wei W, Ohno-Machado L. "Big data" and the electronic health record. *Yearb. Med. Inform.* 2014;9(1):97-104. doi:10.15265/IY-2014-0003.
9. Hrovat G, Stiglic G, Kokol P, Ojstersek M. Contrasting temporal trend discovery for large healthcare databases. *Comput. Methods Programs Biomed.* 2014;113(1):251-7. doi:10.1016/j.cmpb.2013.09.005.
10. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J. Am. Med. Informatics Assoc.* 2010;17(2):124-130.
11. Kellam SG, Werthamer-Larsson L, Dolan LJ, et al. Developmental epidemiologically based preventive trials: baseline modeling of early target behaviors and depressive symptoms. *Am. J. Community Psychol.* 1991;19(4):563-584. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1755436.
12. Klann JG, Buck MD, Brown J, et al. Query Health: standards-based, cross-platform population health surveillance. *J. Am. Med. Inform. Assoc.* 21(4):650-6. doi:10.1136/amiajnl-2014-002707.
13. Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. *Med. Care* 2012;50 Suppl:S60-7. doi:10.1097/MLR.0b013e318259bff4.
14. Turner AM, Reeder B, Ramey J. Scenarios, personas and user stories: User-centered evidence-based design representations of communicable disease investigations. *J. Biomed. Inform.* 2013;46:575-584. doi:10.1016/j.jbi.2013.04.006.
15. Reeder B, Turner A. Scenario-based design: A method for connecting information system design with public health operations and emergency management. *J. Biomed. Inform.* 2011;44(6):978-88. doi:10.1016/j.jbi.2011.07.004.
16. Chang Y, Lim Y, Stolterman E. Personas : From Theory to Practices. In: *NordiCHI '08 Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges.*; 2008:439-442. doi:10.1145/1463160.1463214.
17. Patton MQ. *Qualitative Research and Evaluation Methods*. 3rd Editio. Thousand Oaks, CA: Sage Publications; 2002.