Title

Patient and researcher stakeholder preferences for use of electronic health record data: a qualitative study to guide the design and development of a platform to honor patient preferences.

Journal

Journal of the American Medical Informatics Association, 30(6)

ISSN

1067-5027

Authors

Morse, Brad
Kim, Katherine K
Xu, Zixuan
et al.

Publication Date

2023-05-19

DOI

10.1093/jamia/ocad058

Patient and researcher stakeholder preferences for use of electronic health record data: a qualitative study to guide the design and development of a platform to honor patient preferences

Authors:

Brad Morse, PhD, MA[1]; Katherine K. Kim, PhD[2]; Zixuan Xu, MS[2]; Cynthia G. Matsumoto, PhD, RN-BC[2]; Lisa M. Schilling, MD, MSPH[1]; Lucila Ohno-Machado, MD, PhD, MBA[3]; Selene Mak, PhD;[4] Michelle S. Keller, PhD, MPH[5,6]

1 Division of General Internal Medicine, Data Science to Patient Value Initiative, University of Colorado - Anschutz Medical Campus, Denver, Colorado, USA
2 University of California-Davis, Davis, California, USA
3 Department of Biomedical Informatics, University of California-San Diego, USA
4 Evidence Synthesis Program at the VA Greater Los Angeles Healthcare System, Los Angeles, California, USA
5 Division of General Internal Medicine-Health Services Research, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA
6 Division of Informatics, Department of Biomedical Research, Cedars-Sinai Medical Center, Los Angeles, California, USA

Corresponding Author:

Brad Morse
Assistant Professor

University of Colorado Denver - Anschutz Medical Campus

1635 Aurora Ct 5th Floor

Aurora, CO 80045

Phone: 303-724-1147

Email: BRAD.MORSE@CUANSCHUTZ.EDU

Abstract word count: 250 (250 max)

Word count (excluding title page, abstract, references, figures, and tables):

Patient and researcher stakeholder preferences for use of electronic health record data: a qualitative study to guide the development and design of a platform to honor patient preferences

ABSTRACT

Objective: This qualitative study aimed to understand patient and researcher perspectives regarding consent and data sharing preferences for research and a patient-centered system to manage consent and data sharing preferences.

Materials and Methods: We conducted focus groups with patient and researcher participants recruited from three academic health centers and via snowball sampling. Discussions focused on perspectives on the use of EHR data for research. Themes were identified through consensus coding, starting from an exploratory framework.

Results: We held two focus groups with patients (n=12 patients) and two with researchers (n=8 researchers). We identified two patient themes (1-2), one theme common to patients and researchers (3), and two researcher themes (4-5). Themes included (1) motivations for sharing EHR data, (2) perspectives of the importance of data-sharing transparency, (3) individual control of personal EHR data sharing, (4) how EHR data benefits research, and (5) challenges researchers face using EHR data.

Discussion: Patients expressed a tension between the benefits of their data being used in studies to benefit themselves/others and avoiding risk by limiting data access. Patients resolved this tension by acknowledging they would often share

their data but wanted greater transparency on usage. Researchers expressed concern about incorporating bias into datasets if patients opted out.

Conclusions: A research consent and data-sharing platform must consider two competing goals: empowering patients to have more control over their data and maintaining the integrity of secondary data sources. Health systems and researchers should increase trust-building efforts with patients to engender trust in data access and use.

Keywords: patient privacy, data sharing, design, transparency, secondary data

BACKGROUND AND SIGNIFICANCE

Secondary use of electronic health record (EHR) and administrative claims data, which can include demographics, clinical diagnoses, laboratory and imaging tests, and billing information, is often used for health-related research. In the U.S., researchers are often able to use such data with a waiver of informed consent under the Federal Policy for Protection of Human Subjects (a.k.a. the “Common Rule”). An Institutional Review Board (IRB) may approve a request to waive all or some of the elements of informed consent in cases where (a) the research involves minimal risk to subjects, (b) the research could not be carried out practicably without the waiver or alteration, (c) the waiver or alteration will not adversely affect the rights and welfare of the research participants, and (d), where appropriate, subjects are provided with additional information regarding their participation (e.g., in the case of research where the procedures include deception). This waiver of informed consent has allowed researchers to conduct large-scale analyses with millions of patient records without contacting individual patients; such a task would be otherwise prohibitively time-intensive and expensive research to conduct without a waiver. Several major advantages of secondary data for health-related research, which have led to ubiquity of its use, are the relative ease of access and size of the data sets that could not otherwise be collected.[1,2] For example, during the Covid-19 pandemic, researchers have used large secondary datasets to examine cardiovascular outcomes associated with SARS Cov-2 infections[3], health utilization associated with post-acute sequelae of Covid[4], and mortality risk prediction associated with Covid-19 infection.[5] Conducting these studies would have been infeasible without the use of such datasets and a waiver of informed consent.

Many patients are eager to share their data motivated by the desire to advance health-related research[6–9], yet patients have also expressed a desire to have more control over whether some or all of their data are shared with researchers.[10] Most health organizations present patients with the choice to opt-out of research to ensure more robust participation of research participants and to reduce potential selection bias. Researchers are therefore not allowed to access or use the medical record and related data of patients who opt-out of research. For patients who do not opt-out, researchers may have access to their records via chart reviews or electronic health record data queries after approval of their study by an IRB. Prior to releasing the study data, IRBs review the study and assess which data elements are required to perform the scientific analysis. An “Honest Broker” might also review the data request, ensuring privacy and confidentiality safeguards are met before providing the data.[11] Organizations and IRBs also have protections in place to limit breaches of confidentiality, including rules about how the data are accessed, stored, transmitted, shared, and destroyed. However, as many studies using secondary data rely on waivers of informed consent, patients often have little to no control over which researcher or institution accesses their data and how data are used for which types of studies.

Patients often do not know or control whether their data are shared only with researchers at the health system where they received care, at other outside institutions, or with private (e.g., health technology companies or pharmaceutical firms) or governmental institutions. Moreover, patients have little control over whether sensitive data, such as data related to stigmatizing conditions (e.g., mental

health and substance use), history of interpersonal violence, genetic data, immigration data, pregnancy terminations, and sexual orientation, are shared. These scenarios are more than hypothetical. Several high-profile examples have occurred in recent years demonstrating that health systems are partnering and sharing patients' EHR data with a variety of private institutions. In 2019, Google partnered with Ascension Health, which included 2,600 hospitals, doctors' offices, and other health care facilities.[12] Ascension Health shared the EHR data of tens of millions of patients to Google without patients' knowledge.[12] In 2022, the Mayo Clinic began assessing a Natural Language Processing product from Google to process EHR data more easily. Mayo shared the use of the personal health information from millions of patients with Google, along with at least 16 other companies eager to access data from the Mayo Clinic's EHR and apply insights from those data to the commercialization of digital products and services.[13,14] These cases demonstrate that across health systems inside the U.S., patients' data are actively being shared with outside institutions for a variety of use cases.

Several studies have examined patient preferences for sharing their EHR data, finding that patients would like more control over the types of data (e.g., demographics, clinical conditions, genetic information) shared and with whom their data are shared, and for what purpose.[15–17] In a previous study, we found that approximately 1 in 10 patients preferred certain types of EHR data not be shared with researchers, and 4 out of 10 patients preferred that their EHR data were only shared with the institution from which they received care or with non-profit institutions.[18] Moreover, individuals have expressed significant concerns about the privacy and safety of their EHR data and how it is shared with other institutions.

[10,19,20] However, it is critically important to balance these patient concerns with the benefits of using secondary data for biomedical research. Given these perspectives, we began developing a platform that empowers patients to make decisions about their preferences for data sharing, expedites delivery of data to researchers, and allows representatives (such as ethics or IRB experts) to closely monitor what happens with the data.

OBJECTIVE

The objectives of this study were to (1) understand patients' perspectives regarding consent and data sharing preferences for research and their needs for a patient-centered system to manage consent and data sharing preferences; (2) understand researcher perspectives and related needs; and (3) develop requirement and design recommendations for a data-sharing platform.[21] Importantly, given that the data sharing platform (iAgree) may give patients additional control over their data, we wanted to understand how and why patients and researchers viewed and resolved the competing goals of controlling/limiting data access and privacy and contributing data for the greater good/advancing science.

MATERIALS AND METHODS

Study design

We used a qualitative focus group study design with patients and researchers. We also asked both patients and researchers to react to low-fidelity mock-ups of the data-sharing platform to identify system requirements. The study was reviewed and

approved by the University of California-Davis IRB. A description of the study team is in the author's contribution justification.

Recruitment and sampling

Eligibility criteria included: patients, clinical or healthcare researchers, who used secondary data in epidemiological, clinical, or biomedical research; 18 years of age or older; and able to converse in English. To recruit patient participants, the qualitative study team asked a large multi-site study team of researchers from UC San Diego, UC Davis, Cedars-Sinai Medical Center, University of Colorado, University of Southern California, and UCSF to provide a list of potential contacts. These could include patients who had previously participated in research and agreed to be contacted again, personal contacts, and community board members affiliated with or who had worked with research teams. This list was used to identify a preliminary list of participants; once participants agreed to participate, they were asked to provide referrals of other potential participants (i.e., snowball sampling).[22,23] The same approach was used to recruit researcher participants. Potential participants were emailed a description of the study. Upon agreement to participate, the study staff sent a confirmation email and scheduled a time for the focus group. Each participant received a $45 gift card for their participation. Table 1 provides the characteristics of the participants. Demographics were not collected from seven patients.

Table 1. Research participant demographics

Participants	Patient = 5 (total 12)	Researcher = 8	Total = 13 (total 20)
Age			
30-49	0	4	4
50-46	1	3	4

65+	4	1	5
Gender			
Female	4	4	8
Male	1	4	5
Race			
White	2	4	6
Asian	0	2	2
Black/African American	3	2	5
Education			
High School Graduate	1	0	1
College Graduate	1	2	3
Postgraduate Degree	3	6	9

Focus group design

Focus groups were conducted separately for patients and researchers. The focus groups (See Appendix 1 and 2 for interview guides) were facilitated by two researchers (KK and CGM). The first section of the focus group for patients and researchers presented low-fidelity mockups (Figures 1-3) of the data-sharing platform to collect insights and preferences on signing in, setting default data-sharing preferences, and consenting to new studies. The second section of the focus groups focused on participants' experiences and challenges in EHR data collection, storage, sharing, and their perspectives on individual control over EHR data sharing, as well as system requirements for the data-sharing platform. The same set of topics were used in all focus groups.

In the patient participant focus groups, we first asked participants to share their research participation experience and any benefits and concerns they have about sharing their "medical records" for research. Next, participants were asked to share

their perspectives on giving patients more granular control over the sharing of patient data versus creating efficient processes for researchers to access data. Finally, we discussed issues surrounding the benefits of opting in or out to individuals and society at large, and patients' considerations when deciding whether to share their data for a study such as the type of organization accessing the data (e.g., private industry, academic center) or the type of data shared (e.g., potentially stigmatizing data).

In the researcher participant focus groups, we first asked participants to share their research experiences and benefits and challenges they face using secondary data for research. Next, participants were asked to share their perspectives on providing patients individual control over data-sharing preferences versus remaining with existing opt-out methods, which, while more efficient, may not always honor patient preferences. Specifically, we wanted to explore the tradeoffs that are made between patient granular control of the data and creating efficient and effective access to data sets for researchers, given that giving patients granular control over the sharing of their data could potentially have a negative effect on researchers wanting to access complete and unbiased data sets. Finally, we asked participants about their perceptions of how much control patients should have when deciding whether to share their data for research.

During the design considerations section, we walked patient participants through a series of low-fidelity mockups and asked participants about their thoughts on the log-in process, setting default data sharing preferences, and setting data sharing preferences for individual studies. During this portion of the focus group, we used a

semi-structured focus group approach, which allowed us to ask further questions based on participant feedback. Patients were show a variety of different mockup screens and were asked to provide their feedback to guide refinements. In terms of consenting to new studies, patient participants were asked to select between two interfaces: one, which showed fewer data sharing options over multiple pages (for example, each page include data sharing preferences for a different category, such as demographics, diagnoses, medications, etc.) and a second that displayed a grid structure that had all the data sharing preferences on one screen. Patient participants were also shown a final screen that would enable them to review their data that had been accessed by professional researchers. We asked questions about the usability of the platform, features (e.g., logging in with a social media account such as Facebook, Twitter, or Google), acceptability of the language used in the mockups, processes used (e.g., the use of two-step authentication), and display choices for sharing data. Focus groups were conducted using video-conferencing software (Zoom), were audio-recorded, and were professionally transcribed. To ensure confidentiality, all names were stripped from the transcripts, and participants were assigned a number in the transcripts.

Qualitative analysis

The focus group transcripts were analyzed both inductively and deductively. We started with *a priori* codes and added them to the codebook as we went through the consensus coding process, following an exploratory framework methodology with collaborative consensus coding.[24] Transcripts were coded for any issues, concerns, perspectives, or attitudes participants expressed. Two members from the

research team (BM, ZX) independently coded the first patient and researcher focus group transcripts. Subsequently, the coders came together and discussed their coding results until they reached consensus, and this was used to refine a preliminary codebook. Any disagreements between the two researchers were reviewed and resolved by a third researcher (KK). Then using the preliminary code book, researchers (BM/ZX) independently coded each of the remaining transcripts and met again to resolve discrepancies, discuss newly observed codes, and further refine the code book. After agreement between both researchers regarding the interpretive codes and their operational meaning, these two same researchers then identified the patterns and emergent themes that were discussed in both patient and researcher focus groups, such as ethical concerns, specific needs in data sharing, and perspectives on individual control. We aimed to reach conceptual saturation, in other words, when we felt that the themes had enough breadth of opinions, with enough depth to confirm that we correctly understood them. As the research was aimed at a narrow set of concepts, conceptual saturation was reached after the second focus group with each participant group. Codes were then categorized into main themes. Data for the thematic analysis were analyzed and organized using Dedoose (2019 Los Angeles, CA: SocioCultural Research Consultants, LLC www.dedoose.com). In line with usability and rapid prototyping studies, mockup data was rapidly analyzed without the use of thematic analysis, but rather to identify potential changes to the platform.

RESULTS

We conducted four focus groups, including two patient participant groups (N=12 patients total) and two researcher groups, (N=8 researchers total); each focus group was approximately 60 minutes. Due to data collection error, participant demographics were not correctly captured during the focus groups, and we are unable to report complete demographic data. Partial participant demographics are reported in Appendix 3. Due to data collection error, the demographics of the participants were not correctly captured during the focus groups, and we are unable to report complete demographic data. Partial participant demographics are reported in Appendix 3.

Focus Group Results

We identified five unique themes. Two themes unique to patients were: (1) *motivations for sharing data* and (2) *perspectives of the importance of data-sharing transparency. O*ne theme common to both patients and researchers was: (3) *Individual control of personal EHR data sharing*. Two unique themes to researchers: *(4) how EHR data benefits research, and (5) challenges researchers face using EHR data.*

We provide exemplar quotes in Table 2 and then summarize the results.

Table 2. Themes and sub-themes from patients and research stakeholders about the use of and considerations around electronic health records used for researcher

Theme	*Sub-Theme*	Stake-holder	Exemplar Quote
(1) Motivatio	***Personal Knowledge/***	Patien t	"For me, it was to have access to the data, to the information that would help me make decisions about my health and my

ns for sharing EHR data	***Personal Benefit***		healthcare.”
	Access to Medical Procedures /Resources	Patient	“I wanted the echocardiograms because the study protocol required these very frequent echocardiograms and that’s what was important to me.”
			“I participated in a study… for Elequazine, which was a drug specifically for dilated cardiomyopathy and hypotrophy cardiomyopathy and I was in that study, and I said before, the reason I often participate in these studies is so that I can have access.”
	Societal good	Patient	“As a patient and patient advocate, [I] recognize that I have a very rare cancer so I’ve sought out projects that were looking, studying people like me and then I’ve learned about studies across the country and then contacted them and they’ve done medical records requests to me.”
			“And so, if I can contribute to that, I feel like I’m not only helping me, but I’m helping other people that may not be able to participate because they don’t have a lot of time.”
			“I would say yes, because of the immediate feedback and the immediate sense of contribution, accomplishment that participating in a study such as this. Because it’s a worldwide epidemic. We’re not just talking about a particular neighborhood or a section of the country. And so that sense of benevolence and caring, I think would encourage me to do it and I would say yes, in that instance. But I perhaps would not do it in other healthcare conditions.”
			“In terms of all kinds of things and so I’m not going to send rockets to moon and I’m not gonna make world peace, but if my medical information can benefit someone, whether I know them or not, it’s kind of like donating a kidney kind of.
(2) Highlighting importance of data use transparency	***Clear communication about data use***	Patient	“I would, as long as the criteria for participation in terms of trust and transparency and how the data would be used, and security and things like that, if those were fulfilled, I would not see a difference between flu or cancer or a mammogram.”
			“I think asking for permission is important. Disclosure is important in terms of what the need is and how your data might fit.
			“I would want to understand how my data were gonna be used in a meaningful way for each thing.”
			“I think is extraordinarily important because there’s a lot of things, we no longer control because of the virtual world we live in.”
			“I shouldn’t even say control, but in the very least, I would like to know who has had access to my health data. And for what purpose? I would like to know that.”
			“And you know, again, it’s really about being clear about how they collect the data, how they store the data, who has access to the data, how they’ll use the data and if you personally feel that you have a way to get information back that is personally relevant and helpful as well as do the greater good to the community.
	Concerns	Patien	“And I just think what your medical information, not saying

	about data access	t	that it would be used for nefarious means, but there is definitely that fear of it getting bogged down into, you gave blanket consent for these type of organizations, but you didn't have full transparency what that really gave them the right to do with that information…that would be my fear…within the fine print, you gave them the option to do something else with that information." "I would like to know who has had access to my health data. And for what purpose? I would like to know that. I think some of the portals…you can look at a history of access. But whenever I looked at the history of access, all I see is my logins into the portal. So, it's not real – we all know that somebody's accessing that data within the hospital…it's kind of like what we do for credit, not financial, your credit score or credit history. So, I think something like that would be… important so that people have the ability to know who is using their data and how they're using it, for what purpose." "I had a lot of apprehension in terms of joining the study because of sharing information. They wanted DNA, they wanted stuff I guess historically, as people of color, there's an apprehension with sharing information, not knowing that's gonna come up. Who gets it? You know, is it gonna end up going to the insurance company that may increase my life insurance. Or, you know, are they even though I'm in a study for breast cancer, are they gonna use the information for something else that I don't know about? I mean to me, a lot of it is behind the scenes, so I don't really know who's getting it, when they're getting it and what they're doing with it." "We live in a world in which the healthcare system doesn't always do things on behalf of patients. Really it stems from a sort of mistrust that there – healthcare is a commodity, it's a business, you know, it's you know, they don't always do things on behalf of patients. So, I want to make sure that I agree with how my health data is being used and in the very least, well, and if I suppose if I don't agree, I would like to prevent use of my health data in a way that I may think is unethical or improper or in a way that is not beneficial to me or collectively to my community or to the people who live in my, I don't know, so it's really about having – it's about transparency, right, to build some sort of trust in the relationship." "It's important for the participant and the researchers to come to an agreement as to that handshake."
(3) Individual control of personal EHR data sharing	***Wanting control over which entities access data***	Patient	"Someone said [Research Institute], they do quality research, I would say yes, you know, academic centers. Moving along down the line, I don't know how I feel about pharma. I don't know if they want them making money off of me. So, that's like a weird space where I'd either want to see who's asking and either say I'm good with these guys or not good with these guys to have the option to opt in and out." "I don't particularly have concerns with sharing my electronic health record as long as I'm sharing it with an academic institution that I perceive to be reputable. I do have and would have concerns with sharing my electronic health record with for-profit companies and/or organizations I probably would consider sharing it with a for profit company if they are a part of a study consortium or some sort of consortium or group where I would – where academic institutions would be a part

			of. I find that I trust academic institutions, but not private companies." "I just think I'm not concerned. I'm kinda like the opposite. I haven't had a lot of medical issues or anything, several of my family members have had ovarian cancer, so that might be a concern, but I guess just 'cause my thinking is I have so much, everything's on the internet that our lives have just become like wide open. So, I'm just like, you know, at least if there's some benefit, as long as I'm informed."
	Individuals should have control over which data elements to share	Researcher	"Certain socioeconomic status tends to be more fearful of their data being used, then any dataset you're analyzing from the EHR excludes a very important set of individuals who might have less education about the use of their data or less understanding about why it's important or maybe it's a certain religious group. So, I feel like it would really bias the data, potentially, to allow people to say no, but I do respect participant 's point of view that, you know, everyone probably should at the end of the day be asked or educated…it's just I personally really feel like having the data is extremely important."
(4) Importance of using EHR data	***Researcher Benefits when using EHR Data***	Researcher	"I would say that the 40 or so years that I've been doing research, it's almost always included some aspect of looking at a medical record in electronic health records." "It's absolutely essential to my kind of work to be able to access this kind of information and it's easiest to access, obviously, from a health record."
	Ability to leverage Big Data	Researcher	"I would say mostly it's convenience and larger data sets. So, when power and sample size is an issue, if you can, use the EHR data that are already existing and do a retrospective study, you tend to have more power."
(5) Challenges researchers face using EHR data	***Challenges of harmonization***	Researcher	"I also work in observational multi-site studies where we're trying to combine data sets from different places, and even if you're all using Epic, there seems to be a unique Epic at every hospital and that can be kind of a nightmare…you don't know really if you're comparing apples to apples…" "…the data, the way that it's stored in electronic health record, a lot of times you have one type of thing that's stored in many, many different ways, or in different ways for different patients based on the setting."
	Challenges of de-identification and data sharing	Researcher	"I'm interested in how information is stored in unstructured parts of the medical record. Like the physician notes, the stories that are there. I've been wanting to do some multi-institutional studies, but it's really hard to share that kind of data across institutions because it's really hard to deidentify patient's stories."
	Potential for bias or error in the data	Researcher	"Nobody talks about this for the NIS [National Inpatient Sample]. But in the end, there's still a diagnosis code that requires a clinician to put that in there, right. And whether they're missing a comorbidity, whether they coded the diagnosis wrong, all these other factors, EHR has all of the same flaws that go into those." "I've always had my records transferred and as I get a copy of it. I read through it and there's so many inaccuracies when you're explaining to a doctor, and it's basically how they're transcribing the information…between the medical assistant and the doctor, they had put in incorrect information and that's on my record…this was vital information, important information, so again, having your record out there and it's not

correct is even worse.”

1. Motivations for sharing data

Patient participants identified numerous reasons for sharing their EHR data and joining studies in which their EHR data was being access and described being motivated by a variety of factors. Motivation came in the form of gaining information through participating in a study, being able to access medical treatments or resources that otherwise would not be available and experiencing the desire to contribute data that could potentially benefit society. One patient described having a very rare condition and noted the particular importance of sharing their data in order to improve care and treatment for similar patients. Another patient noted the ability to participate in a study that accessed their EHR data potentially benefited not only them, but also "I’m helping other people that may not be able to participate because they don’t have a lot of time.” Accessing medical procedures and resources was particularly important motivation to some of the patients in the study, especially when the procedures were needed often and would otherwise be expensive or very difficult to obtain. Medical research studies were opportunities for some to get care that otherwise might not be available. Some patients expressed more hesitancy about sharing their EHR data when they did not feel that there was a personal benefit, citing the potential for harms such as data breaches or their data being used in a ”nefarious way,” while others noted the importance of allowing researchers to have a representative sample: “And I would like it to be as representative of reality as possible. So, that’s my motivation from a community point of view.”

Several patient participants described their concerns about the lack of trust in healthcare institutions, which translated to a lower propensity to want to share electronic health record data. Noted one patient participant: “It’s kind of like blind faith, blind trust, and historically, that hasn’t been very helpful for us as a people.” Despite this historical lack of trustworthiness in health institutions, another participant expressed balancing this issue with the fact that minoritized individuals are often underrepresented in studies: “But on the other hand, I understand, [the other participant] referred to the apprehension that people of color have and I understand that, but I am a realist and I understand that we as a people tend to be underrepresented in research studies and that we need to be represented because I’m sure there are things that are unique to us as a people… and we wouldn’t know unless someone is taking a look at it and finding out how certain diseases affect us differently.”

2. The importance of data-sharing transparency

The concept of data use and access transparency was brought up by patient participants repeatedly and was revisited throughout the focus groups. For patient participants, transparency was a straightforward concept: one should know how, when, why, and by whom their data was being accessed and used. Additionally, some felt that when they gave broad permission to the health system to use their data, they did not fully understand what that broad consent entailed and wished that health systems would provide more detail into the types of research studies that might use their data. Most patient participants in our focus groups were unaware that health systems had broad “opt-out” policies for electronic health data

and were surprised that most patients' EHR data was regularly shared. Expressed one participant at the idea of their records being regularly accessed for research:

> *I think that something like healthcare is very personal and… since it's so personal, that consent needs to happen. Even if you're anonymous, just somebody using some information about you without telling you feels like a mass violation of your right to privacy.*

Several participants expressed apprehension about sharing certain types of data such as genetic data. Others brought up the issue of lack of data-sharing transparency as particularly problematic given the history of research in marginalized groups. These stakeholders noted that increasing transparency on research done using secondary data could increase trust for all patient groups and particularly marginalized communities.

Data security and access by researcher teams was another important topic discussed by patient participants. They described how the consent process needed additional features to describe how their data are collected, stored, and protected, and how and when their data destroyed. Patient participants also noted that having information about the privacy and security measures that study teams have in place was paramount. One patient participant discussed how they had a historical foundation for having some "paranoia" with how personal data are handled because they consumed news about hospitals selling deidentified patient data and organizations being hacked. Patients also referred to stories about large technology companies collecting and using their data without being explicit about the use of that data. Explained one patient: "Any time you seem to be a part of something, giving a person access, I think is extraordinarily important because there's a lot of things we no longer control because of the virtual world we live in." Patient

participants also wanted to know how and if their data was going to be handed over to another entity/partner, such as another academic entity or company. However, one participant noted that since so much of their data was already on the internet and collected by various entities, they had fewer concerns about sharing their data with researchers. For this patient participant, not being informed when their data was being used “was the biggest issue.” They noted: “I’m just like, at least if there’s some benefit, as long as I’m informed.”

3. Individual control of EHR data sharing

Patients and researchers both discussed the issue of how much/whether patients should have individual control of EHR data sharing. Most patient participants expressed a strong desire to have individual control over their own EHR data and emphasized a preference for having opt-in consenting models for data sharing for research. They wanted to have the ability to monitor data usage and prevent unethical use of data. For example, one patient who had participated in one study selected not to share their genetic information with the researchers, expressing concern over whether insurance companies or other actors could access their data and potentially increase their insurance rates.

Preferences for control over data varied among patients coming from different backgrounds, but increased data transparency and assurances that data was safeguarded were common factors that would make them less concerned about the type of study or organization conducting the study. Patient participants described the “compromise” they struck between their personal motivations to share data

(e.g., being able to participate in studies that might benefit them directly, advancing science) versus the risks they perceived when allowing their personal medical records to be accessed:

> *People will have different ideas of what that is for them and so for me, it was to have access to the data, to the information that would help me make decisions about my health and my healthcare. So, that's what mattered and was important to me. So, if that means that puts at risk the study of unblinding the study, we have to figure out mechanisms where we can have participants get what they seek out of it. In some ways, it's kind of a compromise.*

On the other hand, although most researchers expressed their willingness to honor patients' preferences of having control over when and how patients' data was used, many were concerned about the practicality of doing so and suggested adopting opt-out consenting models. Some researchers believed that the necessity of providing control over secondary data should depend on the risk level of an individual study.

An important consideration was the ability to select with which type of organization the data would be shared. Certain organizations – health systems, academic institutions – engendered more trust, whereas opinion to share data was more mixed for for-profit institutions. For for-profit institutions, patients noted that having "safeguards" or "at least there is some sort of agreement" was critical. Explained one patient participant: "And it may be a reason for me to still join it if I'm desperate or need something that's important they may offer and I may say, well, I have something to gain here as well and I don't mind you sharing that data. As long as it's -- there's transparency and there's openness." Safeguards were also important with academic institutions for some patients, as one patient noted the historical lack of data infrastructure or data security in some health systems.

Overall, both patients and researchers showed openness to balancing individual needs and research needs. Researcher stakeholders hoped to educate patients about the importance of their data for advancing knowledge and implications of not having a representative sample for a study. Patients also expressed willingness to have closer communication with researchers to help researchers have a better understanding of the context of patient data they collected.

An important consideration was articulated by one of the researcher stakeholders. The notion of a representative sample is critical in making sure that the advancements made in medical research, which eventually impact patient outcomes, benefit all of society. Representative samples increase the value, generalizability, and value of research, the researcher noted. Researchers noted that this representativeness could be compromised with a solely opt-in model.

4. How EHR data benefits research

Access and use of EHR data was endorsed as very important for medical research by researcher participants. One notable characteristic of EHR data identified by the researcher stakeholders was the convenience of accessing EHR datasets. The added efficiency of accessing this secondary data can increase sample sizes and expedite many of the research steps, from forming hypotheses to data validation procedures, hence improving study efficiency. Researchers shared various benefits of using EHR data, which included: perceptions that the EHR is more accurate than self-reported data, convenience of data report automation, ease of hypothesis construction and

testing, availability of quick data capture and access, benefits of the continuity/longitudinal aspects, and data granularity.

5. Challenges researchers face using EHR data

Researchers noted several challenges with the use of EHR data, including biased EHR data collection and entry, challenges with data harmonization across health institutions, difficulties associated with deidentification of data (particularly with free-text notes), and the potential for using data that is riddled with miscoding or other types of human error. In one example, a researcher described a situation in which different medical staff, both administrative staff and clinicians, entered data differently even within the same clinical practice. Therefore, research stakeholders noted that it was important for the data analyst and/or researcher to know the practice workflow so that these inconsistencies could be identified and corrected. Another researcher participant explained that it was important to check individual cases to make sure that the data in the large datasets did not have significant errors introduced either during data entry or data retrieval.

Data validation emerged as the most significant concern of researcher -participants. Researchers noted that the EHR structure was designed for billing purposes and was not meant for research purposes. For example, researchers noted that during a busy day, it was possible and likely that clinicians might miscode a diagnosis or forget to add comorbidities. Moreover, without standardized workflows, clinicians and staff might not use EHR data entry fields in a standardized manner. For these

reasons, researchers emphasized the importance of data validation throughout the research process.

Mockup Results

Login

Patient and researcher participants agreed that using social media to login to the platform was problematic, given their stated negative perceptions of social media platforms such as Facebook/Meta, who they noted have contributed to the spread of misinformation. One patient participant commented, “I don’t think it’s a good idea. I’d be very concerned…I also sense that the public opinion has been shifting against these large companies.“ Brand perception played an important role in the trust of large technology companies, for some as one patient noted:

> *I have a different relationship with Google than I have with Facebook. So, my perception of Google is different because I am a paying customer, I pay for services, I pay for storage, I pay for a number of services. Even YouTube, I pay for -- I pay a subscription to YouTube, I guess they used to call it YouTube Red, so there’s no ads. So, I consider my relationship with Google to be different than my relationship with Facebook. For Facebook, I perceive myself as being the product, not as much with Google.*

Patients praised the use of multi-factor authentication, as patients noted that it builds confidence in the system.

Establishing default data sharing preferences

Patients expressed that they liked the having the ability to set default data sharing preferences, but only if they were able to see what institution or organization was conducting the research. “I may be okay with the doctor’s office, but another procedure, another study, I may be okay with the bio company, so just to blanketly

say [sharing with] all doctor's offices, all insurance companies to start to me would be difficult." One patient described that sharing their data would come down to the individual study:

> *But my personal preference would be to look at the study and go back to the individual study and look at it from a standpoint of who's running it, why are we doing it, where's the money coming from, where's it gonna go? In other words, look at it on a case-by-case basis and offer my permission based on the trust and transparency and the outcomes*

After reviewing the platform mockups, research participants expressed concern that creating so many opportunities for patients to take granular control of their data-sharing preferences could have a negative impact on data sets, one researcher participant commented: "[It's a] Paradox of Choice," referring to studies which have demonstrated that more choice can lead to greater dissatisfaction[1]: "Is it going to make them [patients] more satisfied about their data sharing? Or is it just going to add time and make them walk away?"

Setting and reviewing data sharing preferences with individual studies

Patients noted that they appreciated the transparency available when selecting whether to share their data with individual studies. For example, they noted that it would be "enlightening" to know about multi-site collaborations or academic-private partnerships and how their data was being shared between different agencies. They noted that they appreciated being able to opt-in and opt-out of different studies, depending on the condition, institution, and sponsor. Patients noted that they preferred being able to decide on which elements to share with specific institutions rather than always using the default data sharing preferences. Explained one patient when considering with which institutions to share their data (e.g., health system, academic institution): "I don't know, it feels a little bit too big of a bucket."

One patient wanted to know the specifics of the collaboration, specifically how the study was being funded. Another patient appreciated being able to see the name of the principal investigator:

> *I also like how you added who the researcher was or lead researcher was just so I can do background research on her also and see what she -- what else she might have done, what's her credentials. So, I would definitely go in, see what's going on with that research, so that way it keeps track of where my information has gone, who it's been shared with and then if I see them in the news or something horrible, I know to be on watch for it.*

Finally, patients noted that they wanted to know when their data was included in a particular study: "think it would be good if there are any interim results or progress reports that are shared between the major participants that the patients or the study participants would see those as well." Agreed another patient participant about information that they wished was provided to patients whose data had been accessed:

> *You know, a timeline for this, you know, the expected data of conclusion of the study, publishing or results, or I don't know, targeted for Spring of 2022. Some information back… that would be very nice, that would be a very… ethical and polite and nice transparent way to provide some feedback to people.*

Patient participants noted that they preferred fewer options over multiple pages when consenting to new studies: "It makes you more focused on each choice individually rather than when you're giving it all at once," and "It's more of a step-by-step approach, whereas the grid to me, people can just zoom through it."

Overall perceptions of the iAgree system

Patients had overwhelmingly positive perceptions of the iAgree system, although they expressed some concern that patients with lower digital or health literacy, such as older persons, would have a difficult time navigating such a system. Patients appreciated the ability to consent to studies where their data would be

accessed, appreciated the granular control of sharing different data categories, and specifically liked the ability to share or not share their data with different entities. Noted one patient participant:

> *To me, it's a no-brainer. It's better than anything that... there's nothing like this. So, this would be, as far as I know, I never came across anything like this. So, if this were to be like the -- the only sort of way in which I would consent and not consent to all of my studies, that would be a very attractive way to manage my consents.*

Patients praised the additional "layer of trust and transparency" and also described the ability to be able to learn more about the types of research that was being done. Explained one participant:

> *And I think it's not only an opportunity, but it's also an opportunity to do good for yourself and for the community, but also to learn more about a particular disease or condition that might be getting research. You know, I'm gonna learn more about breast cancer because I'm in the [existing study…You learn more about yourself, you learn more about your condition, you learn more about your options, so I mean it's great to have, you know, kind of a one stop shop to do that.*

Finally, several patients expressed the option to get an occasional notification to update their default sharing preferences. Explained on patient participant: "It would be nice occasionally for like an alert to popup to say check your settings, do you still feel this way? To like re-review and check your head and stuff like that."

DISCUSSION

A tension between advancing science and the risks of sharing personal data

In this qualitative study with patient and researcher participants, patient participants expressed concerns about data access, privacy, and transparency, while researcher participants noted the importance of representativeness in the sample, efficiency, and data quality and validation. Both patient and researcher

participants agreed that while giving patients increased choice over the types of data shared was preferable, researchers worried about potential bias being introduced into studies, which could influence the results of biomedical studies. Patients and researchers attempted to deal with this tension between the two competing goals of sharing personal data to advance science and privacy/trust concerns. Patient participants negotiated this tension by agreeing that it was important to contribute to science and/or gain some personal benefit by sharing their data but expressed that if they do so, researchers, health systems, and companies should provide further transparency about how the data are shared and used. Researchers, on the other hand, noted that the benefits of data sharing are outweighed by the benefits of advancing science and expressed serious concerns about allowing patients individual control of their data sharing. Importantly, researchers worried about the issue of representativeness, particularly if patients of different ages, socioeconomic levels, and other characteristics systematically decided to opt in or out of contributing their data for research. Indeed, this may occur if individuals may not have been made aware of the value of contributing data; lack necessary resources, such as time or health literacy, to engage in an activity aimed at setting their data preferences in an opt-in system; or have experience with prior data breaches which reduce their willingness to share data. This speaks to the critical need for researchers to engage in much more intensive trust-building and science communication efforts with patients so that they can express their viewpoints about why EHR data sharing is important.

Considerations for the design of the iAgree platform

We used our findings from patient and research participants regarding their competing goals, (i.e., advancing science and privacy/trust concerns) to inform our design approach. The iAgree platform, currently under development, aims to simplify data-sharing preferences while presenting enough information to ensure transparency, such as study information, the data that will be accessed, the study funder, and how the data will be shared with other institutions. First, the platform will allow patients both to set default data-sharing preferences, but importantly, patients will be able to review new studies and select data-sharing preferences for each study. Second, the platform will include individual study details including the institution where it is based, the funder, the name of the principal investigator, and whether the data are being shared with other institutions, such as in the case of a multi-site study or company. As the participants noted that certain data elements (such as genetic information) are highly sensitive, patients will be able to select which data categories they do or do not want to share and with what type of organization (such as academic or for-profit). Our next steps in this project are to conduct user testing of the iAgree platform with a larger group of patients, incorporate patient feedback into the design of the platform, and then conduct simulated studies to examine how empowering patients to state their data-sharing preferences may affect the data made available for the simulated studies.

Relevance to the existing literature

Our findings about patients' perceptions of data sharing and access echo those of other studies,[27–29] and demonstrates broader implications for the design of systems that empower patients in how their data is used and shared. Prior literature has found that patient groups with a higher prevalence of certain health conditions

(e.g., substance use) may opt-out of sharing these categories of data, introducing bias into a sample. These biases may be particularly concerning when researchers analyze large datasets using machine learning algorithms.[30] These tradeoffs should be considered when building platforms that assist patients with making choices about sharing EHR data. One of our findings was that patients – even highly educated patients – had little awareness of how their EHR data are accessed and used for research. Prior studies have found that individuals who have little knowledge of how secondary data are used for biomedical research have more concerns about data privacy [27], demonstrating a need for health systems and researchers to improve community outreach educational campaigns about the importance of EHR data for medical advances.

Patient participants in our study also worried about having control of their data and transparency, such as data being shared without consent with pharmaceutical companies and other for-profit institutions. These concerns have been mirrored by patient participants around the world.[27,31] A 2009 survey of 4,600 U.S. adults found that while 92% of respondents would be willing to allow academic researchers to access their genetic data stored in a biobank, only 75% would be willing to allow access to industry researchers.[32] A 2018 survey of 771 participants in clinical trials across three academic medical centers in the U.S. found very similar proportions of patients willing to share their data with academic researchers versus industry researchers.[33] Collectively, these findings demonstrate a need to increase transparency across health systems about how patient data are being shared with various entities. This might include additional information in the patient portal and other patient-facing sites about the value and use of secondary data in

research and use of language and icons that convey trust in the safety of how patient data is stored and transferred. For example, online vendors might use links to the Better Business Bureau to engender trust or might include statements about how the users' data is shared (or not shared) with other commercial sites. Such language could be used in the health system informatics setting to increase transparency and trust.

Incorporating a conceptual model of trust when patients' data are being shared or accessed

Patient participants in our study spoke about the need for transparency to increase trust. Anderson and Griffith (2022) recently created a conceptual model to understand and explain the constructs of trust and trustworthiness in the context of healthcare.[34] In this model, the authors describe how *perceived trustworthiness* – a patients' assessment of the probability of experiencing gains or losses from a provider – is an important element of trust. In the realm of research data sharing, patients may assess whether how and to whom their data are shared may lead to gains (e.g., improving biomedical research, advancing science, benefitting society) or losses (e.g., experiencing a lack of data privacy where sensitive data are accessed by malicious actors). Anderson *et al* also describe how direct or vicarious experiences may affect trust. Indeed, we found that patient participants in our study referenced negative experiences from others about data breaches that may impact their trust in healthcare institutions. Moreover, if patients hear only about how their data are shared with for-profit institutions from headline-grabbing news reports, which are almost always negative, this may decrease their trust in healthcare

institutions. Another important construct in the conceptual framework is the *propensity to trust*, which may be influenced by historical events such as the mistreatment or abuse of certain populations in biomedical research. Finally, patients' goals may influence trust in healthcare institutions. If patients feel that they are both contributing to societal benefits and receiving individual benefits (e.g., increased access to screenings, treatments, etc.), these goals may shape a person's propensity to trust and intention to trust. Anderson and Griffith's model can serve as an important model for informaticists to consider when designing a platform for honoring patient data sharing preferences

Strengths and limitations

A strength of this study is the use of a similar focus group guide for both the patient and research participants, which allowed a contrasting of participant perspectives on EHR sharing and researcher requirements. A further strength is the use of the patient perspectives in the design of the iAgree platform. Our paper also has several limitations. Our recruitment strategies may have led to higher representation of highly educated study participants. However, our study included several community board members, who are committed to representing their respective communities in addition to themselves. Future researchers might partner with community-based organizations to recruit viewpoints of individuals with less knowledge about the use of biomedical data. We also had a relatively small sample; however, our findings have been echoed in other studies, demonstrating their transferability to other patient populations. Moreover, as the purpose of the study had a relatively narrow scope, identification of core issues has been demonstrated to require fewer focus groups overall. Finally, our study aimed to reach conceptual

saturation, and given the narrow scope of the study aims, we felt that conceptual saturation was reached by the two focus groups per participant group. [26]

Future Directions

Patient participants noted that it is important for the research institutions to describe how their EHR data are protected. These findings are similar to those of focus groups and surveys held in other countries, where patient participants have expressed preferences for greater transparency over the process of data use, access, and sharing.[19,27,31,32,35,36] One way of operationalizing transparency might be to provide details on how data-sharing works and then accessed by researchers, or who data will be shared with (e.g., with external institutions to create a multi-site study), and why certain data are needed (e.g., describing the research question(s) the study seeks to answer), and how it fits into the research project (e.g., will be combined with other data sources). These recommendations go beyond the simple opt-out model currently used in many health systems. While they may increase complexity for researchers, a model that operationalizes transparency and patient choice may better honor patient preferences.

Researchers in our sample also noted that sharing data across institutions was complicated given varying standards around current data protection, storage, use agreements, and governance. Indeed, IRBs across organizations have vastly different requirements for governing the processes for accessing and sharing data. [37] Increased harmonization of data access and sharing processes can both increase efficiency for researchers and improve transparency for patients about how

their data are used, increasing the need for better and more efficient data standards.

CONCLUSION

In conclusion, our paper outlines important considerations for researchers and health system administrators building patient-centric platforms aimed at improving patient choice over how their data are shared for research. While we found that patients may be willing to share their data for a variety of reasons – improving science/clinical evidence related to their own conditions, recognizing the importance of contributing to science overall – they have important needs regarding transparency and trust. Broadly, health care systems need to recognize that patient data is an important contribution to their scientific endeavors and need to improve communication about the value of such contributions to patients to further engender trust. Moreover, patients need additional information about how and with whom their data is shared and data privacy protections in place to prevent breaches of confidentiality. Such efforts can increase the likelihood that patients will feel empowered to share their EHR data with researchers and others.

ACKNOWLEDGMENTS

We acknowledge our research participants.

COMPETING INTERESTS

Dr. Keller is a scientific advisor and holds equity in RecoverX, a diagnostic clinical decision support company. Dr. Kim was previously at UC Davis and is now a Principal in Consumer Health Informatics and Health Science at MITRE.

FUNDING

This study was funded by NIH/National Human Genome Research Institute (NHGRI) grant number R01HG011066 (PI: Ohno-Machado). Dr. Keller received funding from the National Institute on Aging (K01AG076865).

DATA AVAILABILITY

Data cannot be shared for ethical/privacy reasons. The data underlying this article cannot be shared publicly due to privacy concerns of individuals that participated in the study.

AUTHOR CONTRIBUTION JUSTIFICATION

Researchers involved in the team have trained in qualitative research and/or biomedical informatics. The research tasks related to data collection, data analysis, and discussion of results for authoring the manuscript were divided among the members of the research team as follows in Table 3.

Table 3. Study team and research task division

Authors:	KK	CM	ZX	BM	MK	SM	LS
Informatics Experts	X				X	X	X
Project Coordination	X	X		X			
Methodological supervision	X	X		X			X
Recruitment	X	X	X		X		
Conducting the focus groups	X	X					

Coding of the focus groups			X	X			
Analysis of the groups	X		X	X	X		
Discussion of the results in preparing publication	X		X	X	X		X
Writing and revision of article	X	X	X	X	X	X	X

References

1. SØRENSEN HT, SABROE S, OLSEN J. A Framework for Evaluation of Secondary Data Sources for Epidemiological Research. *International Journal of Epidemiology*. 1996;25(2):435-442. doi:10.1093/ije/25.2.435

2. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med*. 2007;33(2):155-161. doi:10.1016/j.amepre.2007.04.007

3. Block JP, Boehmer TK, Forrest CB, et al. Cardiac Complications After SARS-CoV-2 Infection and mRNA COVID-19 Vaccination — PCORnet, United States, January 2021–January 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71(14):517-523. doi:10.15585/mmwr.mm7114e1

4. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. *Nature*. 2021;594(7862):259-264. doi:10.1038/s41586-021-03553-9

5. Schwab P, Mehrjou A, Parbhoo S, et al. Real-time prediction of COVID-19 related mortality using electronic health records. *Nat Commun*. 2021;12(1):1058. doi:10.1038/s41467-020-20816-7

6. Cancer Patient Perspectives on Sharing of Medical Records and Mobile Device Data for Research Purposes - Elizabeth F Franklin, Helen M Nichols, Linda House, Joanne Buzaglo, Kim Thiboldeaux, 2020. Accessed November 28, 2022. https://journals.sagepub.com/doi/full/10.1177/2374373520923837

7. Weitzman ER, Kaci L, Mandl KD. Sharing Medical Data for Health Research: The Early Personal Health Record Experience. *Journal of Medical Internet Research*. 2010;12(2):e1356. doi:10.2196/jmir.1356

8. Ford E, Oswald M, Hassan L, Bozentko K, Nenadic G, Cassell J. Should free-text data in electronic medical records be shared for research? A citizens' jury study in the UK. *Journal of Medical Ethics*. 2020;46(6):367-377. doi:10.1136/medethics-2019-105472

9. Kalkman S, Delden J van, Banerjee A, Tyl B, Mostert M, Thiel G van. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of Medical Ethics*. 2022;48(1):3-13. doi:10.1136/medethics-2019-105651

10. Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers' views on electronic data sharing for healthcare and research. *Journal of the American Medical Informatics Association*. 2015;22(4):821-830. doi:10.1093/jamia/ocv014

11. Boyd AD, Hosner C, Hunscher DA, Athey BD, Clauw DJ, Green LA. An ‘Honest Broker’ mechanism to maintain privacy for patient care and academic medical research. *International Journal of Medical Informatics*. 2007;76(5):407-411. doi:10.1016/j.ijmedinf.2006.09.004

12. Google’s ‘Project Nightingale’ Gathers Personal Health Data on Millions of Americans - WSJ. Accessed November 28, 2022. https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790

13. Ross C. At Mayo Clinic, sharing patient data with companies fuels AI innovation–and concerns about consent. *STAT News*. 2020;3.

14. Mayo Clinic Assessing Google AI Tool as It Seeks More Insights From Patient Records - WSJ. Accessed November 28, 2022. https://www.wsj.com/articles/mayo-clinic-assessing-google-ai-tool-as-it-seeks-more-insights-from-patient-records-11648159790

15. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. *J Am Med Inform Assoc*. 2013;20(1):7-15. doi:10.1136/amiajnl-2012-001023

16. Whiddett D, Hunter I, McDonald B, Norris T, Waldon J. Consent and widespread access to personal health information for the delivery of care: a large scale telephone survey of consumers’ attitudes using vignettes in New Zealand. *BMJ Open*. 2016;6(8):e011640. doi:10.1136/bmjopen-2016-011640

17. Whose Data Are They Anyway? Can a Patient Perspective Advance the Data-Sharing Debate? | NEJM. Accessed November 28, 2022. https://www.nejm.org/doi/full/10.1056/NEJMp1704485

18. iCONCUR: informed consent for clinical data and bio-sample use for research | Journal of the American Medical Informatics Association | Oxford Academic. Accessed November 28, 2022. https://academic.oup.com/jamia/article/24/2/380/2631502

19. Hollis KF. To Share or Not to Share: Ethical Acquisition and Use of Medical Data. *AMIA Jt Summits Transl Sci Proc*. 2016;2016:420-427.

20. Haddow G, Bruce A, Sathanandam S, Wyatt JC. ‘Nothing is really safe’: a focus group study on the processes of anonymizing and sharing of health data for research purposes. *Journal of Evaluation in Clinical Practice*. 2011;17(6):1140-1146. doi:10.1111/j.1365-2753.2010.01488.x

21. Schuler D, Namioka A. *Participatory Design: Principles and Practices*. CRC Press; 1993.

22. Naderifar M, Goli H, Ghaljaie F. Snowball Sampling: a purposeful method of sampling in qualitative research. Strides Dev Med Educ [Internet]. 2017 [access at: June 17, 2020]; 14 (3): e67670.

23. Parker C, Scott S, Geddes A. Snowball Sampling. *SAGE Research Methods Foundations*. Published online September 9, 2019. Accessed February 24, 2022. http://methods.sagepub.com/foundations/snowball-sampling

24. Rendle KA, Abramson CM, Garrett SB, Halley MC, Dohan D. Beyond exploratory: a tailored framework for designing and assessing qualitative health research. *BMJ open*. 2019;9(8):e030123.

25. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*. 2006;18(1):59-82. doi:10.1177/1525822X05279903

26. Hennink MM, Kaiser BN, Weber MB. What Influences Saturation? Estimating Sample Sizes in Focus Group Research. *Qual Health Res*. 2019;29(10):1483-1496. doi:10.1177/1049732318821692

27. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland - PMC. Accessed November 28, 2022. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6402072/

28. Miller FG. Research on Medical Records without Informed Consent. *Journal of Law, Medicine & Ethics*. 2008;36(3):560-566. doi:10.1111/j.1748-720X.2008.304.x

29. Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue | Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. Accessed November 28, 2022. https://royalsocietypublishing.org/doi/full/10.1098/rsta.2016.0130

30. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. JAMA Intern Med. 2018; 178: 1544–1547. Published online 2018.

31. Robling MR, Hood K, Houston H, Pill R, Fay J, Evans HM. Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study. *Journal of Medical Ethics*. 2004;30(1):104-109.

32. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. *The American Journal of Human Genetics*. 2009;85(5):643-654.

33. Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. *New England journal of medicine*. 2018;378(23):2202-2211.

34. Anderson A, Griffith DM. Measuring the Trustworthiness of Health Care Organizations and Systems. *The Milbank Quarterly*. Published online 2022.

35. Kass NE, Natowicz MR, Hull SC, et al. The use of medical records in research: what do patients want? *Journal of Law, Medicine & Ethics*. 2003;31(3):429-433.

36. Dhopeshwarkar RV, Kern LM, O'Donnell HC, Edwards AM, Kaushal R. Health care consumers' preferences around health information exchange. *The Annals of Family Medicine*. 2012;10(5):428-434.

37. Willison DJ, Emerson C, Szala-Meneok KV, et al. Access to medical records for research purposes: varying perceptions across research ethics boards. *Journal of Medical Ethics*. 2008;34(4):308-314.

1. Tversky A, Kahneman D. Rational choice and the framing of decisions. *Multiple criteria decision making and risk analysis using microcomputers*. Springer; 1989:81-126.