



Stakeholders' views on data sharing in multicenter studies

Journal of **Comparative Effectiveness Research**

Aim: To understand stakeholders' views on data sharing in multicenter comparative effectiveness research studies and the value of privacy-protecting methods. **Materials & methods:** Semistructured interviews with five US stakeholder groups. **Results:** We completed 11 interviews, involving patients (n = 15), researchers (n = 10), Institutional Review Board and regulatory staff (n = 3), multicenter research governance experts (n = 2) and healthcare system leaders (n = 4). Perceptions of the benefits and value of research were the strongest influences toward data sharing; cost and security risks were primary influences against sharing. Privacy-protecting methods that share summary-level data were acknowledged as being appealing, but there were concerns about increased cost and potential loss of research validity. **Conclusion:** Stakeholders were open to data sharing in multicenter studies that offer value and minimize security risks.

First draft submitted: 7 February 2017; Accepted for publication: 25 April 2017; Published online: 14 August 2017

Keywords: comparative effectiveness research • data sharing • distributed research networks • electronic databases • multicenter studies • PCORnet • privacy-protecting methods

Multicenter research networks support a wide range of patient-centered outcomes research, comparative effectiveness and safety research, and public health surveillance activities [1,2]. They allow stakeholders to generate timely and actionable information, study treatment effect heterogeneity in large and diverse populations, and produce generalizable results. In the past, it has often been necessary to share highly granular and potentially identifiable patient-level information across healthcare systems to perform the desired statistical analysis. Even when organizations are willing to collaborate and share information, they must address issues surrounding patient privacy and confidentiality, data security, data control and proprietary interest to meet federal, state and institutional requirements. Meeting these requirements can result in real or perceived loss of efficiency associated with extensive, time-consuming negotiations

and the administrative paperwork burden (e.g., Institutional Review Board [IRB] approvals, data use agreements).

The advent of several new analytic and data-sharing methods offers a more efficient way of tackling these requirements [3–9]. For certain analyses, these methods require only summary-level data, such as propensity scores or intermediate statistics from regression models, to produce results identical or highly comparable to those from pooled patient-level data analysis [3–9]. These newer methods are considered more 'privacy-protecting' as they do not require exchange of potentially identifiable information. They have the potential to improve the efficiency of research through more streamlined security and privacy protection requirements, and could enhance stakeholders' willingness and ability to collaborate in multicenter studies.

Existing research suggests that patients and the public are concerned about the

Kathleen M Mazor^{1,2}, Allison Richards¹, Mia Gallagher³, David E Arterburn⁴, Marsha A Raebel⁵, W Benjamin Nowell⁶, Jeffrey R Curtis⁷, Andrea R Paulino⁵ & Sengwee Toh^{*3}

¹Meyers Primary Care Institute, Worcester, MA 01605, USA

²Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA

³Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, MA 02215, USA

⁴Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA

⁵Institute for Health Research, Kaiser Permanente Colorado, Denver, CO 80231, USA

⁶Global Healthy Living Foundation, CreakyJoints, Upper Nyack, NY 10960, USA

⁷University of Alabama at Birmingham, Birmingham, AL 35294, USA

*Author for correspondence:

Tel.: +1 617 867 4818

Fax: +1 617 867 4276

darren_toh@harvardpilgrim.org

privacy of their electronic health information, but also value research that has the potential to improve care [10–12]. At the same time, most patients and members of the public are not familiar with how their data may be shared, and how their privacy is currently protected [10,13,14]. The new privacy-protecting methods are especially unfamiliar to the public, and relatively unfamiliar to most stakeholders involved in research. These methods may also lack the capability to address some stakeholders' needs and preferences. Regardless of how robust or secure, methods are of limited value if not known to, understood by and proven to be useful to stakeholders. The goal of this qualitative study was to explore and describe various stakeholders' views on sharing of electronic health information in multicenter comparative effectiveness research studies and on privacy-protecting methods in particular.

Materials & methods

Stakeholder groups interviewed

For the purposes of this study, we defined our stakeholders as individuals contributing data to multicenter studies, individuals responsible for stewardship of patient data and the requirements associated with engaging their institutions in data sharing, individuals involved in overseeing and conducting multicenter studies, or individuals involved in using the results of multicenter studies [15]. We identified and invited a purposive sample of stakeholders to participate in the study, including patients, healthcare system leaders, experts in the governance of multicenter studies, researchers and experts who review or oversee compliance, confidentiality and regulatory requirements of research studies.

We recruited patients from two existing groups: a bariatric surgery patient advisory panel previously convened to advise on a research application and patients who participated in the Arthritis Partnership with Comparative Effectiveness Research (known as ArthritisPower™; Upper Nyack, NY, USA), a Patient-Powered Research Network within the National Patient-Centered Clinical Research Network (PCORnet) [16]. We chose these two existing groups because this study was conducted in the context of a larger project that involves patients who have undergone or are considering bariatric procedures and patients with autoimmune diseases. We identified healthcare systems leaders, experts in the governance of multicenter studies and experts in research compliance, confidentiality and regulatory requirements from three delivery systems: Group Health Research Institute (now Kaiser Permanente Washington Health Research Institute), Kaiser Permanente Colorado and Kaiser Permanente Northern California. We

enrolled researchers from the attendees of the Patient-Centered Outcomes Research Institute (PCORI) Annual Meeting in 2015. In the following text, we refer to patient stakeholders as 'patients' and to all other participants as 'organizational stakeholders'.

Data-sharing & analytic methods of interest

We were interested in stakeholders' views on various data-sharing and analytic methods used in multicenter studies, including pooled patient-level data analysis, patient-level or summary-level data analysis that leverage confounder summary scores (e.g., propensity scores), risk set-based analysis, and meta-analysis of site-specific effect estimates [3–6]. Each method requires sharing specific information across sites and offers various degrees of analytic flexibility. See [Supplementary Appendix 1](#) for examples of information typically shared by a participating site in a multicenter study using these analytic methods. Detailed description of the strengths and limitations of each method is available in other published articles [3–6].

Interview process & content

Prior to each interview, we sent stakeholders a fact sheet that described the purpose of the study, potential risks of the interview (which were minimal) and their expected level of participation (see [Supplementary Appendix 2](#) for a version shared with the healthcare system leaders). We conducted the interviews in person or via telephone, as a group or individually, based on the preference and availability of the stakeholders. One author (S Toh) conducted all interviews. At least one other member of the research team was also present for all interviews. Each interview began with a review and clarification of the fact sheet. The interviewer then described various data-sharing and analytic methods in multicenter studies using educational materials (see [Supplementary Appendix 3](#) for a version used for the interviews with the healthcare system leaders) tailored to the background of the interviewees. The presentations and interviews focused on data typically captured in electronic health records and administrative claims databases, rather than biospecimens or genetic data. The interviewer then asked the interviewees a series of questions based on the domains developed by the study team ([Box 1](#)). The specific interview questions varied depending on the interviewee's role and familiarity with data sharing, and evolved over the course of the interview (see [Supplementary Appendix 4](#) for an interview guide used for the healthcare system leaders). We recorded all interviews with permission from the interviewees and professionally transcribed them for analysis. We did not collect any identifiable information about the interviewees during the interview.

Box 1. Interview domains.

- Familiarity and experience with multicenter research and data sharing
- Attitudes toward multicenter research and data sharing
- Perspectives on privacy and data security
- Perspectives on sharing aggregate-level versus individual-level data
- Recommendations for improving processes around data sharing
- Reactions to privacy-protecting analytic and data-sharing approaches

Note: Specific interview questions varied across interviews depending on stakeholders' roles and responses.

Analysis

We used an integrated approach to the qualitative data analysis as described by Bradley *et al.* [17]. The interview domains provided an initial organizing framework, consistent with a directed content analysis approach [18]. However, we were attentive to unanticipated content as we reviewed the transcripts and applied the evolving coding scheme, integrating new codes and concepts as they emerged inductively, consistent with conventional content analysis [18]. One investigator (KM Mazor) created an initial coding framework after observing four interviews. Two other investigators (S Toh, DE Arterburn), who had participated in the interviews provided feedback on the framework, and suggested additional themes or sub-themes. The first investigator (KM Mazor) elaborated the framework through ongoing review of the transcripts as additional interviews were completed. Four team members (S Toh, DE Arterburn, MA Raebel and A Richards) each reviewed at least one transcript, with the coding framework at hand. These second readers checked for the completeness of the framework, and suggested new codes or modifications based on their

review. The full qualitative team (KM Mazor, S Toh, DE Arterburn, MA Raebel and A Richards) reviewed and reached consensus that the final coding framework captured all relevant themes and subthemes expressed in the interviews (Supplementary Appendix 5). One team member (A Richards) coded all transcripts; a second team member (KM Mazor) reviewed the coded transcripts to confirm accuracy, and to resolve any questions that emerged during the final coding. The team entered the transcripts and codes into the Statistical Package for the Social Sciences (SPSS version 22) in order to facilitate data management, manipulation and reporting.

Results

We interviewed 34 stakeholders between June 2015 and February 2016 (Table 1). The average interview duration was approximately 61 min (range: 36–109 min). The analysis identified three major themes which emerged inductively from the qualitative analysis: perceived benefits and value of research, cost and perceived risks. Figure 1 provides a conceptual model of how these major themes relate to stakeholders' willingness

Table 1. Stakeholder groups interviewed.

Stakeholder group	n	Interview type	Interview mode
Patients			
Arthritis patient panel	10	Group	In person
Bariatric patient panel 1	4	Group	In person
Bariatric patient panel 2	1	Individual	Telephone
Healthcare systems leaders			
Vice president for governmental external relations	1	Individual	Telephone
Executive medical director	1	Individual	Telephone
Medical director for quality	1	Individual	In person
Consultant for research compliance and ethics	1	Individual	In person
Multicenter research governance experts			
– Multicenter research governance expert 1	1	Individual	In person
– Multicenter research governance expert 2	1	Individual	In person
Researchers	10	Group	In person
Compliance, confidentiality and regulatory experts	3	Group	Telephone

Each row of this table represents a separate interview session, either group or individual.

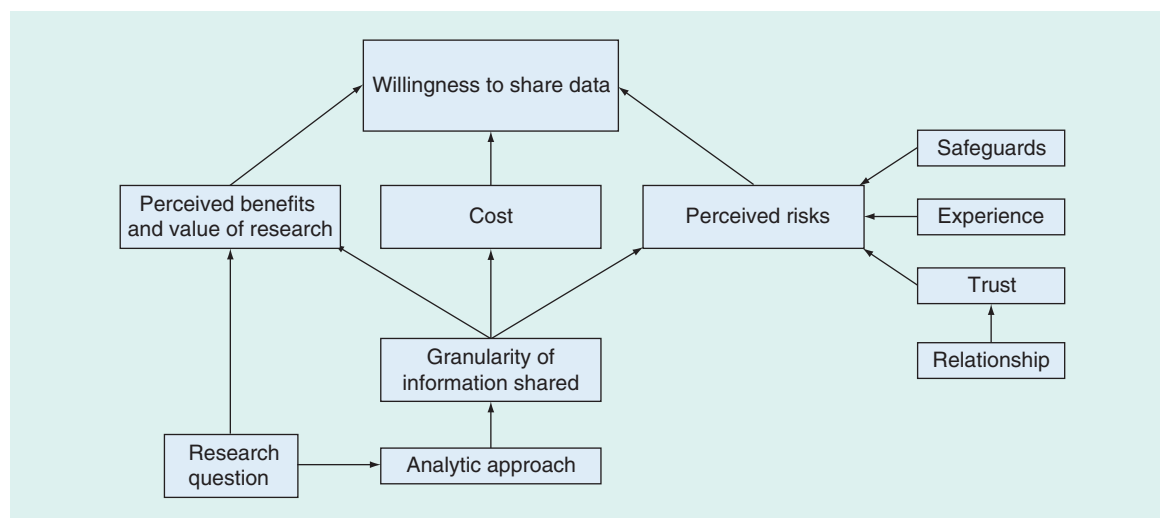


Figure 1. Major themes identified from the stakeholder interviews.

to share data in multicenter studies, which was a central focus on this study. Each of these major themes (perceived benefits and value, cost and perceived risks) was influenced by the granularity of the information to be shared, as well as by other factors (e.g., perceptions of risk were also influenced by past experiences). We noted varying levels of stakeholder familiarity with privacy-protecting analytic and data-sharing methods, as well as differences in views on the usefulness of these methods; these findings are presented last.

Factors that influenced willingness to share data in multicenter studies

Perceived benefits & value of the research

Stakeholders' perceptions of the purpose, benefit and value of the research were the strongest influences toward data sharing. Both patients and organizational stakeholders referred to the need for research that would answer questions that they perceived to be important, relevant and likely to improve care or outcomes for patients. They indicated they would be more likely to support data sharing in pursuit of those goals. As one patient said, *"If it's improving the general knowledge in service of people like me, that's a good thing."* In contrast, patients were unwilling to share data if they perceived the request was motivated by financial gain or profit. Patients considered it both possible and highly objectionable that an entity might profit by selling their data.

Organizational stakeholders valued data sharing as a means of improving patient care, and of advancing understanding of treatment risks and side effects. They referred to the fact that multicenter data sharing necessarily results in larger datasets and thus enhances the ability to study rare diseases and rare outcomes (i.e., increased statistical power). Organizational

stakeholders also referred to improving generalizability of study findings. As one organizational stakeholder stated, *"It seems like more data is better... more generalizable, more scientific."* Another noted that data sharing allows healthcare systems to *"provide richer data to the world."*

Patients' comments indicated a desire for their data to be helpful, and to lead to valid and actionable findings with the potential to improve care for others. Patients referred to the need for *"good science,"* and recognized that not all studies achieve this. As one patient said, *"I suppose it goes back to the risk/reward, ... we're getting good science out of these studies. And if we're not, I think that's a bigger problem than the privacy issue."*

Costs

Cost was a factor identified as influencing organizational stakeholders away from data sharing. Organizational stakeholders' comments implied that financial consequences and costs of decisions were important in their decision-making, including their decision making related to their organization's participation in research and data sharing. These stakeholders were cognizant of the costs associated with data sharing, and considered these when making decisions about data sharing. They noted that data sharing requires resources, most notably programmer or analyst time and expertise, which are often limited. As one organizational stakeholder noted *"everything's an opportunity cost."* None of the stakeholders commented on how the costs of data sharing using privacy-protecting methods might be covered, though one organizational stakeholder commented that building on existing research networks, where the foundational work, such as the creation of shared data models, *"lowers the burden"* of data sharing. This

stakeholder went on to say that in the short- to near-term the additional costs associated with developing and implementing privacy-protecting methods would be “an investment in methods development,” but also noted that if a project did not fully cover the costs of participation, then “we can’t do it.”

While patients did not refer to the cost of data sharing *per se*, some mentioned compensation, believing that they should be compensated for the use of their data, with compensation being broadly defined to include financial compensation, expressions of appreciation and recognition and sharing of results. Patients also expressed concern that their data might be used for commercial purposes.

Perceived risks

Perceived risk was also identified as influencing stakeholders away from data sharing. The most prominent concern identified by organizational and patient stakeholders was loss of control of the data, with the associated risk of unauthorized use or disclosure. Interviewees expressed concerns that sensitive health information, including information about patients’ diagnoses and treatments, might be divulged to those who should not have access, and ultimately result in harm to patients. While patients were concerned about loss of confidentiality and unauthorized release of their information, few were explicit in identifying the downstream consequences of disclosure they were most worried about. One patient was somewhat specific, expressing a concern about the possible impact on employment, saying, “Twenty years ago, and you have HIV, you’re fired, ... Today, not as much, but, like, I think that’s a factor.” Another patient referred to “my insurance company or somebody’s going to use that against me,” while another said simply “it’s a stigma...it’s nobody’s business.”

Organizational stakeholders also alluded to the risk of data sharing resulting in damage to an organization’s reputation, or loss of competitive advantage. One organizational stakeholder referred to using the litmus test “if this were released and it ended up on the front page of the (newspaper name), what would that do? To our patients, to our reputation, et cetera.” Organizational stakeholders appeared concerned about the possibility that disclosed data could suggest that a given provider, clinic, or organization might be portrayed as a poor performer, referring to “issues around quality outcomes, competition,” in this context. One organizational stakeholder referred to concerns about “a dataset and that gets in the wrong hands and you suddenly discover that, you know, this one clinic is horrible.” Another organizational stakeholder referred to the potentially competing interests of researchers within

an organization, noting “you also have the researcher who might be trying to do, you know, kind of establish themselves in a particular topic area, and may feel some level of protectiveness over the data.” Overall, organizational stakeholders were acutely aware that harm could result from a data breach or loss of patient confidentiality secondary to data sharing, though none reported direct experience with such events. One organizational stakeholder referred to the widely publicized data breach at the Veteran’s Health Administration, saying “a data breach in VA research, as you may remember, completely shut down the VA research enterprise for a couple of years...It was horrible.”

It is noteworthy, however, that some patients and organizational stakeholders were not concerned about data sharing, and made explicit their belief that there was little risk of harm. One patient asked directly whether unauthorized disclosure was a problem with research data, saying “I guess I would want to know how rampant a problem it is,” later noting “The risk is much smaller than say, just me buying something with my credit card.”

Several factors influenced organizational and patient stakeholders’ perception of risk, as described in Figure 1.

Safeguards: Organizational stakeholders identified a number of safeguards and strategies used to minimize risk of data breaches and to maximize data security. Some organizational stakeholders indicated that such safeguards are currently in place; others indicated that they would require that such safeguards be in place prior to data sharing. Approaches referenced included technological approaches (e.g., use of encryption, firewalls), policies and contractual practices (e.g., data use agreements) and oversight for ensuring compliance with agreed upon practices.

Some organizational stakeholders noted their organizations required that an internal researcher be involved in all studies involving data sharing to reduce the risk of inappropriate use. Involvement of an internal researcher was also sometimes necessary to ensure that the nuances of the data were taken into account in analyses and reporting. Some organizational stakeholders were apparently acutely aware of the complexities of operational data, and the potential for naive users to make incorrect assumptions about the data which could, in turn, lead to erroneous and invalid results.

Organizational stakeholders also referred to restricting data access (again referring to the current implementation of such practices), and the need to obtain assurances about limits on access whenever data were shared. Patients also brought up the importance of restricting data access, oversight of such

restrictions and voiced specific questions about data security, for instance, wanting details on how the data would be transferred. Some patients expressed uncertainty about current practices; as one patient said, “*I don’t know who has access to my information.*”

Prior experience: Stakeholders’ prior experience with data sharing influenced their views on the potential risks. Several organizational stakeholders referred to sharing data for research without problems or concerns. Successful experiences appeared to reduce the perception of risk, at least for data sharing in similar contexts. No interviewees reported direct personal or organizational experience of negative consequences of data sharing, though one organizational stakeholder referred to a “*near miss*,” – for example, an event where identifiable data were almost shared, but were detected and prevented. Organizational stakeholders also noted that if a data breach were to occur, it would be likely to have a major impact on the organization’s willingness to share data in the future.

Two patients mentioned personal experience working with data (one in a work setting, and one in an educational setting), and indicated that this experience had increased their comfort with data sharing. One patient noted “*we would get datasets like this, and I mean, there was absolutely no way you could tell, you know, even what region the person was from ... I can say as someone who has seen how it’s presented, you know, I feel safe.*” Another patient also referred to being more comfortable when “*everything’s just a number.*” An organizational stakeholder also raised this issue, noting, “*I don’t think the patients have a clear sense of when we go into a data warehouse and extract data, what that’s like, that they’re a string, with a random ID.*”

Trust and relationships: both patients and organizational stakeholders referred to the need to trust the researchers or organization requesting data, both with respect to how the data would be used and in the users’ ability to ensure the data security. The degree of trust appeared to be influenced by familiarity, and whether there was an existing relationship with the organization or individuals involved. As one patient stated “*...with (organizations), you know, there’s years of trust there, and so forth. So that comes down to the people, knowing the people that are behind the scenes, working with that information.*” Organizational stakeholders were less willing to share with unfamiliar requestors. As one organizational stakeholder stated: “*if we were approached by some other, new group we’ve never heard of, that our delivery systems or health insurers or whatever that we don’t know and they said, ‘Trust us, we would (laughs) have some trouble with that’.*”

Type & granularity of data shared

The type of data to be shared and the degree of aggregation also influenced stakeholders’ views on the value, costs and risks of data sharing and their willingness to share. In all multicenter studies, the research question drives the analytic approach which, in turn, dictates the type and granularity of information to be shared. Organizational stakeholders, especially those with oversight or regulatory responsibility, focused on whether the requested data elements were relevant to the research questions, and were unwilling to approve sharing of data elements that were not relevant. Organizational stakeholders were also reluctant to approve sharing of sensitive information such as HIV status, mental health status or alcohol use, and referred to requests for medical record numbers as ‘red flags’. Patients were generally unwilling to share names, birth dates, social security numbers and financial information; it was implicit in organizational stakeholders’ comments that these would typically not be shared. Some patients wanted to specify as to which data elements would be shared, and the conditions under which these could be shared; others indicated they would want to be informed when their data were shared. In general, some research topics and data elements were considered more sensitive than others, and would receive greater scrutiny.

Both patients and organizational stakeholders made statements and asked questions about the relative advantages and disadvantages of summary- versus individual-level data. The risk reduction obtained by sharing summary-level data rather than individual-level data was attractive to some stakeholders. However, a repeated theme across several interviews was whether aggregating data resulted in a loss of information that would reduce the value or validity of the research. As one organizational stakeholder asked, “*How much more generalizable knowledge can be obtained through – from the scientific perspective – in analyzing the patient-level data?*” A patient asked a similar question, with a slightly different focus, saying, “*Does this type of method, where you have less granular information, lead to a less actionable result?*” and later “*To me, actionability of research outweighs my privacy anxiety, significantly.*” Some questioned whether summary-level data would allow as complete and nuanced exploration of the research question as individual-level data.

Organizational stakeholders expressed concerns about the costs involved, noting that creating summary-level data files may require more technical and programming expertise and additional resources to create. Devoting resources to aggregating data files was seen as having opportunity costs as well, as programmers and

analysts were viewed by some as a relatively scarce staff resource within their organization.

Some organizational stakeholders opined that summary-level-based approaches would be appropriate if the goal of the study was to answer a single, well-defined research question, but that these approaches would be less useful if the goal was to gain a nuanced understanding of a phenomenon. As one organizational stakeholder put it, *"...if you get something that's surprising, you'd want to know why and that means you have to unpack it ... you probably can't do that because some of those problems are in the way the propensity score was constructed."*

Some organizational stakeholders indicated that summary-level data approaches would not influence their willingness to share data. As one put it, *"If I'm not comfortable giving you the individual stuff, I'm not going to be comfortable giving you the propensity score."* This leader went on to say, *"It seems a trade-off and the question is, what do I gain for that trade-off and do I think that that was already at risk? If I saw the data as at risk, I don't know that I'd be wanting to participate."* However, individual-level data were not preferred by all organizational stakeholders: *"Yes, you have more ability to do analysis on patient-level data, but it comes at a cost, right? Of security and privacy."* Another saw an advantage in planning and decision-making needed to assemble data for aggregate approaches, suggesting that specifying the variables to be included and the analysis prior to data sharing would result in a more 'honest and transparent' approach.

Familiarities with & views on the privacy-protecting analytic & data-sharing methods

Patient stakeholders were unfamiliar with privacy-protecting analytic and data-sharing methods; organizational stakeholders expressed limited understanding. Most interviewees had never heard of one or more of these newer methods, but some researchers had used some of the methods (e.g., propensity score-based methods) in their studies.

Stakeholders' reactions to the privacy-protecting analytic and data-sharing methods, as we described them during the interviews, were mixed. Some interviewees did not perceive a need for these methods, and others did not view these methods as providing significantly greater privacy protection. Overall, organizational stakeholders considered current safeguards sufficient. However, as one interviewee noted, if someone *"made a big mistake"* those views might change, resulting in a greater need for privacy-protecting methods. Some were uncertain of the relative advantages of the newer privacy-protecting methods (i.e., the approaches which were the ultimate focus of this investigation).

Other organizational stakeholders felt that use of privacy-protecting methods were clearly preferable to sharing patient-level data. As one organizational stakeholder said, *"I believe that the cultural resistances to patient-level data sharing are so deeply embedded in organizations that the best approach is privacy-protecting methods ... I think privacy-protecting methods allow us to patiently but persistently figure out better approaches to multi-site data."*

Some interviewees suggested that privacy-protecting methods would be more acceptable to specific stakeholder groups. For instance, researchers predicted that IRBs would find these methods more acceptable. This was confirmed by a comment from an organizational stakeholder with IRB experience who said, *"From an IRB perspective it's great. It's definitely better, there's no question."* Another organizational stakeholder predicted *"Our patients are going to be viscerally more comfortable with it."*

Patients' comments were more equivocal. One patient, apparently unconvinced of the need for or value of privacy-protecting methods, commented *"It's a lot of trouble simply for me to feel a little more secure. And for my vote, it's insignificantly more secure."* Another patient appeared not to perceive a need for privacy-protecting methods personally, but thought other patients might: *"You know, there's information I'm willing to give and information I'm not willing, you know, to – as long as you let me know. I don't care. But I can see that there's going to be a lot of people who aren't so open, and I think this method would probably make them much more comfortable."*

Discussion of ways to increase the acceptance of privacy-protecting methods identified recommendations for providing additional evidence of the value of the approach. Some organizational stakeholders wanted to see examples of the application of these methods, and demonstrations of the equivalence of results obtained when using these methods compared with standard approaches. As one organizational stakeholder put it *"...since these methods are opaque by design, I think the only way to overcome that is a series of studies that basically have access to both the full data-set and the privacy protected methods and to show that across a wide array of questions, data structures, analytic techniques, that the results are identical."* Another stated *"...we have to have confidence as a reader of the literature, that they (privacy-protecting methods) actually are correct."* A patient made a similar recommendation, saying, *"If you can say that you can get the same quality results from the summary, then I'd go with that. But my – I question whether or not that's true."*

The possibility that proposals using privacy-protecting methods might make it through the institutional review

process relatively quickly was noted as an advantage by some organizational stakeholders, and examples of instances where proposals using these methods resulted in more timely IRB approval would help to convince stakeholders of their value. Similarly, recognition of the resources needed to produce the summary-level datasets used in privacy-protecting analyses led to recommendations to find ways to make these methods cheaper, faster and more efficient.

One organizational leader referred to the ‘downside’ of privacy-protecting methods as “...the fact that all reputable researchers like to get the data under their fingernails. You like to get dirty with the data. And when you can’t do that, then, then you get apprehensive and you should. That’s an instinct that was trained into all of us in graduate school. And so not being able to see the raw data makes us viscerally uncomfortable.”

Patients’ questions and comments also highlighted the need to inform and educate patients about current practices and protections. As one patient stated “I think part of it comes down to, it’s just patients getting enough education about the process, and the outcomes that we’re looking for, to feel comfortable sharing that information, and to realize that you know what? Guess what? Maybe we, as a generation, need to go out on a limb a little bit here, but it’s that proper education of how this is going to be used, and proper thanks.”

Discussion

Multicenter research studies that leverage various existing data resources have the potential to generate timely, actionable and generalizable results. Our findings from these interviews elucidate the factors that influence stakeholders’ views on data sharing in multicenter research. Consistent with prior studies conducted in the USA, the UK, Canada and elsewhere [10–12,19,20], our findings suggest that while stakeholders generally recognize the value of research and are motivated to contribute to better patient care and outcomes, many have reservations about data sharing for research. New analytic methods may reduce concerns about privacy and anonymity, but our findings suggest that other factors influence stakeholders’ views and must be considered.

Our findings extend what is known by providing insights into organizational stakeholders’ perspectives that were generally consistent with patients’ views, perhaps because most have responsibility for protecting patient confidentiality and data security. While both patients and organizational stakeholders voiced questions and concerns, most were open to data sharing as long as the research was addressing questions that were important to patient care.

A particular focus of these interviews was to explore stakeholders’ reactions to the newer privacy-protecting

methods analytic and data-sharing methods. Our experience in these interviews highlights the fact that these methods were not well-understood by the stakeholders. The methods were also difficult to explain, especially to less technical audiences. As we did explain them, reactions were mixed. While stakeholders acknowledged that privacy-protecting methods enhanced privacy and reduced the risk of reidentification of patients, these benefits were weighed against the cost of preparing the datasets, and the perception that such approaches might limit the value of the research by reducing generalizability, validity and the ability to explore nuances in the data. There are several ways to make multicenter studies more efficient, for example, by standardizing the databases in advance so that the analytic code can be developed by the study team and executed with minimal modifications at other participating sites [1,21–23]. Recent simulation and empirical studies have also shown that these methods produce results statistically equivalent to the results from pooled patient-level data analysis for certain study settings [5,6,8,9]. The feedback from the stakeholders in this study highlights the need for better education and more research in these methods.

Trust emerged as an important influence in our interviews, for both patients and organizational stakeholders. The importance of trust to patients has been reported previously [11,20,24–26]. Our finding that organizational stakeholders also consider trust and relationships when deciding about data sharing in the context of multicenter studies is not surprising. The stakeholder interviews provided insights into ways to build and maintain trust, including familiarity with the data requestor and proper safeguarding of the data.

The patients participating in this study were already engaged with the research process in some way and thus potentially more open to data sharing than other patients, but most did not convey a solid understanding of existing practices and safeguards around the use of their personal health information. All of the patient stakeholders in this study were familiar with research wherein individuals choose whether or not to participate, provide informed consent and know generally what information they are providing to the researchers. However, many patients were not familiar with studies where electronic health data might be deidentified and shared without documentation of patients’ permission. Patients with previous exposure to data analyses or reporting in the context of work or school appeared much less concerned about the risks associated with data sharing. Typical patients without exposure to research or data analysis processes likely have an even more limited understanding of how data may be used and

what data sharing entails. This may influence their willingness to share their information for research. Our findings are consistent with prior studies that have documented patients are poorly informed as to current practices, safeguards and implications of data sharing [10–14,20,27,28].

Further, our findings highlight the need to improve patients' awareness and understanding of the risks and benefits of research in general. Future studies are needed to identify the best methods of educating the public about existing safeguards and data sharing practices, as well as the need for, and potential value of, comparative effectiveness research using real-world data. For organizational stakeholders, who are likely to weigh the potential value of proposed research against both the costs and the potential risks, additional research to determine the actual costs of data sharing using different methods, as well as further evidence regarding the comparability of the findings obtained, may help these stakeholders as they consider the trade-offs associated with each method.

A major strength of the study is the inclusion of a group of stakeholders with diverse backgrounds, who may be involved in multicenter studies. Their participation in these interviews offered a more comprehensive view on the complex issues around data sharing in multicenter studies. However, our findings should be interpreted in the context of the following limitations. This was a qualitative study with a relatively selected group of stakeholders; while participants brought different perspectives, we do not know the extent to which their views are representative. We conducted both group and individual interviews, which may have influenced our findings. The study was not designed to provide generalizable findings. Patients were selected because of their engagement with research; their views may not be representative of naive patients. On the other hand, their feedback may be relevant to PCORnet, which includes a network of patients who are actively engaged in research activities. Finally, our focus was on sharing information currently available in electronic healthcare databases, such as diagnoses, pharmaceutical or surgical treatments, healthcare encounters and laboratory test results. We did not explore issues related to sharing genetic or genomic data, and so cannot comment on whether stakeholders' views on that topic would be similar or different.

Conclusion

In this study, we found that stakeholders are open to data sharing in multicenter studies if the research offers benefits and value to patient care, minimizes data security risks, and can be done at reasonable cost. The gains in privacy protection associated with the use of privacy-protecting analytic and data-sharing methods in multicenter studies were attractive to some stakeholders, but others were concerned about increased cost and potential loss of research validity when using these methods. Most stakeholders were not familiar with these newer methods and their validity, highlighting the need for better education and more research into these methods.

Acknowledgements

The authors thank the following individuals for their contribution to this manuscript: J Fraser, who helped organize a stakeholder meeting; S Gruber, who provided comments on an earlier draft of the manuscript; and L Lagreid, who served as a patient advisor to the project.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/cer-2017-0009

Financial & competing interests disclosure

This study was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1403-11305). All statements in this article, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI or PCORI's Board of Governors or Methodology Committee. Dr. Toh is also supported by the National Institute of Biomedical Imaging and Bioengineering (U01EB023683). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Summary points

- Data sharing is a fundamental step in multicenter studies, allowing stakeholders to generate timely and actionable information, study treatment effect heterogeneity and produce generalizable results. However, data sharing entails costs and risks.
- Newly developed privacy-protecting analytic and data-sharing methods offer an approach to sharing data and conducting multicenter research that eliminates the need to share potentially identifiable patient-level information.
- We conducted semistructured group and individual interviews with diverse stakeholders to gather a variety of perspectives on data sharing. Interviews were audio-recorded and professionally transcribed. Using content coding followed by thematic coding, we sought to identify factors affecting stakeholders' willingness to share data.
- We completed a total of 11 stakeholder interviews, involving patients (n = 15), researchers (n = 10), Institutional Review Board and regulatory staff (n = 3), multicenter research governance experts (n = 2) and healthcare system leaders (n = 4).
- Stakeholders' perceptions of the benefits and value of the research was the strongest influence toward data sharing; perceived value was related to the relevance of the scientific question and the methodological rigor.
- Influences against data sharing were primarily cost and data security risks; the latter could be mitigated by various safeguards (e.g., encryption, data use agreements and oversight), successful data sharing experience, established relationships and trust.
- The risk reduction obtained by sharing aggregate-level data rather than individual-level data was acknowledged as being potentially more acceptable to some stakeholders, but some stakeholders expressed concerns about the increased cost and potential loss of research validity.
- Our findings highlight the need for better education and more methodological research in privacy-protecting analytic and data-sharing methods.

References

Papers of special note have been highlighted as:

• of interest; •• of considerable interest

- Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff. (Millwood)* 33(7), 1178–1186 (2014).
- An overview of recent efforts to use multiple databases for rapid evidence generation in the USA.**
- Toh S, Platt R, Steiner JF, Brown JS. Comparative-effectiveness research in distributed health data networks. *Clin. Pharmacol. Ther.* 90(6), 883–887 (2011).
- Rassen JA, Avorn J, Schneeweiss S. Multivariate-adjusted pharmacoepidemiologic analyses of confidential information pooled from multiple health care utilization databases. *Pharmacoepidemiol. Drug Saf.* 19(8), 848–857 (2010).
- Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med. Care* 51(8 Suppl. 3), S4–S10 (2013).
- A useful overview of various data-sharing and analytical methods available in multicenter studies, focusing on patient privacy protection and analytic flexibility.**
- Toh S, Reichman ME, Houstoun M *et al.* Multivariable confounding adjustment in distributed data networks without sharing of patient-level data. *Pharmacoepidemiol. Drug Saf.* 22(11), 1171–1177 (2013).
- Toh S, Shetterly S, Powers JD, Arterburn D. Privacy-preserving analytic methods for multisite comparative effectiveness and patient-centered outcomes research. *Med. Care* 52(7), 664–668 (2014).
- A real-world data analysis that compared the statistical performance of several data-sharing and analytic methods in a multidatabase study. The study showed that some privacy-protecting methods produce statistically equivalent or highly comparable results to the results from pooled patient-level data analysis (benchmark).**
- Fireman B, Lee J, Lewis N, Bembom O, Van Der Laan M, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. *Am. J. Epidemiol.* 170(5), 650–656 (2009).
- Karr AF, Lin X, Sanil AP, Reiter JP. Secure regression on distributed databases. *J. Comput. Graph. Stat.* 14(2), 263–279 (2005).
- Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary LOGistic REGression (GLORE): building shared models without sharing data. *J. Am. Med. Inform. Assoc.* 19(5), 758–764 (2012).
- Hill EM, Turner EL, Martin RM, Donovan JL. 'Let's get the best quality research we can': public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. *BMC Med. Res. Methodol.* 13, 72 (2013).
- A useful systematic review that summarizes both qualitative and quantitative studies of the public's views on the use of existing health data in research. Findings from eight countries are represented. Results are complemented by focus group findings (primary data collection).**
- Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. *Soc. Sci. Med.* 64(1), 223–235 (2007).
- Patients from Veterans Affairs facilities in the USA participated in small group deliberations (with access to experts) about sharing their medical records for research. Both qualitative and quantitative methods were used to assess participants' views.**
- Willison DJ, Schwartz L, Abelson J *et al.* Alternatives to project-specific consent for access to personal information

- for health research: what is the opinion of the Canadian public? *J. Am. Med. Inform. Assoc.* 14(6), 706–712 (2007).
- **A national telephone survey of the Canadian public's attitudes toward allowing access to medical information for research. Quantitative data on attitudes are presented conveying the complexity of the public's views on this topic.**
- 13 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. *J. Med. Ethics* 30(1), 104–109 (2004).
 - 14 Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. *Int. J. Med. Inform.* 75(7), 530–541 (2006).
 - 15 Concannon TW, Meissner P, Grunbaum JA *et al.* A new taxonomy for stakeholder engagement in patient-centered outcomes research. *J. Gen. Intern. Med.* 27(8), 985–991 (2012).
 - **A new framework and taxonomy for engaging stakeholders in patient-centered outcomes research.**
 - 16 Daugherty SE, Wahba S, Fleurence R. Patient-powered research networks: building capacity for conducting patient-centered clinical outcomes research. *J. Am. Med. Inform. Assoc.* 21(4), 583–586 (2014).
 - 17 Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv. Res.* 42(4), 1758–1772 (2007).
 - 18 Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual. Health Res.* 15(9), 1277–1288 (2005).
 - 19 Luchenski SA, Reed JE, Marston C, Papoutsis C, Majeed A, Bell D. Patient and public views on electronic health records and their uses in the United Kingdom: cross-sectional survey. *J. Med. Internet Res.* 15(8), e160 (2013).
 - 20 Stevenson F, Lloyd N, Harrington L, Wallace P. Use of electronic patient records for research: views of patients and staff in general practice. *Fam. Pract.* 30(2), 227–232 (2013).
 - 21 Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med. Care* 48(6 Suppl.), S45–S51 (2010).
 - 22 Toh S, Platt R. Is size the next big thing in epidemiology? *Epidemiology* 24(3), 349–351 (2013).
 - 23 Ross TR, Ng D, Brown JS *et al.* The HMO research network virtual data warehouse: a public data model to support collaboration. *EGEMS (Wash. DC)* 2(1), 1049 (2014).
 - 24 Kass NE, Natowicz MR, Hull SC *et al.* The use of medical records in research: what do patients want? *J. Law Med. Ethics* 31(3), 429–433 (2003).
 - 25 Weitzman ER, Kaci L, Mandl KD. Sharing medical data for health research: the early personal health record experience. *J. Med. Internet Res.* 12(2), e14 (2010).
 - 26 Paolino AR, McGlynn EA, Lieu T *et al.* Building a governance strategy for CER: the Patient Outcomes Research To Advance Learning (PORTAL) network experience. *EGEMS (Wash. DC)* 4(2), 1216 (2016).
 - 27 Bell EA, Ohno-Machado L, Grando MA. Sharing my health data: a survey of data sharing preferences of healthy individuals. *AMIA Ann. Symp. Proc.* 2014, 1699–1708 (2014).
 - 28 Stone MA, Redsell SA, Ling JT, Hay AD. Sharing patient data: competing demands of privacy, trust and research in primary care. *Br. J. Gen. Pract.* 55(519), 783–789 (2005).