



Published in final edited form as:

*Genet Med.* 2010 August ; 12(8): 486–495. doi:10.1097/GIM.0b013e3181e38f9e.

## Genomic Research and Wide Data Sharing: Views of Prospective Participants

Susan Brown Trinidad, MA<sup>1</sup>, Stephanie M. Fullerton, DPhil<sup>1</sup>, Julie M. Bares, MPH<sup>1</sup>, Gail P. Jarvik, MD, PhD<sup>1</sup>, Eric B. Larson, MD, MPH<sup>2</sup>, and Wylie Burke, MD, PhD<sup>1</sup>

<sup>1</sup> University of Washington, Seattle, WA

<sup>2</sup> Group Health Research Institute, Seattle, WA

### Abstract

**Purpose**—Sharing study data within the research community generates tension between two important goods: promoting scientific goals and protecting the privacy interests of study participants. The present study was designed to explore the perceptions, beliefs, and attitudes of research participants and possible future participants regarding genome-wide association studies (GWAS) and repository-based research.

**Methods**—Focus group sessions with (1) current research participants, (2) surrogate decision-makers, and (3) three age-defined cohorts (18–34 years, 35–50, >50).

**Results**—Participants expressed a variety of opinions about the acceptability of wide sharing of genetic and phenotypic information for research purposes through large, publicly accessible data repositories. Most believed that making de-identified study data available to the research community is a social good that should be pursued. Privacy and confidentiality concerns were common, though they would not necessarily preclude participation. Many participants voiced reservations about sharing data with for-profit organizations.

**Conclusions**—Trust is central in participants' views regarding GWAS data sharing. Further research is needed to develop governance models that enact the values of stewardship.

### Keywords

data sharing; genetics; electronic medical records; privacy; participant perspectives

## INTRODUCTION

Recent technological advances have decreased the expense and increased the feasibility of genome-wide association studies (GWAS), and still more comprehensive genomic investigation, in the form of whole-exome research and full genome re-sequencing, is on the horizon. Because the contribution of individual gene variants to common diseases tends to be small, and because more definitive mutations tend to be quite rare, these forms of research require very large sample sizes – in some cases, tens of thousands of participants.<sup>1–2</sup> Sharing study data within the research community is an attractive solution to the problem of amassing sufficient datasets; it also promises to increase research efficiencies, maximizing the utility of existing datasets while minimizing participant burden. These

Corresponding author: Susan Brown Trinidad, MA, University of Washington, Health Sciences Building, A204, Department of Bioethics & Humanities, Box 357120, Seattle, WA 98195-7120, tel: (206) 543-2508, fax: (206) 684-7515, sbtrini@uw.edu.

**CONFLICTS OF INTEREST:** None to declare

benefits have informed policies of the National Institutes of Health (NIH) aimed at promoting data sharing.<sup>3-4</sup>

However, making such data available to the research community generates tension between two important goods: advancing scientific goals and protecting the privacy interests of study participants.<sup>5-9</sup> Because every person's DNA is unique, the traditional means of safeguarding research participants' privacy – de-identification of study data and biospecimens – does not guarantee protection.<sup>10-14</sup> In addition, trade-offs exist between de-identification and other possible participant concerns, such as the ability to receive individual research findings or the ability to withdraw from research participation.<sup>9, 15-16</sup>

Numerous prior studies have characterized potential participants' views about willingness to participate in biobanks and related forms of population-based genomic research and how informed consent ought to be handled.<sup>17-33</sup> There are also some published reports regarding participants' views about research access to medical record data.<sup>34-39</sup> However, relatively little is known about participants' and the general public's attitudes and perceptions regarding newer data-sharing mechanisms, such as the Federal database of Genotypes and Phenotypes (dbGaP), designed to make large amounts of genotypic and phenotypic information about individual participants available to any qualified researcher. McGuire and colleagues have described research participants' preferences regarding informed consent for public release of data,<sup>40</sup> while Kaufman et al. have investigated public opinions regarding a large prospective genetic cohort study being contemplated by the National Human Genome Research Institute (NHGRI).<sup>41-42</sup> Lemke et al. recently reported the results of focus groups with biorepository participants and members of the general public, in which they found "varying views" with respect to data sharing.<sup>43</sup> The present study was designed to explore the perceptions, beliefs, and attitudes of research participants and possible future participants regarding GWAS and repository-based research. In this paper, we report study findings with respect to participants' views about data sharing.

The study was carried out as part of the electronic Medical Records and Genomics (eMERGE) Network, a research consortium funded by the National Human Genome Research Institute (NHGRI) and the National Institute of General Medical Sciences (NIGMS) to explore the feasibility of using electronic medical record (EMR) data to derive reliable phenotypic data for use in GWAS. Our project, a partnership between the Group Health Research Institute, the University of Washington, and the Fred Hutchinson Cancer Research Center, is using an existing dataset from the Adult Changes in Thought (ACT) Study to perform proof-of-concept GWAS of dementia, carotid artery atherosclerotic disease, and adverse events associated with statin use. The project also includes an aim specifically targeted at understanding the ethical and social implications of such research, with the ultimate goal of informing policy development.

The ACT Study is cohort study of aging and dementia and the successor to a model Alzheimer's Disease Patient Registry funded since 1986 by the National Institute on Aging.<sup>44</sup> The cohort study began in 1994 with the enrollment of a randomly selected population of 2,581 persons over age 65 who were known not to have dementia at the time of enrollment. The project has been focused on detection of markers and risk factors for Alzheimer disease and related dementias, as well as age-related cognitive decline. Related studies address the relationship of mild cognitive impairment and insulin resistance<sup>45</sup> and population-based pharmacoepidemiologic neuropathology.<sup>46</sup> Because of ongoing replacement sampling, approximately 2,000 living members are currently enrolled; another 2,000 study participants have died after enrolling in the study. Participants are followed over time, completing periodic study interviews and cognitive tests at study visits every other year. In addition, a

rich array of clinical and pharmacy data about each participant is available through the Group Health EMR and other electronic data systems.

## METHODS

Between March and August 2008, we conducted a series of 10 focus group discussions at Group Health Cooperative, a large health maintenance organization based in the Seattle metropolitan area. Two separate sessions were held with representatives of 5 selected populations: (A) current research participants in the ACT Study, (B) individuals with decision-making authority on behalf of incapacitated ACT participants, (C) Group Health members aged 18–34 years, (D) Group Health members aged 35–50 years, and (E) Group Health members aged >50 years who were not in the ACT Study. Because the composition of any given focus group can affect group dynamics in unpredictable ways, we held 2 sessions within each population (e.g., Group A participants were in either Session A1 or Session A2). To be eligible to participate any of the sessions, individuals needed to be able to communicate in English and attend in person. For Groups A, C, D, and E, current enrollment in Group Health was also required. To be eligible for Group B participation, individuals had to recall having given consent on behalf of the ACT Study participant under their care.

### Study Design

The ACT Study participants (Group A) were included in this investigation because this is the study group being used in the eMERGE project, and we wanted to understand ACT participants' thoughts and questions about GWAS and data sharing for use in future communications. In addition, ACT participants represent individuals who have enrolled in a long-term study that includes a genetic component, rather than general members of the public who may or may not be willing to participate in such research.

Our study setting also afforded an opportunity to explore the perceptions of surrogate decision-makers (Group B) with respect to sharing study data. The ACT Study population includes many participants who have experienced cognitive decline since enrollment or died while being followed for the study. When an ACT participant has been diagnosed with dementia, their continued study participation is authorized (or not) by a legally authorized representative or surrogate. Participants in Group B either held current decision-making authority for a living ACT participant or had previously been responsible for a participant who was deceased at the time of the focus group session.

Inclusion of the three age-stratified groups (Groups C, D, and E) was designed to address potential concerns about the generalizability of our findings. In addition, we wanted to understand whether differences in beliefs, attitudes, and perceptions about data sharing may be correlated with age. Prospective observational studies with young adults represent a valuable research resource for high-throughput genomic investigations, but little is known about this group's attitudes toward such research. In particular, controversy exists over the question of whether younger adults' adoption of web-based communications and social networking tools has resulted in a lower level of concern regarding personal privacy.<sup>47–48</sup> We were unable to find peer-reviewed reports that considered research participation and wide data sharing in this light.

Focus groups enable researchers to observe how opinions about the issues under study coalesce or diverge within a relatively homogenous group.<sup>49</sup> These guided discussions are an effective and time-efficient means of gathering data for the purposes of policy development and public education, particularly when questions of acceptability are salient and the subject under investigation is complex.<sup>50–51</sup> Importantly for this study, focus groups

can elicit information from people who may be intimidated by or unwilling to participate in interviews, who have trouble responding to written surveys, who feel they “have nothing to say,” or who may not believe they have sufficient subject-area expertise to share their thoughts about the topic of interest. This method is also well suited to gathering potentially critical feedback, as individuals may feel more comfortable sharing negative comments when they are part of a larger group.<sup>52</sup> All plans and study instruments for the focus groups were approved by the Group Health Human Subjects Review Committee and were developed in accordance with accepted methods for this type of research.<sup>53–54</sup> Written informed consent was obtained from all focus group participants.

### Focus Group Pilot

To test the planned recruitment approach and refine the draft discussion guide, a pilot focus group was convened in March 2008 with 5 Group Health members >50 years of age. Light refreshments and a participation incentive of \$50 were provided, and participants received paid parking or taxi service. The session lasted 2 hours, followed by 1 hour of debriefing and discussion with focus group participants. Substantive revisions were made to the focus group guide based on the trial discussion and on participants’ feedback. Changes included starting with a few open-ended questions to assess the group’s familiarity with basic genetic concepts and health research more generally; providing education on genetics, GWAS, and informed consent as needed; sharpening the hypothetical examples posed for discussion; and reordering the discussion topics to promote participants’ comprehension and facilitate the flow of conversation more effectively.

### Recruitment

Prospective participants in each of the 5 population groups were randomly identified using Group Health automated records. Prior to recruitment, ACT Study staff screened the list of candidates for Group A and removed those who had experienced cognitive decline, would have difficulty traveling to downtown Seattle, or were otherwise inappropriate to contact in the study timeframe (e.g., current hospitalization, recent death of a spouse). The initial recruitment contact was a letter that described the study, explained what would be involved in participating, and told potential participants to expect a follow-up call inviting them to participate. Up to 3 attempts were made to contact candidates by telephone to ascertain their willingness to take part. Those who agreed to participate were then contacted by telephone or e-mail to schedule the evening focus group sessions and provide logistical information. A packet of written materials, including the consent form, study description, and directions to the Group Health Research Institute, was mailed to all participants prior to their scheduled session. Participants were offered the same payments and reimbursements as for the pilot session.

A total of 969 letters were mailed to prospective participants. We were unable to contact 293 of these individuals by telephone. Of the 676 who were successfully contacted, 124 (18%) were ineligible. Ineligible candidates were those who had disenrolled from Group Health (23% of ineligibles), were mentally or physically unable to participate (18%), had moved out of the area (15%), had language barriers (7%), had died (5%), or – for Group B participants – did not recall having given consent for the ACT Study (18%). Another 14% were classified as “other.” Of those who were contacted and eligible to participate, 355 (64%) declined and 197 (36%) agreed to participate. Reasons reported for declining participation were time/too busy (39%), lack of interest (24%), location inconvenient (6%), timing inconvenient (5%), and caring for a sick family member (1%). The remaining 25% declined to state a reason. Coordinating the schedules of those who wished to be in the study led to a total of 79 participants (14% of contacted eligible candidates) being recruited into the focus groups.

Demographic characteristics for the 5 groups are shown in Table 1. Focus group participants ranged in age from 18 to 89, with the mean age of ACT Study participants (Group A) approximately 20 years older than the oldest group of general Group Health members (Group E), 80.4 vs. 62.7. Surrogate decision-makers for ACT participants were approximately the same age as Group E members, whereas Groups C and D were, on average, 40 and 20 years younger respectively. Overall, focus group participants were evenly balanced with regard to sex, with more men participating in Groups D and E and more women participating in Group B. Focus group participants were overwhelmingly likely (89% overall) to identify their race as white, and the majority (60%) reported annual household incomes exceeding \$50,000. Focus group participants were also, as a group, very highly educated, with 83% reporting post-secondary levels of education. Forty-two percent of all participants reported taking part in health research in the past; excluding the ACT Study participants in Group A, 25% of focus group participants reported prior research involvement. Group Health has not routinely collected data on enrolled members' race/ethnicity, socioeconomic status, or educational attainments. The demographic data reported here were collected as part of this study; comparable data are not available for those we could not reach or who declined to participate. General demographic information, however, suggests that focus group participants were representative of Group Health members: 85% of current enrollees are white and 84% have at least some college education (K. Ehrlich, personal communication).

### Focus Group Discussions

The focus group discussions were held at the Group Health Research Institute during early evening hours between May and August 2008. Each session lasted 2 hours and included 5–9 participants. Two members of the research team (SMF and SBT) co-facilitated the group discussions, and another (JMB) took notes and provided logistical support. The first 15 to 25 minutes of each session were spent introducing the aims and mechanics of GWAS, with particular emphasis on the need for large datasets (and thus the importance of data sharing), the nature and comprehensiveness of genetic data generated in the course of GWAS, and the role of EMR-derived phenotypic data. The full discussion guide is available from the authors on request.

To frame the data-sharing discussion, we asked participants to consider a series of hypothetical scenarios in which they were to imagine that they were taking part in an ongoing genetic study. The data-sharing portion of the discussion guide is shown in Table 2. Participants were asked to consider questions such as whether they would want to limit research access to their EMR data, whether it would be acceptable for their de-identified genetic information to be shared outside Group Health, and under what circumstances (if any) they would wish to be contacted by the research team. All sessions were audio-recorded and transcribed for analysis. Transcripts were proofed against the audio-recordings, which were then destroyed.

### Analysis

Immediately following each session, the on-site investigators debriefed in person and one of us was assigned to draft field notes. The other team members who attended the session added their impressions, and then the complete field note was reviewed by another team member (WB) who was not present in the sessions. This step gave us the opportunity to determine whether there were any gaps in the discussion guide and offered an interim “reality check” on consistency of approach across the 10 sessions.<sup>55</sup> As more sessions were completed, a comparative element was included in the field notes; this form of collaborative memoing helped us to identify emerging themes and concepts and to reach analytic consensus within the research team.<sup>56–57</sup> When the transcripts had been proofed and the

sessions completed, we closely read and re-read the transcripts and our field notes, writing margin notes and conceptual memos as we went. We then performed a summative content analysis with the aim of describing participants' views regarding data sharing.<sup>58–59</sup>

## RESULTS

The focus groups were designed to elicit participants' views on a number of issues that can arise in the conduct of GWAS, including when re-consent and the return of individual research findings may be appropriate. This report focuses on our results with regard to wide data sharing; other findings are being prepared for publication. Summary findings are presented in Table 3. All quoted text in this section represents participants' words.

Overall, participants endorsed the value of data sharing and, while they recognized some risks, most considered the potential benefit of high-throughput genomic research to outweigh the possible harms. As one participant put it, "At the same time as I can see some tremendous assets to having [dbGaP], because you can really do something powerful, I think there's always risk. In this case, I tend to think, well, with that potential of where we are in terms of understanding the genome, maybe that's a benefit and maybe, if it's securely regulated and actually looked after, maybe that's a risk worth taking."

### Acceptability of Wide Data Sharing and Willingness To Participate

Most participants saw the pooling of research resources as a reasonable approach to enhancing efficiency, avoiding duplication of effort, hastening the development of outcomes that would benefit public health, and creating a reference of "historical value" for future generations. Participants told us, "I think there does have to be an open exchange of information in order for some of these really significant things to happen for people's benefit," "I think some very interesting things may turn up because of that. That vast amount of information has got to have some really positive effects for everybody," and "I think the whole thing's just a marvelous idea." One participant remarked, "To me, the more information researchers have, the better, as long as you [can protect against discrimination]. I mean, that's what research is, and you're crippling it by not allowing them to share. And they can't make advances, you know, if they can't – I mean, they can advance quicker [if they share], I would think. I would hope." Participants believed that the value of such resources lies in (1) the completeness and accuracy of the data and (2) its accessibility to many different researchers investigating many different questions.

We asked participants whether knowing that study data would be deposited in dbGaP would affect their willingness to participate in a genetic study. Most did not see data sharing as a reason not to participate, and some said that it would encourage them to sign up. As one person commented, "It would be another reason to do it." Some told us that they would be gratified to know that their contribution would continue to be useful: "It's rewarding to know that I didn't just dabble a bit, got in the one study, but ... roses keep on growing." Others saw a practical benefit to participating in a study that would maximize the utility of their contributions. The longitudinal and ever-growing nature of dbGaP was also viewed favorably by most participants. This was especially so in Groups A and E, in which older participants spoke of the continued use of their research data as a "legacy, living on in the lab," and a way to contribute to society even after death. Another participant said, "It makes me a little less mortal. Not immortal, but a little less mortal."

There was general agreement among surrogate decision-makers (Group B) that the ACT Study participants they represented would have had no problem with having their study data sent to dbGaP. "I'm the power of attorney for a family friend who is in the late stages of dementia, and she actually volunteered for this study, recognizing that her father had

dementia, she wanted to participate and, you know, have her body donated in any and all research to gain more information. So I think she would have been very supportive of this.” Another said, “I know my mother would say fine. She definitely would have gone for it.” In another exchange, one participant said, “My aunt would have helped,” and another responded, “My mother-in-law too.” Surrogates were able to separate their own views and preferences from those of their charge: while some Group B participants said that while they personally may have misgivings about GWAS participation, these concerns were not shared by the ACT Study participants they knew. As described below, the surrogate decision-makers’ perceptions of ACT enrollees’ views were consistent with what we heard directly from ACT Study participants (Group A).

A minority of focus group participants considered having research data deposited in dbGaP as a reason not to take part in GWAS. These individuals saw research involving data sharing as a qualitatively different, and riskier, activity compared with other kinds of studies. One participant remarked, “It’s a leap of faith to go from a bunch of researchers to a Federal database, and it’s not one – if I knew, I would never have signed up for that [hypothetical] study if I thought even any of that information was going to go off ...”

### Who Should Have Access to Data

ACT Study participants (Group A) were largely in favor of data-sharing with researchers outside of Group Health in the name of efficiency. The surrogate decision-makers in Group B, most of whom were not Group Health members, were more cautious about the potential sharing of their own data, and they expected to be informed if their loved one’s information were to be widely shared. The youngest group we spoke with, Group C, expressed a range of opinions, from no concern about data sharing to requiring detailed information as part of the informed consent process (and the possibility that data sharing would be a reason not to participate in research). In Group D, there was some disagreement about whether any sharing not specifically described in the consent form was acceptable, even within Group Health. Participants in Group E generally felt that data sharing was a good thing and noted that even international sharing should be encouraged, both because “the same diseases affect us here in this country that affect people around the world” and because of an expectation of reciprocity: “If everybody keeps secrets ... They may know something that will save my great-grandkids, and if I don’t share mine, why should they share theirs? So it’s in everybody’s interest to have as much information [as possible] out there in the pool.”

Participants generally agreed that sharing with other Group Health investigators and close collaborators (such as those at local academic institutions) would be acceptable, as would sharing with non-profit, public-interest research organizations (e.g., the American Cancer Society). Such organizations were viewed as “more legitimate,” because participants believed that these kinds of institutions conduct “pure science” aimed at benefiting the general public and advancing knowledge, rather than generating financial returns. (Some participants identified exceptions to this rule, e.g., corporately funded non-profit organizations, such as research institutes funded by the tobacco industry, whose financial interests could be advanced or impeded by certain research results.) A few people expressed misgivings about the potential for insurance discrimination to occur within Group Health, which has functions in clinical care, insurance, and research.

Current research participants, who generally expressed altruistic motivations for research participation as well as strong trust in Group Health, were willing to rely on Group Health’s internal review processes and trusted Group Health to “be selective” about granting access to outside entities. For most participants, concerns began to arise as they considered more “distant” users of the data. Many participants expressed misgivings about sharing data with for-profit entities; in half of the sessions (A2, B1, B2, D1, D2) participants raised the issue

before we asked about it. These participants often perceived a mismatch between the altruistic motivations of research participants and the fiscal goals of for-profit companies, as reflected in this comment from a person who had participated in a breast cancer study: “I gave all my medical records, I signed permission – ‘Use anything you want.’ It was in a Group Health context. Yes, they could have gone to [*a local cancer research institute*], yes, they could have gone to [*a local research university*], yes! Could they have gone to [*a large pharmaceutical company*]? No!”

We also heard that some participants felt that genetic information should not be patentable, and that it would be unethical to use public resources in “profit-seeking” activities. Although our questions were generic (we asked about sharing with “for-profit organizations”), participants in all groups expressed distrust of the motives, ethics, and research and marketing practices of pharmaceutical companies. Some thought it was unfair that research participants could be made to “pay twice” (or more) for commercial products resulting from the use of their data, once through their study contributions, and again through their taxes, pocketbooks, or insurance. There were counterbalancing opinions on this point, with some noting that industry partners are needed in order to translate research results into tangible products: “I don’t see how you could avoid giving this out to for-profit companies. If this study is of any use at all, they are going to have to make it available to a wide group of experimenters, and there are no wide groups of experimenters that don’t have something to do with for-profit companies.” Several participants commented that perhaps for-profit users could be required to pay NIH or Group Health for data access.

### Governance Concerns

While some participants trusted the Federal government to manage dbGaP and similar repositories in a responsible manner, others worried about the potential for abuse. Distrust with regard to the possibility of Federal agencies’ obtaining research data for purposes other than research was expressed in every session. In some sessions, strong trust in Group Health was contrasted with a lack of trust in the Federal government. As one participant stated, “This is the privacy issue: that there’s no failsafe, as far as I’m concerned. And I would trust researchers, but I don’t trust the insurance industry, and I don’t trust the government.” Participants voiced two kinds of concerns. One was with regard to inappropriate use of data by law enforcement or national security agencies, and the other was regarding the possibility of a “tyrannical government” using such data for eugenics or other objectionable purposes: “I don’t really have a problem with it as it stands now, however, the future thought of Big Brother watching you and the government getting involved in doing all these things is scary, just because I think ... trust in the government isn’t real high right now, and if they were to, I mean, if government really got involved and insisted on doing this stuff, I mean, I could see where they could genetically do everything they wanted to do. And it’s scary.”

Participants saw a need for trustworthy governance to ensure that both practical and ethical goals – advancing science and protecting research participants – would be achieved. As one participant noted, “I think the key is finding the right balance between letting science and research go along and make great discoveries and not throttling them back with public policy issues. Ideally, we could kind of work them together so that science could move ahead and the Congress and other bodies could work alongside to make sure the protections are there.” Another said, “I would want to do more than trust [the managers of the data repository]. I would hope that the Group Health institution and the National Institutes of Health and others would also be very aggressive about safeguards.” A related concern had to do with what participants saw as the inevitability of changes in law and regulations. “You just don’t know what your ‘yes’ really means down the line. We’ve all grown up realizing how nothing seems to be sacred, and how the most secure information somehow gets found and used and abused,” according to one participant.



The obligations of users of shared data came up without prompting from the facilitators, with concerns about whether such users would be held to the same standards as Group Health: “My question would be, do the rules that the first group signed on with, apply to the group that gets handed the new information? If we sign consent information forms and all that kind of stuff, what’s the obligation of the second group to follow those guidelines?”

### **Inclusion of Data from Participants’ Electronic Medical Records**

One focus of the eMERGE Network is to assess the feasibility of GWAS using EMR-derived phenotypes, requiring the sharing of some clinical information with the dbGaP repository. Participants understood the research value of such information. However, many participants saw medical record information as potentially more sensitive than genetic data, in part because of the potentially stigmatizing nature of information that could be contained in the EMR, such as reproductive health information and mental health history: “I can see [that] your personal health record, if it’s a carte blanche to share anything that’s in there, a lot of us might have reservations.” Some participants were uncomfortable with the idea that information they had shared confidentially with their healthcare provider could be made available to researchers.

Participants in Group A and E agreed that there were no specific parts of the medical record that they would like to withhold, and some mentioned that sharing the entire medical record without reservation would be of greatest utility to research. Group B participants had a range of opinions regarding the use of EMR data. Some were comfortable with allowing open access to the complete record, while others questioned the utility of such access. Some individuals were agreeable to making their records freely available, while others would not personally consent to such access. (“I doubt that I would participate in a study that involved universal access to my healthcare records.”) Participants in Group C, the youngest group we spoke with, advocated direct control over how much and which parts of their medical records would be available for research purposes (“I think it would also be helpful to have some way of my being able to take control of that process and being able to check boxes of, like, ‘It’s fine to have this information, but you don’t have permission – when it comes to, like, you don’t have the right to my reproductive health [information], but you do to my blood pressure.”) Others noted that it could be difficult to operationalize such an approach, particularly given that researchers may not know in advance which variables would prove to be needed for a future study for research purposes. In Group D, several participants thought that the sharing of physician’s notes would not be appropriate; even the most altruistic participant thought that the text of patient-provider conversations held in confidence should be off limits.

Notwithstanding these concerns, most participants were generally willing to have some EMR data shared for research purposes, provided that the data were fairly limited (e.g., if only “strictly scientific things ... like diagnostic codes and medications” were included), well defined (e.g., if they knew ahead of time what data would be extracted for research use), and de-identified (with links to personal identifiers maintained at Group Health): “If it’s anonymized, and Group Health is the protector, I wouldn’t have a problem.”

### **Privacy and Confidentiality Concerns**

Although we did not raise the issue of privacy directly, it was an underlying theme throughout the discussions, and most participants had at least some privacy concerns. Participants in Groups A and E were substantially less worried about privacy and confidentiality than other groups. Many of those we talked with, however, said that they believed the potential benefits of wide data sharing outweighed potential risks: “I guess it comes down to a balance. How much good is expected from it, against that extreme risk that

might – might – happen. You can't weigh that. I'd say you can't even weigh it today, let alone 5 years from now. So you just kinda take it on faith, and do it."

A recurring theme in all groups was the inevitability of "protected" data being accidentally released (several participants mentioned stolen laptops containing confidential data) or otherwise accessed by unauthorized persons. Focus group participants simply did not believe that data security can be guaranteed, despite researchers' good intentions: "There's no realistic way of controlling [the data], once you share it. Let's face it." and "Unless we go back to working out of a shoebox, there's no security at all." At the same time, most people felt that the risk of breach of confidentiality is commonplace in modern life, as demonstrated by this exchange in Session B1:

SPEAKER 1: "As soon as there's a database, and it's on a computer, sooner or later there is a thing where bank records all of a sudden get lost, or somebody steals them, or somebody hacks them, or somebody's personal computer gets stolen out of their home, and all of a sudden it's gone. And it's bank records or it's hospital records, and this happens several times a year, it's in the paper. There's ten thousand records that were supposed to be private, are now unaccounted for."

SPEAKER 2: "However, having said that, and knowing that this happens, we don't stop using banks! No, we don't. We don't stop those kinds of things. We do everything we can to divorce our personal information from uses we haven't authorized, but we still, just because of the complexity of life, are involved with insurance companies and banks and employers – and, hopefully, health research."

Some participants believed that health information would be a less attractive target for ill-intentioned individuals than other kinds of data (such as financial records or credit-card information). To a number of participants, a confidentiality breach regarding banking information or other personal information that could be used for purposes of identity theft would be a greater cause for concern than would unauthorized access to their de-identified genetic information.

Some participants saw the very large size of typical GWAS (and associated databases) as conferring a certain degree of privacy protection, citing "safety in numbers" as reducing the risk that they would be personally identified or harmed as a result of research participation. "You know, there's something that feels more comfortable about a huge study. You're kind of lost in that huge sea of information, and it really seems like fewer risks." "It seems to me, as you increase the amount of data, your individuality is really getting more and more lost. You are just a much smaller part of a large data pool." But not all participants agreed with this idea: "If I were to share my DNA and medical record, I would add one drop to this ocean of statistics. But if something were to go wrong, that would have a great effect on my life."

Participants felt that robust privacy protections would be necessary to ensure the quality of the data to be deposited, for two reasons: first, enrollment would be higher and the ultimate value of the resource maximized if potential participants believed that appropriate steps would be taken to protect the data (acknowledging that such protections are not absolute). One person commented, "Don't you think that if the safeguards get lessened, people will stop saying, 'Ok, I'll give my DNA?' They'll stop being a part of the study if they perceive it isn't safe, and then our information will be just kind of dead-ended, because they're going to have to have information over a long period of time to see what changes. It's never going to be static. It's going to have to go on and on and on." Second, research participants may be tempted to hedge or withhold potentially important self-reported information if they do not trust that it would be kept confidential: "One of the thoughts that comes to mind is the

validity of the data somewhat could depend on the confidentiality, because a person might be a little hesitant to be really honest and outright if they felt uncomfortable about it.”

## DISCUSSION

Participants in the focus groups understood the rationale for wide data sharing, especially in the context of GWAS and related genomic approaches, and believed that making de-identified study data available to the research community is a social good that should be pursued. Advantages identified by participants fell into three broad categories: increased research efficiency, benefit to patients and society, and respect for research participants. The value of maximizing research efficiency was embraced by all the groups, and participants favored efforts to reduce unnecessary duplication of effort, control costs, promote collaboration, and make the most of available resources. Participants also expressed the belief that broad data access would increase the potential for meaningful findings to be uncovered, and for health benefits to be realized in a more timely fashion. Most saw altruism as the primary reason anyone would agree to participate as a study subject; because of this, they saw researchers’ maximizing the use of subjects’ contributions as a respectful recognition and realization of subjects’ goals.

Privacy and confidentiality concerns were also common, though they were not necessarily a deal-breaker when it came to willingness to participate. While some participants considered the possibility of breach sufficient reason not to participate in genetic research, most considered the risk to be relatively small and worth taking in view of the potential benefits of such investigations. Several participants perceived the risks involved in data sharing to be substantially less concerning than the (often unavoidable) privacy risks they encounter in daily life. This finding raises interesting questions about whether such research may – at least in the minds of possible study participants – be classifiable as “minimal risk” under current regulatory definitions.<sup>60</sup>

While younger people may have been expected to be more comfortable with the technology that allows data sharing and thus express fewer privacy concerns, we found that older people were least worried about the potential for loss of confidentiality. Younger participants were more inclined to desire direct control over which data could be shared. This could reflect greater interest in privacy per se, or perhaps greater knowledge of the technological feasibility of user-controlled privacy settings, such as those used in online social networking applications. Older participants, by contrast, told us that they had nothing to hide and little to fear. These findings may be consistent with survey results reported by Kaufman and colleagues, in which younger respondents (under 60 years of age) were more concerned that research data could be used against them.<sup>41</sup>

A few limitations of this study should be noted. While our participants were fairly representative of Group Health Cooperative membership, that population tends to be slightly older, more highly educated, and less racially and ethnically diverse than in the Northwestern United States more generally and in the US as a whole. Even within that context, our focus group participants were very well educated: nearly three-quarters (73%) had earned at least a bachelor’s degree, and more than a third (37%) held advanced degrees. Some degree of selection bias was unavoidable, as Group Health Cooperative members who are unlikely to support genetic research and associated data sharing may have been unwilling to participate in these focus groups. It may also be the case that Group Health Cooperative members are generally more favorably disposed toward research, given its history as a consumer-governed organization and the Group Health Research Institute’s long-standing and locally well-publicized track record of health research. Many of our participants (25% excluding ACT Study enrollees) reported that they had taken part in

health research in the past. However, rather than limiting the generalizability of our findings, the unique features of the study population may offer an informative “extreme case.”<sup>61</sup> If a population with very high trust in both the researchers and the research institution nonetheless has significant reservations about data sharing, groups that do not have this kind of relationship with researchers may well be more likely to have such concerns.

In contrast with the academic community’s emphasis on human subjects protections and regulatory controls, focus group participants viewed trust between researchers and subjects as central. While they did consider informed consent important, most focus group participants seemed to believe that this tool is but a small part of the governing relationship between researchers and subjects, especially in a context in which future recipients and downstream uses of data may be difficult to predict. Considering the potential harm that could result from breach of confidentiality, for example, participants did not want – and did not place much if any credence in – assurances that their information would be kept private. Paradoxically, strong assurances regarding privacy were seen by some participants as generating less trust in researchers. They did, however, want researchers to promise that they would protect study data to the very best of their ability, provide an honest accounting of any breaches that may occur, and make their “best effort” to mitigate any negative effects. This shift from legalistic guarantees to personal commitments signals a fundamental difference in how the relationship between researchers and study participants could be construed.<sup>62–63</sup>

Our findings point toward the need for investigation regarding governance models that enact the values of stewardship. To develop research practices that foster trust and trustworthiness, more dialogue between the research community and the lay public is needed; and the issue of trust (or lack of trust) in the Federal government must also be addressed. These engagements should explore ways of maintaining participants’ and public trust around potentially contentious issues, including data access by for-profit entities, procedures for Federal oversight and accountability, the return of individual research findings to participants, and the conditions under which researchers should re-contact participants to seek permission for a new use of existing study data. Such efforts will help the research community to proactively address participant expectations for the coming era of high-throughput population-based genomic research.

## Acknowledgments

The eMERGE Network was initiated and funded by NHGRI, in conjunction with additional funding from NIGMS through grant no. U01-HG-004610. GPJ received additional support from a Washington State Life Sciences Discovery fund grant to the Northwest Institute of Genetic Medicine. The Group Health/University of Washington Alzheimer’s Disease Patient Registry/Adult Changes in Thought study is supported by NIA grant no. U01-AG-06781. The authors thank Kelly Ehrlich for her help with focus groups arrangements and demographic data; and Darlene White, Cheryl Wiese, and Dorothy Oliver for their assistance in recruitment. We are grateful for helpful comments and suggestions from the journal editor and an anonymous reviewer, which improved this paper substantially.

## References

1. Chasman DI, Pare G, Mora S, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet* Nov;2009 5(11):e1000730. [PubMed: 19936222]
2. Ganesh SK, Zakai NA, van Rooij FJ, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* Nov;2009 41(11):1191–1198. [PubMed: 19862010]
3. Final NIH Statement on Sharing Research Data. 2003 [Accessed December 22, 2009, 2009]. <http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html>

4. Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). 2007 [Accessed December 22, 2009]. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html>
5. Bathe OF, McGuire AL. The ethical use of existing samples for genome research. *Genet Med* Oct; 2009 11(10):712–715. [PubMed: 19745750]
6. Church G, Heeney C, Hawkins N, et al. Public access to genome-wide data: five views on balancing research with privacy and protection. *PLoS Genet* Oct;2009 5(10):e1000665. [PubMed: 19798440]
7. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics--reshaping scientific practice. *Nat Rev Genet* May;2009 10(5):331–335. [PubMed: 19308065]
8. Ashburn TT, Wilson SK, Eisenstein BI. Human tissue research in the genomic era of medicine - Balancing individual and societal interests. *Archives of Internal Medicine* Dec;2000 160(22):3377–3384. [PubMed: 11112229]
9. Eriksson S, Helgesson G. Potential harms, anonymization, and the right to withdraw consent to biobank research. *Eur J Hum Genet* Sep;2005 13(9):1071–1076. [PubMed: 15986039]
10. McGuire AL, Gibbs RA. No longer de-identified. *Science* Apr;2006 312(5772):370–371. [PubMed: 16627725]
11. Lin Z, Altman RB, Owen AB. Confidentiality in genome research. *Science* Jul;2006 313(5786):441–442. [PubMed: 16873628]
12. Homer N, Szlinger S, Redman M, et al. Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. *PLoS Genet* 2008;4(8):e1000167. [PubMed: 18769715]
13. Malin BA. An evaluation of the current state of genomic data privacy protection technology and a roadmap for the future. *Journal of the American Medical Informatics Association* Jan-Feb;2005 12(1):28–34. [PubMed: 15492030]
14. Karp DR, Carlin S, Cook-Deegan R, et al. Ethical and practical issues associated with aggregating databases. *PLoS Med* 2008;5(9):e190. [PubMed: 18816162]
15. Greely HT. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu Rev Genomics Hum Genet* 2007;8:343–364. [PubMed: 17550341]
16. Caulfield T, McGuire AL, Cho M, et al. Research ethics recommendations for whole-genome research: consensus statement. *PLoS Biol* Mar 25;2008 6(3):e73. [PubMed: 18366258]
17. Aagaard-Tillery K, Sibai B, Spong CY, et al. Sample bias among women with retained DNA samples for future genetic studies. *Obstet Gynecol* Nov;2006 108(5):1115–1120. [PubMed: 17077232]
18. Bhatti P, Sigurdson AJ, Wang SS, et al. Genetic variation and willingness to participate in epidemiologic research: Data from three studies. *Cancer Epidemiology Biomarkers & Prevention* Oct;2005 14(10):2449–2453.
19. Chen DT, Rosenstein DL, Muthappan P, et al. Research with stored biological samples -What do research participants want? *Archives of Internal Medicine* Mar;2005 165(6):652–655. [PubMed: 15795341]
20. Hoeyer K, Olofsson BO, Mjorndal T, Lynoe N. Informed consent and biobanks: a population-based study of attitudes towards tissue donation for genetic research. *Scandinavian Journal of Public Health* Jun;2004 32(3):224–229. [PubMed: 15204184]
21. Kaphingst KA, Janoff JM, Harris LN, Emmons KM. Views of female breast cancer patients who donated biologic samples regarding storage and use of samples for genetic research. *Clinical Genetics* May;2006 69(5):393–398. [PubMed: 16650074]
22. Kettis-Lindblad A, Ring L, Viberth E, Hansson MG. Genetic research and donation of tissue samples to biobanks. What do potential sample donors in the Swedish general public think? *European Journal of Public Health* Aug;2006 16(4):433–440. [PubMed: 16207726]
23. McCarty CA, Nair A, Austin DM, Giarnpietro PF. Informed consent and subject motivation to participate in a large, population-based genomics study: The Marshfield Clinic Personalized Medicine Research Project. *Community Genetics* 2007;10(1):2–9. [PubMed: 17167244]
24. McQuillan GM, Pan QY, Porter KS. Consent for genetic research in a general population: An update on the National Health and Nutrition Examination Survey experience. *Genetics in Medicine* Jun;2006 8(6):354–360. [PubMed: 16778597]

25. Pentz RD, Billot L, Wendler D. Research on stored biological samples: views of African American and White American cancer patients. *Am J Med Genet A* Apr 1;2006 140(7):733–739. [PubMed: 16523508]
26. Pulley JM, Brace MM, Bernard GR, Masys DR. Attitudes and perceptions of patients towards methods of establishing a DNA biobank. *Cell Tissue Bank* Mar;2008 9(1):55–65. [PubMed: 17960495]
27. Schwartz MD, Rothenberg K, Joseph L, Benkendorf J, Lerman C. Consent to the use of stored DNA for genetics research: A survey of attitudes in the Jewish population. *American Journal of Medical Genetics* 2001;98(4):336–342. [PubMed: 11170078]
28. Stegmayr B, Asplund K. Informed consent for genetic research on blood stored for more than a decade: a population based study. *Bmj* Sep 21;2002 325(7365):634–635. [PubMed: 12242175]
29. Sterling R, Henderson GE, Corbie-Smith G. Public willingness to participate in and public opinions about genetic variation research: A review of the literature. *American Journal of Public Health* Nov;2006 96(11):1971–1978. [PubMed: 17018829]
30. Treloar SA, Morley KI, Taylor SD, Hall WD. Why do they do it? A pilot study towards understanding participant motivation and experience in a large genetic epidemiological study of endometriosis. *Community Genetics* 2007;10(2):61–71. [PubMed: 17380055]
31. Treweek S, Doney A, Leiman D. Public attitudes to the storage of blood left over from routine general practice tests and its use in research. *J Health Serv Res Policy* January 1;2009 14(1):13–19. [PubMed: 19103912]
32. Wang SS, Fridinger F, Sheedy KM, Houry MJ. Public attitudes regarding the donation and storage of blood specimens for genetic research. *Community Genet* 2001;4(1):18–26. [PubMed: 11493749]
33. Wendler D, Emanuel E. The debate over research on stored biological samples - What do sources think? *Archives of Internal Medicine* Jul;2002 162(13):1457–1462. [PubMed: 12090881]
34. 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* Jan;2007 64(1):223–235. [PubMed: 17045717]
35. 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* Feb;2004 30(1):104–109. [PubMed: 14872086]
36. Kass NE, Natowicz MR, Hull SC, et al. The use of medical records in research: what do patients want? *J Law Med Ethics* Fall;2003 31(3):429–433. [PubMed: 14626550]
37. Purdy S, Finkelstein JA, Fletcher R, Christiansen C, Inui TS. Patient participation in research in the managed care environment: Key perceptions of members in an HMO. *Journal Of General Internal Medicine* Jul;2000 15(7):492–495. [PubMed: 10940136]
38. Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. *Int J Med Inform* Jul;2006 75(7):530–541. [PubMed: 16198142]
39. Willison DJ, Keshavjee K, Nair K, Goldsmith C, Holbrook AM. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. *Bmj* Feb 15;2003 326(7385):373. [PubMed: 12586673]
40. McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A. DNA data sharing: research participants' perspectives. *Genet Med* Jan;2008 10(1):46–53. [PubMed: 18197056]
41. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. *Am J Hum Genet* Nov;2009 85(5):643–654. [PubMed: 19878915]
42. Kaufman D, Murphy J, Scott J, Hudson K. Subjects matter: a survey of public opinions about a large genetic cohort study. *Genet Med* Nov;2008 10(11):831–839. [PubMed: 19011407]
43. Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. Public and biobank participant attitudes toward genetic research participation and data sharing. *Public Health Genomics*. 2010
44. Kukull WA, Higdon R, Bowen JD, et al. Dementia and Alzheimer disease incidence: a prospective cohort study. *Arch Neurol* Nov;2002 59(11):1737–1746. [PubMed: 12433261]
45. Sonnen JA, Larson EB, Brickell K, et al. Different patterns of cerebral injury in dementia with or without diabetes. *Arch Neurol* Mar;2009 66(3):315–322. [PubMed: 19139294]

46. Li G, Shofer JB, Kukull WA, et al. Serum cholesterol and risk of Alzheimer disease: a community-based cohort study. *Neurology* Oct 11;2005 65(7):1045–1050. [PubMed: 16217057]
47. Kornblum, J. Online privacy? For young people, that's old-school. *USA Today*; Oct 27. 2007
48. Kirkpatrick, M. Facebook's Zuckerberg Says The Age of Privacy is Over. MacManus, R., editor. Vol. 20102010. ReadWriteWeb;
49. Kahan J. Focus groups as a tool for policy analysis. *Analysis of Social Issues and Public Policy* 2001;1(1):129–146.
50. Powell RA, Single HM. Focus groups. *Int J Qual Health Care* Oct;1996 8(5):499–504. [PubMed: 9117204]
51. Morgan, DL. *The Focus Group Guidebook*. Thousand Oaks, CA: Sage Publications; 1998.
52. Kitzinger J. Qualitative research. Introducing focus groups. *BMJ* Jul 29;1995 311(7000):299–302. [PubMed: 7633241]
53. Fern, EF. *Advanced Focus Group Research*. Thousand Oaks, CA: Sage Publications; 2001.
54. Stewart, DW.; Shamdasani, PN.; Rook, DW. *Focus Groups: Theory and Practice*. 2. Thousand Oaks, CA: Sage Publications; 2007.
55. Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods* 2002;1(2):1–19.
56. Montgomery P, Bailey PH. Field Notes and Theoretical Memos in Grounded Theory. *West J Nurs Res* February 1;2007 29(1):65–79. [PubMed: 17228061]
57. Corbin, J.; Strauss, A. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 3. Thousand Oaks, CA: Sage Publications, Inc; 2008.
58. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* Nov; 2005 15(9):1277–1288. [PubMed: 16204405]
59. Sandelowski M. Whatever happened to qualitative description? *Research in Nursing & Health* 2000;23(4):334–340. [PubMed: 10940958]
60. Code of Federal Regulations. Protection of Human Subjects. Vol 45 CFR 46.102(i)2005.
61. Gerring, J. *Case Study Research: Principles and Practices*. New York: Cambridge University Press; 2007.
62. McDonald M, Townsend A, Cox SM, Paterson ND, Lafreniere D. Trust in health research relationships: accounts of human subjects. *J Empir Res Hum Res Ethics* Dec;2008 3(4):35–47. [PubMed: 19385755]
63. Yarborough M, Fryer-Edwards K, Geller G, Sharp RR. Transforming the culture of biomedical research from compliance to trustworthiness: insights from nonmedical sectors. *Acad Med* Apr; 2009 84(4):472–477. [PubMed: 19318781]

Table 1

Demographics

|                              | Group A ACT participants | Group B ACT surrogates | Group C ages 18-34 | Group D ages 35-50 | Group E ages >50 | Total (%)         |
|------------------------------|--------------------------|------------------------|--------------------|--------------------|------------------|-------------------|
| N                            | 18                       | 16                     | 16                 | 14                 | 15               | 79 (100%)         |
| Age (yrs) <sup>d</sup>       |                          |                        |                    |                    |                  |                   |
| Mean                         | 80.4                     | 59.5                   | 27.1               | 42.8               | 62.7             | 56.4 <sup>b</sup> |
| Range                        | 71-89                    | 46-74                  | 18-34              | 37-49              | 51-84            | 18-89             |
| Sex                          |                          |                        |                    |                    |                  |                   |
| Male                         | 9                        | 5                      | 7                  | 10                 | 10               | 41 (52%)          |
| Female                       | 9                        | 11                     | 9                  | 4                  | 5                | 38 (48%)          |
| Race <sup>c</sup>            |                          |                        |                    |                    |                  |                   |
| White American               | 14                       | 16                     | 14                 | 13                 | 13               | 70 (89%)          |
| Indian/Alaska Native         | 0                        | 1                      | 0                  | 0                  | 0                | 1 (1%)            |
| Asian Native                 | 2                        | 1                      | 3                  | 1                  | 1                | 8 (10%)           |
| Hawaiian/Pacific Islander    | 0                        | 0                      | 1                  | 0                  | 0                | 1 (1%)            |
| Black                        | 1                        | 1                      | 1                  | 0                  | 1                | 4 (5%)            |
| Declined                     | 1                        | 0                      | 0                  | 0                  | 0                | 1 (1%)            |
| Household income             |                          |                        |                    |                    |                  |                   |
| \$0-\$25k                    | 3                        | 0                      | 2                  | 0                  | 1                | 6 (8%)            |
| \$25k-\$50k                  | 5                        | 5                      | 6                  | 1                  | 4                | 21 (27%)          |
| \$50k-\$75k                  | 5                        | 5                      | 3                  | 1                  | 0                | 14 (18%)          |
| \$75k-\$100k                 | 2                        | 3                      | 1                  | 5                  | 5                | 16 (20%)          |
| \$100k+                      | 2                        | 3                      | 3                  | 7                  | 2                | 17 (22%)          |
| Declined                     | 1                        | 0                      | 1                  | 0                  | 3                | 5 (6%)            |
| Education                    |                          |                        |                    |                    |                  |                   |
| High school or equivalent    | 4                        | 4                      | 3                  | 1                  | 1                | 13 (17%)          |
| Some college <sup>d</sup>    | 2                        | 1                      | 1                  | 0                  | 1                | 5 (6%)            |
| College degree               | 8                        | 5                      | 5                  | 5                  | 8                | 31 (39%)          |
| Postgraduate degree          | 4                        | 6                      | 6                  | 8                  | 5                | 29 (37%)          |
| Prior research participation | 18                       | 5                      | 3                  | 2                  | 5                | 33 (42%)          |

<sup>a</sup>6 participants did not report age



<sup>b</sup> Mean age of all participants (across groups)

<sup>c</sup> Some participants reported mixed heritage

<sup>d</sup> Short of degree requirements

Table 2

## Data-Sharing Portion of Discussion Guide

We've talked about the fact that this kind of research requires very large datasets. One way these studies can be done is to combine information from different studies to create one big dataset for research purposes. The data that would be combined would first be de-identified, meaning that no personal identifiers – such as your name, date of birth, or SSN – would be attached to your records. This is done to protect the confidentiality of private health information.

What do you think about this idea?

If you were a participant in a genetic study at Group Health,<sup>a</sup> what would you think if the researchers wanted to share your de-identified information:

- With other Group Health researchers?
- Outside Group Health, with researchers who are collaborating with Group Health (e.g., from the University of Washington)?
- Outside of Group Health and its research partnerships:
  - With other non-profit research organizations, like the government or disease research foundations?
  - With for-profit research organizations?
  - With researchers outside the US?

As described earlier, ideally a GWA study would consider genetic relationships with many different traits and conditions. Although researchers would gladly collect as much data as possible to achieve this, study participants might wish to have a say in exactly which of their personal health data is used for research, and how much of this is reasonably shared with outside researchers. We'd like to ask you a few questions about the kinds of information you would agree to share and why.

What would you think of having information from your medical record – the electronic chart your doctor keeps about your ongoing care – shared with Group Health researchers? Other researchers?

- Are there parts of the medical record you think should be withheld, or should only be shared with specific permission?

What about your genetic information? Would you be willing to have information about many of your genes (possibly your entire genome) shared broadly for the purposes of research?

- One issue that comes up in this context is the problem of de-identification: because your DNA is unique to you, this information can't truly be de-identified, even if it doesn't have your name on it. So, if someone already had a DNA sample with your name on it, they could (in theory) match that genetic information to a de-identified genetic profile in a repository. Does this affect your thinking about whether this information should be included?

There are many different ways to share the data needed to perform GWA studies. The National Institutes of Health, the Federal agency that funds most health-related research in the US, has recently instituted a giant database to make data-sharing easier: dbGaP, which stands for the database of Genotypes and Phenotypes. "Genotype" refers to the genetic information derived from a DNA sample and "Phenotype" refers to information about a person's health. dbGaP is intended to serve as a resource for researchers around the world, doing all kinds of studies including GWAS. Researchers who are doing Federally-funded genetics studies are being asked to submit their data, in de-identified form, to dbGaP. There are a number of protections in place to ensure that study participants' information is protected:

- De-identification of data stored in the database
- Access to detailed data is limited to researchers who have been thoroughly checked out by a special committee
  - Research purpose is valid and requires access
  - Type of research is consistent with terms of consent
  - Researchers are qualified
- Researchers promise not to try to identify individuals
- Penalties for individuals and institutions who break the rules

Do these protections seem appropriate/sufficient?

- Is anything missing?
- What else would you want to know?

Would you be willing to have your information submitted to dbGaP if it were protected in this way?

- Would knowing that your information would be sent to dbGaP affect your decision about whether to participate in a Group Health GWAS?

dbGaP is intended as a resource researchers can use forever – it will just keep growing as more and more information is added. What do you think of the fact that there's no end date for the research that could be done using dbGaP?

---

<sup>a</sup>The hypothetical example was established earlier in the session

**Table 3**

## Summary of Major Findings

---

|  |
|--|
| <ul style="list-style-type: none"><li>• Participants understood and generally supported the value of sharing de-identified genotypic and phenotypic study data, citing scientific advancement, research efficiency, and health benefits that would ultimately result from research.</li><li>• Older participants generally expressed stronger trust in the research enterprise and greater interest in serving the common good than did younger participants.</li><li>• While younger participants had a more sophisticated understanding of the potential benefits of high-throughput genomic research, they described greater privacy concerns and desire for control of study data.</li><li>• Surrogate decision-makers were willing to abide by ACT Study participants' preferences to allow broad sharing, though they themselves expressed some concerns about the practice.</li><li>• Participants in all groups expressed strong distrust of for-profit organizations as potential users of repository data, as well as concern about Federal oversight of such resources.</li></ul> |
|--|

---