Thought Leader Comparisons of Risks in Precision Medicine Research

by Laura M. Beskow, Catherine M. Hammack-Aviran, and Kathleen M. Brelsford

Table 1. Participant Characteristics (n = 60)		
	n	*(%)*
Perspective*		
ELSI research	6	(10.0)
Ethics	7	(11.7)
Federal government	7	(11.7)
Genome research	7	(11.7)
Health law	6	(10.0)
Historically disadvantaged populations	7	(11.7)
Human subjects protections	7	(11.7)
Informatics	6	(10.0)
Participant-centric approaches	7	(11.7)
Academic degrees		
MPH/MSPH	7	(11.7)
Other master's degree (e.g., MA, MS, or MBA)	23	(38.3)
JD, LLB/LLM	18	(30.0)
PhD	35	(58.3)
MD	16	(26.7)
RN	2	(3.3)
Based in		
United States	58	(96.7)
Other (Canada, United Kingdom)	2	(3.3)
Gender (self-reported)		
Female	31	(51.7)
Male	29	(48.3)
Race (self-reported)		
American Indian or Alaska Native	2	(3.3)
Asian	5	(8.3)
Black or African American	3	(5.0)
Native Hawaiian or other Pacific Islander	1	(1.7)
White	49	(81.7)
Ethnicity (self-reported)		
Hispanic or Latino	2	(3.3)

* The primary perspective for which we identified thought leaders is given; many could have been recognized in more than one category.

Box A.
The Hypothetical "Million American Study"

The Million American Study (MAS) is a federally funded, large-scale research endeavor to improve under-standing of health and to find new ways to predict, detect, diagnose, treat, and prevent disease. Specifically, the aim is to compile comprehensive information from a cohort of one million Americans in a repository that will serve as a rich research resource for a wide variety of studies for decades to come.

MAS will seek to enroll a representative sample of American adults reflecting diversity in terms of race and ethnicity, age, and sex. Those who agree to participate will give broad consent for

- extensive characterization (including whole genome sequencing) of biospecimens, such as blood;

- ongoing access to clinical data (such as medications, test results, and imaging) from electronic health records; and

- real-time monitoring of lifestyle and behavioral information, such as physical activity and environ-mental exposures, through mobile health devices.

At the time of consent, participants will be offered choices about whether they are willing to be re-contacted for various purposes (e.g., to provide additional information or specimens). Participants will be able to withdraw consent for future use of their specimens and data, with the exception that data generated in past studies cannot be withdrawn, nor can specimens and data be withdrawn from studies already begun.

Specimens will be stored in coded form in a repository at a major academic medical center in one state, while the data will be held at the coordinating center in another state. A robust data security framework will be in place, including administrative, technical, and physical safeguards. There will be a centralized governance process, comprising participant representatives, researchers, health care providers, government officials, and other stakeholders to ensure overall accountability and responsible project management.

Multiple tiers of access to MAS data—from open to controlled—based on data type, data use, and user quali-fications will be employed. For example, certain information, such as some aggregate results, will be publicly available. Access to other information will be available to qualified researchers from academic, nonprofit, and for-profit entities, in the United States and around the world, through application to a data access committee. For approved projects, data-use agreements will be used to ensure that data and specimens are used and shared for authorized purposes only and that privacy and security safeguards are maintained.

Information will be publicly available concerning how MAS cohort data and specimens are being used, includ-ing information about ongoing studies and summaries of research findings.

Adapted from F. S. Collins and H. Varmus, "A New Initiative on Precision Medicine," *New England Journal of Medicine* 372, no. 9 (2015): 793-95; M. J. Khoury and J. P. Evans, "A Public Health Perspective on a National Precision Medicine Cohort: Balancing Long-Term Knowl-edge Generation with Early Health Benefit," *Journal of the American Medical Association* 313, no. 21 (2015): 2117-18.

**Figure 1.
Riskiest Data and Level of Risk**
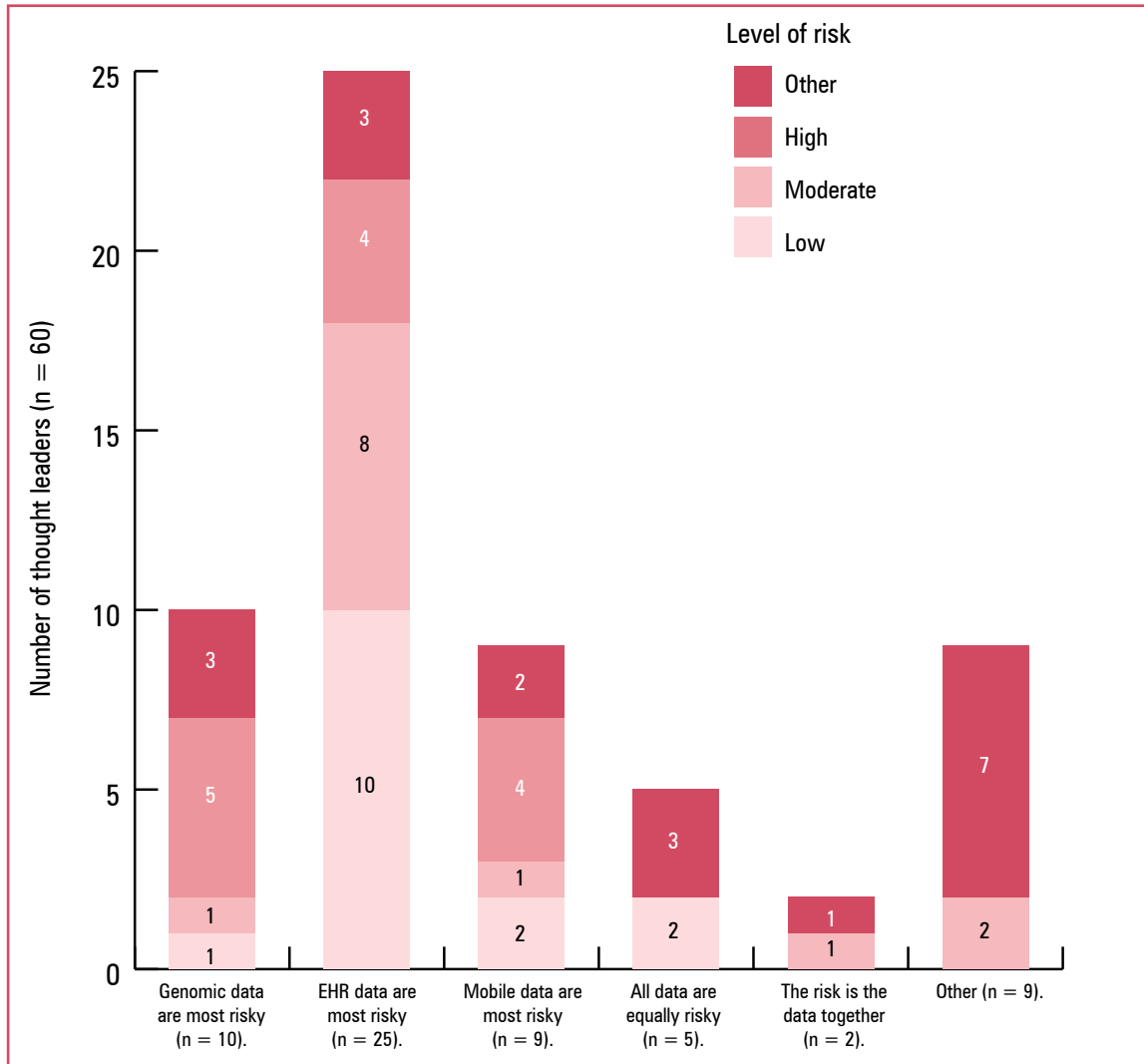
We asked the following:
- Which of these three aspects of the Million American Study do you think poses the most risk: genomic analyses of bio-specimens, ongoing access to EHRs, or streaming data from mobile health devices (and why)?
- You chose _____ as the most risky of the three things we asked you about. On a scale from 1 to 10, where 10 is the most risky, how risky would you say _____ is?

Rating responses are categorized in this way: 1-3 = low, 4-6 = moderate, and 7-10 = high.

Figure 2.
Overall Likelihood of Risks and Severity of Harms
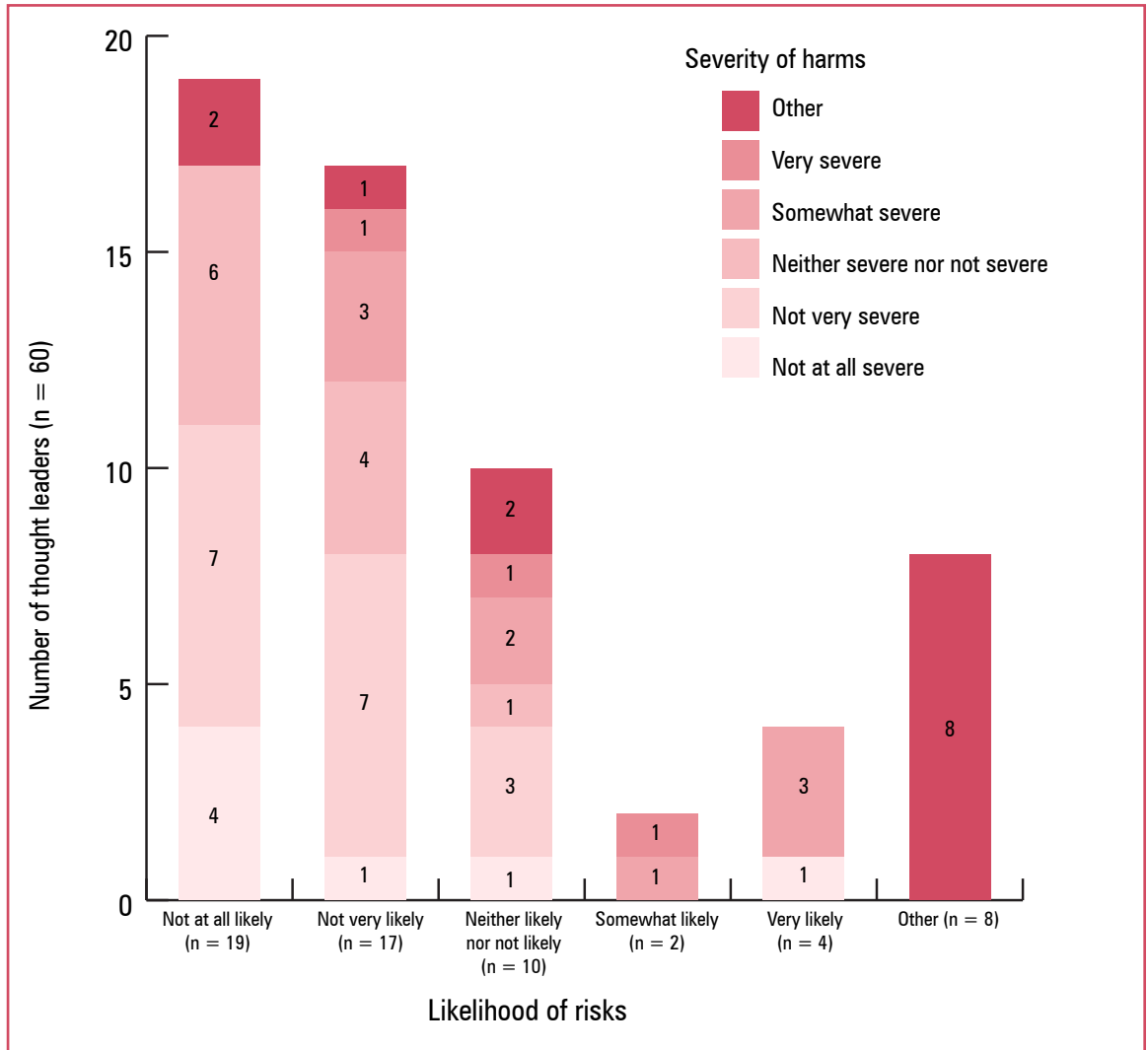
We asked the following:

Thinking about the risks and harms associated with participation in the Million American Study . . .

- how would you rate the likelihood of the risks involved actually occurring, on a scale from 1 to 5, where 1 is not at all likely and 5 is very likely?
- how would you rate the severity of the consequences if the risks actually occur, on a scale from 1 to 5, where 1 is not at all severe and 5 is very severe?

Appendix

Part A: Illustrative Quotations—Riskiest Aspect of Million American Study

Part B: Illustrative Quotations—Likelihood and Severity of Harm in Million American Study

A. Million American Study: Riskiest Aspect

1. EHR data are riskiest

Theme	Examples of illustrative quotations
Nature of data	
Sensitivity	There may be all sorts of things that you wouldn't think about that people regard as highly sensitive. Electronic health records are gonna capture all of that and continually update it. So that, to me, is the greatest risk. (12,* Law)
	There's more sensitive information in your EHR that you might not want other people to have access to. (15, Ethics)
	I think there could be greater sensitivity, even if only from a perception standpoint, about what someone currently has going on with their health, and what implications others might draw from that about them as a person or their future, etcetera. (21, Government)
	There are things within the EHR that I think people do consider to be more sensitive. You know, issues related to potential substance use, mental health issues. Where there is clearly stigmatization from a societal perspective. (58, Research)
"Actual" (versus probabilistic) meaning	I'd be more concerned about the electronic health record access because that's real information, clinical data that's about me. (49, Research)
	I would consider their health data a bigger risk than their genomic or lifestyle/behavioral risk. Because that contains all sorts of information about their actual health. (56, Research)
	I would say ongoing access to clinical data. The characterization of biospecimens, in and of itself, may not tell you anything about a medical condition . . . whereas clinical data is real. In the sense that you already have a disease—it's not just about genetic risk . . . . If you're getting information that somebody has a disease, particularly if it's a stigmatizing disease, to me the risk is much greater of that information. (59, Government)
Identifiability	
	It's probably the medical records piece that is the most at risk of being disclosed, and also relatively severe harms could occur. (37, ELSI)
	In terms of the greatest risk, [it] would be the ongoing access to clinical data. Putting in which medicines a person is on and what their test results are, all of that, just sort of the transfer of that information. If an evildoer can access that information then that seems to me to pose the greatest risk. (40, Human Subjects)
	The greatest source of risk is probably from the EHR data. There's the most amount of data in there. That is the most specific to who you are. (53, Informatics)
	Probably the one that has more concern or could be a potential area of risk is the EMR—ongoing access to that . . . . That's more directly linked to identifying information. (60, Historically Disadvantaged)
Unanticipated use	
General	Most people don't like sharing their health information. They find it highly personal . . . . The risk is in it being misused. So I guess there's two ways to ask the question.

Theme	Examples of illustrative quotations
	What's the potential the information is going to be misused? And, if it were misused, what bad things could happen? I still probably stick with electronic health record as the riskiest, posing very high levels of risk if misuse occurred . . . . If there's not protection against misuse, anything could be stigmatizing. (22, ELSI)
Discrimination	There are things that are probably in many of our medical records that we just wouldn't want to get out. So that would be the thing—whether it's embarrassment or stigma or information actually being used in ways that could harm us—that seems to me to be more likely to be potentially used harmfully than anything else. (16, Ethics)

You can infer things like sexually transmitted diseases, or really bad patterns in weight or depression or abuse of medication—that sort of thing. So that data is much more useable from a discrimination perspective in the short term than genetic information. (23, Participant-centric)

Ongoing access to clinical data—it really goes to issues about privacy to me. You know, yes, you may want to get information about my genome, but why do you need a continuation of information that really might be private? And that might be used in different manners, right? Clinical tests, medications, test results that lead to issues of discrimination. (46, Historically Disadvantaged)

I would say [your clinical data are] probably the most sensitive—the data that someone could most meaningfully make inferences about you and use against your interests. (54, Ethics) |
| **Level of risk** *(among interviewees who chose EHR data as riskiest, but rated level of risk as low or moderate)* | |
| Safeguards | While there are those that might be actively looking to capture health information for individuals to use it for their own purposes, there are many safeguards in place. And I don't perceive it to be the highest target on a list of . . . other ways that people might seek information [such as] financial information. (21, Government) |

2. Genomic data are riskiest

Theme	Examples of illustrative quotations
Unanticipated use	
Law enforcement	Biospecimens can be used for things like—or at least they're perceived to be facilitating activities such as—forensic investigations or the ability to link people to crimes in a way that clinical data might not be. (31, ELSI)
Discrimination	Whole genome sequencing and characterization, in my mind, has a higher risk than the ongoing access to clinical data. Just because you might find something there that's new or unique that affects insurability. (39, Human Subjects)
	You have no idea what's going to happen with your data, who's going to have access to it, what it's going to be used for, whether it could be used to discriminate against you, keep you from getting life insurance. Pharmaceutical companies could find out your disease, bug you to death. There are all kinds of risks—your employer could find out. I know we have GINA, but still. If the data are not controlled and you don't know how they're being used or . . . what they're going to be used for, I think it's a huge risk. (51, Informatics)
Objectionable research	We've not got a good history as a society or a species of dealing with genetic information very well. And especially from certain ethnic groups, that information has had pretty negative consequences. (5, Research)
	There're a lot of reasons why scientists keep on pushing forward with discovery. They don't rein themselves in on the discovery end, including starting biotech firms and making a lot of money. (47, Historically Disadvantaged)
	The risk can be ameliorated, but without further knowledge, I would have to say it's a huge risk for you to give over your genomic data without knowing how they're going to be protected, how they're going to be used, who's going to be using them. (51, Informatics)
Nature of data	
Identifiability	The whole genome sequence is widely viewed as the most accurate identifier of an individual there is. With today's big data analytics, you potentially could derive a person's identity from their clinical data, as well. But if a person's genome were actually stolen, that would be very hard to get around. (51, Informatics)
Familial implications	They all pose significant risks, but they're somewhat different risks. For example, the one with the biggest family risk component is the genomic data because that's what you share most with your biological family, with your genetic family. (14, Law)
	Different participants will have different concerns about their genome sequencing. If they have a condition that they know has a genetic risk, they may be concerned about their family members. Now, obviously they might have some concerns about themselves. But the larger concern for most people that I know is, "I can deal with this, but I have kids or I'm going to have kids, and how is this going to impact them?" (55, Government)
Return of results	
	Genomic testing, again, if it involves returning results or anything of probabilistic interpretation, is the most risky for an individual. (3, Informatics)

Theme	Examples of illustrative quotations
	In today's world of sort of return of research results, I think of the possibility being quite high that the information will come funneling back to people, but not in such a way that will be ultimately well taken care of. (5, Research)
Genetic exceptionalism *(among interviewees who chose another data type as riskiest)*	
	I actually think people might perceive the first one, the sequencing of the tissue, as being the most risky because there's been so much hype around genetics . . . . People have described a special weight to genetic information even though I think that's genetic exceptionalism. Most of the information is of little value or not even getting interpreted. (28, Law)

3. Mobile data are riskiest

Theme	Examples of illustrative quotations
Nature of data	
Volume, granularity	Medical records usually pertain to only information that's collected in the clinical setting. But [streaming data are] going to contain way more information than what you generally have in the current clinical setting. We would be talking about lifestyle information that might be coming from mobile devices, we'd be talking about exposure information that isn't necessarily documented in the medical record. We're basically talking about substantially more information than what you would find in a medical record. (13, Informatics)
Where, when, what	Mobile phones tell you quite a lot, where the person has been, what he did, which hours . . . . So it gives much more rich information and content information by looking at someone's cell phone rather than to look at someone's DNA. (7, Research)
	We are feeding information to people about what you're doing at any point in time. (8, ELSI)
	There's also some geolocation information being gathered, because you're talking about environmental exposures. If there's some sort of tracking of your movements and activities, well, where do we go and what do we do when we get there? People might actually find it more concerning to have the world knowing that than they would that they have a 20 percent chance of some obscure disease. (26, Human Subjects)
Identifiability, security	
	I think it's riskier than anything else, because you're relying on third-party hardware to send out my data. This is either my phone or my Fitbit or my Microsoft band or my Android watch or whatever you're using to collect my health data. That requires me to encrypt it and to encrypt it properly. That, I think, is the riskiest thing. That also tells people that I'm sleeping or I'm awake, or I'm running around . . . . That's a lot of information that's very intrusive. (48, Law)
Unanticipated use	
Law enforcement	It really tells you at any one time where somebody is. Their geographic coordinates are available because of that and that's on a database somewhere and that in itself

Theme	Examples of illustrative quotations
	can be used for law enforcement purposes, so I think that's a problem. (29, Human Subjects)
Government	I'll tell you a whole lot more about yourself with your ZIP code than with your entire genomic code . . . . I'm a big government kind of guy in my personal politics—but we're not used to thinking the government is going to know where we are and this is a government study. (44, Research)
Malicious actors	I think it does raise some issues, especially if people are thinking that all this information is just helping them maintain their exercise regime or pay attention to their diet. There are all sorts of other uses that folks may not be fully thinking of when they sign on to these things. Eventually I think we adjust to that. But we usually adjust because people do really stupid nasty things and we pass a law saying you can't do that anymore. (52, ELSI)

4. Other responses

Theme	Examples of illustrative quotations
Depends on the protections	If it were me, I would want to know how robust the data controls are on the back end. (18, Law)
Depends on the individual	It depends on the individual's situation and what's in their genome. Because a lot of us, so far, have very boring genomes. Which is in general a good thing. If you don't have a boring genome, that's a bad thing—so there are risks to that if that's the situation you happen to be in. (20, Research)
	It really would depend on the individual and their health profile. (24, Government)
	It would really depend on the individual's personal concerns. (36, Historically Disadvantaged)
Risks are different	I think they're all core. They're all sort of built on each other and they have their own unique risk. It's hard to say which one is the most. That's probably specific to what is the risk we're talking about. (17, Participant-Centric)
Combination of data types together	For me, the combination is much more powerful which is of course why it is being proposed at all. . . . I really think it's the combination that matters. . . . That's the fullest picture of that person, so if you're worried about people knowing what you're doing and what's going on in your body and physiologically . . . and what's going on at the genetic level, you're learning a whole lot more by the combining of the three. (11, ELSI)
Risks are inseparable	I would have a hard time separating those things because I think technology is such that genomic data will be appended to electronic health records and it will be streamed on mobile devices. I can't really partition those risks from one another. I think all of these data—electronic health records, phenotype data, genotype data— they are all very portable and very easy to make copies of. So I see it really is kind of one big bolus of data. (43, Participant-centric)

B. Million American Study: Likelihood of Risks, Severity of Harms

1. Likelihood of risks

Theme	Examples of illustrative quotations
Not at all/not very likely	
Track record of safety	You have thousands and thousands of people participating in these studies over the years—millions I should say. If you add it all up, the number of people that have experienced an actual risk playing out, an actual breach playing out, is relatively small. I'm involved with a lot of cohort studies, and every single one of them has had some kind of little issue that has emerged. An information issue that has emerged—but they're usually small, and they're rare. (28, Law)
Protections	I'm making some assumptions about this being done in HIPAA-covered entities, that there would be strong privacy and confidentiality, data security protections, that there would be strong terms in an MTA about reidentification and so on. So, under those conditions, I gave it a 2. (59, Government)
Somewhat/very likely	
Unintended/unwanted use	Just the same way that you can have the issues with direct-to-consumer genetic testing, you could have issues with large databases being used in different ways, right? Let me give you an example. There was an issue a while ago where women going and having babies, of course they do the blood spot for the child to check for certain diseases. That information was being used to populate the mitochondrial DNA database, which they're supposed to use for fallen soldiers, but now it's changed to talk about terrorism and they were using it for other things. My question is, if we can do that for a little small thing in terms of a blood spot where we don't get consent, who's to say once we develop this information that it won't be used for things that are not thought of? (46, Historically Disadvantaged)
It depends	
Trustworthiness	I'm thinking about: In whose hands is the information? Who's holding it? Where's the responsibility for maintaining and protecting it? Because I think we've seen that the federal government is not so able to do that, but it's pretty rare for [prominent universities] to have a breach. (11, ELSI)
Protections	It's hard to know a priori because we haven't done this kind of thing at this scale. You could imagine a lack of infrastructure or oversight or something leading to a breach. (43, Participant-centric)
Return of results	If you're somebody who says, "I don't want lots of information back," it's going to be a 1. If you're somebody who says, "I want a lot of information back," then it's going to be higher than that, even as high as a 5 if you basically said, "I want a very broad scope of information back," and the investigators make that possible. So tough to answer because it depends on choices you make. (16, Ethics) Of course, there could be a little bit of a difference in the risk based on what the provisions for returning findings are and how they're worded and what they might give you. But still I think it would be lower [risk] if it's well protected and not clinical. (39, Human Subjects)

2. Severity of Harms

Theme	Examples of illustrative quotations
Historically disadvantaged perspectives	Pretty high—and I'm thinking of Alaska Native/American Indian people here specifically—I would say 4 or 5. (30, Historically Disadvantaged)
	All of the risks I mentioned are the ones I think should be on the table and we should figure out how to have intense conversations of how to minimize these risks that effect certain populations . . . . There is a thin line between identifying concerns of native people and stereotyping them . . . . The only way to do this ethically is to engage in the process early on with researchers who have expertise in dealing with the ethical issues that native people are facing. (36, Historically Disadvantaged)
Return of results	I think that you're gonna have too many people just monitoring themselves, driving themselves crazy. (48, Law)
Consequences for research endeavour	If it's a risk to the individual, that means the impact it can have on the project is huge—it's that public trust stuff, right? (28, Law)
It depends	I suppose it would depend on how sensitive the information is, objectively as well as subjectively, right? People may view certain things in a more sensitive way . . . that would depend on individuals. (12, Law)
	The magnitude of harm, should that occur—that is a lot harder to judge. Because it's so dependent on exactly what's collected and what your personal circumstances are, right? It just depends. (19, Ethics)
	It depends on what the information is—you know, is it Huntington's disease or something like that? Or is it the fact that you have blue eyes and you might have a propensity towards obesity theoretically? (39, Human Subjects)
	And severity? I honestly can't give a number because that depends on the individual person: like how—if you're entering your career and something deleterious to your future health is made public, that could hurt your career. But maybe you're a retiree and it doesn't matter . . . . So it's so dependent on individual circumstances. It's just so hard. (54, Ethics)
	If we're talking about some relatively benign condition . . . it depends. And then whether you're talking about something like alcoholism or schizophrenia or something like that, it could be much higher. For the average kinds of things, I would say it would be at a 2 to a 3, maybe more towards a 2. Could be as high as a 5, again, depending on what information is being shared. (59, Government)

*Throughout the appendix, numbers in parentheses refer to participant IDs.