NATIONAL INSTITUTES OF HEALTH

Using Public Engagement to Inform the Use of Data in Biomedical Research
Workshop at University of California, San Diego
February 27, 2023 (All Times PST)

Community Conversation Summary of Feedback

Background Information:

The following sections include a summary of facilitated discussions around case studies from each of the community conversations. These descriptions are presented in two forms: 1) summary and main themes by research topic area and 2) bulleted points by case study. These summaries were derived from notes captured by science writers present for each of the events. Quotes from participants and other potentially identifying information have been removed to protect participant privacy. Informal survey results, ice-breaker questions, and other information are included where applicable.

The WG designed case studies and discussion questions based on the types of research questions that they defined in Phase I and to be accessible to lay audiences. After each conversation, the case studies and discussion questions were reviewed by staff and facilitators and some were adjusted slightly to elicit feedback most responsive to the NExTRAC WG's prioritized questions. Additionally, to cover different elements of the research topic, two different case studies were developed for Topic Area #1 and #2. Full versions of the case studies are embedded in the summary info below for context and the table below notes specific case studies used to facilitate conversations in different communities.

Location	Date	**Topic Area #1:** Novel data types used outside of the traditional healthcare system	**Topic Area #2:** New Ways of Analyzing Health Data	**Topic Area #3:** Linking and Aggregation of different Types of health data
Harlem, NY (Northeast USA)	11/9 2022	Case Study 1: Wearable, app, social media data to monitor heart health	Case Study 1: Algorithm to detect lung cancer (imaging data)	Linkage of data from a dementia study with COVID data
Bronx, NY (Northeast USA, en Español)	11/10202 2	Case Study 1: Wearable, app, social media data to monitor heart health	Case Study 2: Algorithm on social media to identify signs of suicide in teens	Linkage of data from a dementia study with COVID data
Alamosa, CO (Southwest USA)	12/5 2022	Case Study 1: Wearable, app, social media data to monitor heart health	Case Study 1: Algorithm to detect lung cancer (imaging data)	Linkage of data from a dementia study with COVID data
Jackson, MS (Southeast USA)	12/14 2022	Case Study 2: In-home sensors, smart devices to monitor an older relative's health	Case Study 2: Algorithm on social media to identify signs of suicide in teens	Linkage of data from a dementia study with COVID data
Dearborn, MI (Midwest, USA)	1/18 2023	Case Study 2: In-home sensors, smart devices to monitor an older relative's health	Case Study 2: Algorithm on social media to identify signs of suicide in teens	Linkage of data from a dementia study with COVID data
Webinar (Rural Health Advocates, Rare Disease Advocates)	1/24 2023	Case Study 2: In-home sensors, smart devices to monitor an older relative's health	Case Study 1: Algorithm to detect lung cancer (imaging data)	Linkage of data from a dementia study with COVID data
Santa Clara, CA (West, USA)	1/31 2023	Case Study 1: Wearable, app, social media data to monitor heart health	Case Study 1: Algorithm to detect lung cancer (imaging data)	Linkage of data from a dementia study with COVID data
Webinars (American Indian, Alaska Native Advocates)*	2/16, 2/17 2023	Case Study 1: Wearable, app, social media data to monitor heart health	Case Study 2: Algorithm on social media to identify signs of suicide in teens	Linkage of data from a dementia study with COVID data

* Science writer summaries of these community conversations were not completed at the time of preparation of this packet, though some feedback was summarized by NIH attendees.

Overall Themes

The following themes were observed across all topic areas/case studies:

- Participants' support for data sharing efforts was dependent on who was receiving the data, whether they were considered trustworthy, and whether there would be a significant return to the public in terms of health benefits.
- Participants expressed a desire to know more about how their data are used downstream from their initial uses (whether collected first for research or for other purposes) and to have more control over those uses.
- While participants expressed some enthusiasm about novel uses of data across the three topic areas, they generally wanted greater transparency in what these uses would entail and reassurances that the benefits would in fact be worth the potential risks.

Topic Area #1: Integrating public feedback into biomedical research and technology involving novel data types used outside of the traditional healthcare system (Wearable technology, Social Media, etc.)

Types of research questions for Topic Area #1 defined in Phase I

- How are personal health data collected from outside of the traditional health system (wearables, fitness trackers, apps (e.g., period tracker apps), social media posts) being used to study health-related questions and predict health risks, at either an individual, family, group, or public health level?
- How can other consumer and lifestyle data from non-health-specific sources (e.g., credit card and consumer rewards data, sensors in the home) be used to study health-related questions and predict health risks?
- Can/how can integration of health data with data on social determinants of health (SDOH) enable better risk stratification of patient populations and better development of predictive algorithms? SDOH include socio-economic status, housing status, education status, geographical environments in which people spend time (e.g., crime rates, environmental pollutants in a neighborhood), & identity factors that advantage or disadvantage health status.

Summary of community conversation feedback on Topic Area #1

- In general, participants supported sharing data to improve their own health and the health of individuals in their community, including for research purposes.
- In general, participants wanted reassurances that their privacy would be protected when sharing personalized, identifiable data in research, especially from social media. If data could be protected against misuse, participants were enthusiastic about the potential to contribute to research and help address health issues in their communities.
- Support for sharing data was dependent on who would receive the data and whether they were trustworthy, as well as whether there were reassurances for individuals' and communities' privacy.

Main themes of public feedback on Topic Area #1

- **Trust**: Participants placed high importance on whether individuals or organizations receiving and using data could be trusted. Companies or researchers making money from sharing data were sometimes viewed as untrustworthy.
- **Transparency:** In general, participants wanted greater transparency about where data are going, how data are being used, and the goals of research using their data.
- **Privacy:** Participants wanted greater reassurances that their privacy would be protected, especially for data used for secondary purposes that could be used against individuals. For example, participants identified the possibility that data could be shared from wearables or use of in-home sensors in ways that could affect insurance, reproductive health, immigration status, or generally could cause stigma for certain groups.
- **Control over downstream uses:** Participants generally wanted more control over future uses of their data, especially when data would be identifiable or unprotected. For example, if a study utilized in-home sensors, data from those not participating in the study could be collected and used for other purposes with few options for limiting uses.

- **Return to individuals and communities:** Participants wanted more of a return in terms of health benefits to participants and communities. For example, there was strong interest in sharing research information with health care providers for individuals. Some participants were also skeptical that social media data would be sufficiently accurate that it could translate to health benefits for individuals.
- **Challenges with informed consent:** There was slight concern with researchers obtaining personal health data from public and/or commercial sources (e.g., social media data) without asking individuals for consent to do so, as well as the practice of obtaining data when individuals perceive they have very little choice over whether to consent to data sharing (e.g. rewards cards, consumer data). Many participants seemed aware that these data are often owned by developers or were in the public domain AND preferred consent or at least notification to use this data in research.

Public feedback on Topic Area #1: main points by case study

Case Study 1: *Use of a wearable, smartphone app, and social media data to monitor heart health (Harlem, Bronx, Alamosa, Santa Clara, AI/AN advocates webinars)*

You have been asked to participate in a research study about heart disease. You are in good health, but heart disease runs in your family. The research study is trying to identify the signs of when or how things start to go wrong. The researchers hope that by tracking several things that you do every day, they can identify "healthy" patterns that people at risk of heart disease should follow and "unhealthy" activities to avoid.

As part of the study, the researchers loan you a watch that can measure your heart rate and blood pressure. They also ask you to download an app on your phone. The app collects information from the watch, as well as other information from your phone (such as how active you were throughout the day). They also ask you to keep a record of the food that you eat and enter that information into the app. The app automatically shares all of this information with the researchers. All you have to do is keep your watch on your wrist and your phone on or near your body (such as in your pocket or a bag that you carry), and the researchers will study the data that have been collected.

To get an even clearer picture of how you're doing, the researchers also ask for permission to monitor the information you post on your social media accounts (if you are active on social media). The researchers hope that by looking at information from all of these places, they will better understand early signs of heart disease.

Case Study 1 Discussion Questions

1. *Are there particular personal health or other (e.g., social media) data that you would be more comfortable sharing with researchers than others? Which data, and why?*
2. *What do you think is the best way for researchers to make use of this data after the study? Who should have permission to access the data? What should people be allowed to do with the data? Who should decide who can access the data?*
3. *Would you let your children participate in this study? Would you have any concerns about children sharing any of the specific types of data that the researchers are looking to collect?*
4. *In this case, researchers have outlined for participants the data that they would collect, and they have asked your permission to collect it. However, some of this data may be publicly available or collected and made available for purchase by a company that makes one of the devices or apps or by a "data aggregator" that compiles and packages information from multiple sources. How would you feel about researchers using this publicly available information or purchasing information to create a picture about your health without asking for your permission?*

Case Study 1: Main points from community participant responses

- Though there were some exceptions, participants generally expressed support for sharing data to improve their own health and the health of individuals in their community.
- There was general support for sharing data to help advance COVID-related research.
- In general, participants were concerned about the use of personalized data in research, especially from social media.

- Support for sharing data was dependent on who would receive the data and whether they were considered trustworthy (e.g., companies or researchers making money from sharing data were viewed as untrustworthy).
- Participants wanted greater transparency about where data are going, how data are being used, and the goals of the research.
- Some participants, particularly younger individuals, were less concerned because so much data about them has already been shared; the data might as well be used for research. But these participants still wanted greater transparency about how data are used.
- The biggest concerns were for privacy and the potential for data to be used against individuals and communities (e.g., if downstream sharing could affect insurance, reproductive health, immigration status, or generally could cause stigma for certain groups).
- There was strong interest in sharing research information with health care providers. In general, participants expected more of a return to communities in terms of health benefits. They were largely skeptical that social media data would be sufficiently accurate and would translate to health benefits for individuals.
- There was slight concern with the practice of researchers being able to get public data without consent. Participants preferred consent or at least notification to use this data in research.
- There was general skepticism toward the inclusion of children in such studies.

Case Study 2: *Use of in-home sensors and smart devices to monitor an older relative's health (Jackson, Dearborn, Rural health and rare disease advocate webinar)*

Over the holidays last year your friends gave you a gift—a set of "smart" speakers, similar to a Google Echo and Amazon Alexa. The speakers not only play music and the radio but also listen to you when you talk out loud. The speakers are connected to the internet and have a virtual assistant who you can talk to, ask questions, and ask to control other devices in your home.

Recently, an elderly family member has moved into your home. They are having some difficulties remembering things clearly and also have some trouble maintaining their balance, which makes you nervous about leaving them alone. You often go to doctors' visits with your relative, and at one such visit, their doctor approaches you and your relative about a study helping researchers at the local hospital increase the safety of elderly individuals in the home.

The researchers offer to outfit your home with a network of sensors that connect to the smart speakers. For example, currently the Google Nest can be set to automatically turn the temperature of your home up or down based on your preferences. In this potential study, some of the sensors the researchers bring could detect the temperature and tell the smart speakers to turn on your home's heating and fans when it gets to a certain temperature inside. There are other sensors that can be used to monitor health conditions, such as a sleep tracker mattress, a microchip in the toilet that detects elevated lab markers in urine, a smart scale that measures weight loss, and a digital mirror that notices changes in appearance. Sensors can also analyze movement patterns to detect whether someone might be injured if they fell or were confused. However, the researchers note that none of these functions are foolproof and that the technology sometimes makes mistakes.

The researchers want to collect data on your relative's speech, movement, and vital signs using an app that will help them identify potential causes of safety issues. The speakers that are already in your home can be connected to the sensors to constantly collect data on you and your family members and send the data back to the company and researchers for analysis. This is how the researchers would monitor your relative's health status. For example, the data might be able to tell the difference between when your relative is having a stroke or just has low blood sugar. The company and researchers claim to keep the data private and protected, but a lot of personal data is shared with the company through the app and the sensors.

Case Study 2 Discussion Questions:

1. *How do you feel about the company getting your and your relative's data? How would you feel if other researchers were able to get this data from the company to help study other conditions?*
2. *Suppose you live with other family members and frequently have family and friends over to your home. Many sensors can't distinguish between different people and just collect data on everyone in the home at any given time. How would this influence (affect) your decision to enroll in the study?*
3. *What concerns would you have (if any)?*
4. *Are there factors that would increase your likelihood of participating?*
5. *How important is it to you to understand the details of your data and how they were analyzed? Do you want/need to understand all of the details? Who would you prefer to help you in analyzing (understanding/interpreting) your data (e.g. your physician, a researcher, or someone else)?*

Case Study 2: Main points from community participant responses

- Participants were enthusiastic about the potential benefits of in-home sensors for health but were concerned about downstream uses of the data.
- Support was dependent on receiving reassurances about privacy, accuracy of any conclusions made, and how the data will be used (e.g., participants were largely opposed to selling data to companies, even if for future research).
- Participants were aware of possible risks with incidental collection of data on others in the home and wondered how that could possibly be controlled for or prevented.
- Participants wanted more of a return in terms of health benefits to participants.
- Some community members felt that their data had not been used very well in the past, and consequently they felt burnt out by the amount of requests they received that didn't seem to go anywhere.
- There were concerns about whether companies who develop the technologies or the researchers would be liable in the case of faulty data.
- It was important to many participants that the data and any results would be explained in an accessible way.
- People would be more likely to participate if trusted partners to the community were involved.

Topic Area #2: Integrating public feedback into biomedical research and technology involving new types of analyses (e.g. Artificial Intelligence, Machine Learning)

Types of research questions for Topic Area #2 defined in Phase I

- What is the role of computer-based technologies, such as artificial intelligence (AI), machine learning (ML), and automated image analysis, in advancing health decision-making?
- Can/how can natural language processing be deployed to analyze data held in health systems (e.g., Electronic Health Record (EHR), health insurance data, data from pharmacies) to provide insights about patient symptoms and disease classification?

Summary of public feedback on Topic Area #2

- Support for AI was dependent on the reliability of algorithms over time. If algorithms could demonstrate reliability, participants were excited about the idea of sharing their data in order to receive more accurate diagnoses, better care, and to address public health issues.
- Participants largely supported the use of AI in clinical contexts but trusted the decisions of human doctors more, based on longstanding relationships and existing trust.
- Participants were aware of the possibility of bias and misuse of algorithms. They accordingly wanted to know more about the possibility of stigma and how else the data and algorithms might be used.

Main themes of public feedback on Topic Area #2

- **Trust:** While open to the idea of trusting AI in clinical contexts, many participants ultimately wanted to hear from their own trusted doctor to provide context and explain AI decision-making.
- **Transparency:** When receiving diagnoses from an algorithm, participants said that they would want greater transparency around how the algorithm was developed, who developed it, how it comes to conclusions, and which populations were studied during development. Many participants wanted to hear from both researchers and health care providers to understand more.
- **Accuracy and reliability:** Participants wondered how inaccurate conclusions made by AI might be corrected, whether those conclusions could be harmful, and how those conclusions and any associated data would be used in the future. For example, many participants were enthusiastic about the potential for using algorithms on social media to help address mental health issues but were skeptical that this would translate to reliable interventions.
- **Return to individuals and communities:** Many participants strongly wanted the research to be translated into benefits for the people involved in the research and their communities. They also wanted greater involvement from communities (including children as appropriate) in deciding how data would be used. There was some concern that this sort of research could cause stigma to individuals and communities.
- **Privacy risks require significant health benefits:** Many participants perceived a potential loss of privacy in the use of AI but expressed that the tradeoff would be worth it if the health benefits were significant. For example, an early cancer diagnosis may be worth it.
- **Control over downstream uses:** Participants generally wanted more control over data collected and used in AI research, especially data that may be completely unrelated to the main topic of a study (e.g., social media data). Some participants indicated that they would be happy to receive potential personal health information that AI might have discovered beyond the scope of the research for early detection, but that the possibility that this could happen should be shared in advance and individuals should have more options to control how the data are used.

Public feedback on Topic Area #2: main points by case study

Case Study 1: *Use of an algorithm on imaging data to detect early-stage lung cancer (Harlem, Alamosa, Rural health and rare disease advocates webinar, Santa Clara)*

Lung cancer is a complex disease. There are people who smoke daily throughout their adult lives and remain cancer-free. Others live their entire lives in clean-air environments, avoid smoking, and have no family history with lung disease but for some reason develop lung cancer. There's still a lot to learn, and researchers continue to search for early risk factors to prevent people from getting lung cancer or to diagnose that cancer at an early stage to enable early treatment.

At your yearly checkup with your health care provider, you see a flier seeking healthy volunteers in your age group to participate in a research study looking for early risk factors of lung disease. It's a minimal time commitment. You have a family history of lung cancer, and while your affected relatives were heavy smokers and you've never smoked, you decide to participate in the study.

When you talk to the recruiter for the study, they inform you that the researchers are developing a way to compare the medical records of healthy people with the medical records of people with lung cancer, using artificial intelligence (AI). As part of the study, you will undergo a screening procedure (computerized tomography, or CT, scan). Researchers will compare your scan to scans of both healthy patients who never got cancer and patients with a variety of different types and stages of lung cancer. The researchers inform you that you have the option to receive your results from the study, and they will schedule an appointment with a lung doctor for you to discuss these at your convenience.

A few months after enrolling in the study, you see the lung doctor, who gives you the following results:

- *There are a few factors in your medical records that are considered risk factors. While you haven't smoked, your parents did in your home throughout your childhood. Additionally, it was discovered that you work and have lived near some industrial factories that release chemicals that may cause cancer.*
- *The doctor informs you that the scans were analyzed with an AI algorithm that was specifically designed to detect early-stage lung cancer. However, the doctor cannot tell you much more about how the algorithm works. The doctor only knows the outcome: no tumor detected. The doctor has no further concerns at this time. However, given the risk factors discovered, they recommend more frequent screenings (yearly as opposed to every 5 years).*

Case Study 1 Discussion Questions

1. *Algorithms (i.e., computerized calculations) consider a lot of information assisting your doctor. Are you more or less likely to trust your doctor's diagnoses and guidance than that of an AI-based tool? What else, if anything, would you want to know about the researchers' process of scanning social media posts? Are there factors that would increase or decrease your trust in their use?*
2. *Who would you want to hear from to help you understand this sort of predictive or diagnostic health information?*
3. *What, if anything, would you want to know about the algorithms being used?*
4. *Are there factors that would increase or decrease your trust in their use?*
5. *Do you have any concerns about privacy?*
6. *How do you balance the benefits of knowing this health information with any concerns about privacy that you may have?*

Case Study 1: Main points from community participant responses

- Participants largely supported the use of AI but trusted the decisions of human doctors more, based on longstanding relationships and existing trust.
- Support for AI was dependent on the reliability of algorithms over time. If algorithms could demonstrate reliability, participants were excited about the idea of sharing their data to receive more accurate diagnoses and better care.
- While open to the idea of trusting AI, participants ultimately wanted to hear from doctors to provide context and explain AI decision-making.
- When receiving diagnoses from an algorithm, participants said that they would want greater transparency around how the algorithm was developed, who developed it, how it comes to conclusions, and which populations were studied during development. Participants wanted to hear from both researchers and health care providers to understand more.
- Participants were aware of the possibility of bias and wanted to know more about how the data and algorithms might be misused.
- There were concerns about privacy but attendees expressed that the tradeoff would be worth it if the health benefits were significant (e.g. an early cancer diagnosis would be worth it).
- Some community members indicated that they would be more supportive of AI if it enabled easier access to healthcare.
- Some might be more familiar with the use of AI from social media and pop culture generally, so they might be likely to misunderstand the use of AI for health purposes.
- There are unique complexities of AI and rare diseases, such as the algorithms' ability to distinguish specific subtypes of a rare disease. So, some individuals from those communities might want to hear about the experiences of patients who received an early, accurate diagnosis from an algorithm. There was also a concern that a patient could be diagnosed with a condition that had no treatment options.

Case Study 2: *Use of an algorithm on social media data to identify signs of suicide in teens (Bronx, Jackson, Dearborn, AI/AN webinars)*

Suppose you live in a town where there has been a recent spike in teen deaths. The pandemic has been hard on everyone—people have been stuck in their homes, the economy has taken a big hit, and the pressure on middle and high school students with virtual school is higher than ever. Because these teens haven't been able to interact with each other in person for some time, they're mostly expressing themselves and connecting with each other through social media apps (e.g., Instagram, TikTok, Facebook). Your neighbors have all kinds of theories about what might be causing so many deaths among younger individuals, but it's hard to know for sure. Regardless, everybody agrees that it's a problem and something needs to be done.

Researchers have developed a way to rapidly scan and interpret social media posts, including video, images, sound, and text. Public health officials in your town have contacted these researchers to determine whether this resource could be used to get at the underlying causes of teen deaths, identify signs of suicidality in teens, and develop ways to solve this problem. The researchers agree to test their process in your community and will scan social media posts and use the data to obtain information such as median (average) income and education level of the neighborhood and crime rates.

Your public officials have organized a town hall with these researchers to engage with you and your neighbors so that you can discuss the pros and cons of using this technology.

Case Study 2 discussion questions:

1. *What is your first question for the researchers? For the public health officials?*
2. *What else, if anything, would you want to know about the researchers' process of scanning social media posts? Are there factors that would increase or decrease your trust in their use?*
3. *If you agree to work with the researchers, they would like to publish their results and share the data for reuse in future studies or by other researchers. Typically, any identifying information would be removed from the data before publication or sharing, though this can limit the ability to use the data to help people. Is it enough to remove identifying information? Are there other conditions that you feel would be necessary before you would consent to publication and sharing of these data about your and your neighbors' children?*
4. *Would you feel differently about your response if we were talking about adults and not kids?*

Case Study 2: Main points from community participant responses

- Many participants were enthusiastic about the potential for such algorithms and supported using social media data to help address mental health issues, but they were skeptical they would actually translate to reliable interventions.
- Participants wondered how inaccurate information might be corrected, whether data could be used for harmful purposes, and how data would be used downstream. Some participants suggested that it would be helpful to see evidence of the AI's efficacy.
- There were questions about how the AI itself worked, including the criteria it measured and how it could capture the complexity within a social media post.
- Participants strongly wanted the research to be translated into benefits for the people involved in the research and their communities. They also wanted greater involvement from communities (including teens) in deciding how data would be used. There was some concern that this sort of research could cause stigma to individuals and communities.
- Participants were generally skeptical and somewhat concerned about the inclusion of children in studies analyzing social media.
- Some participants were skeptical that such data could ever be fully de-identified and so were concerned about risks of reidentification.
- There was support for having results returned to research participants and their communities.

- There was concern about how incidental data collection, unrelated to the research, would be treated and potentially used downstream.
- Participants were strongly supportive of getting consent for this type of research.

Topic Area #3: Integrating public feedback into biomedical research and technology involving data linkage and aggregation

Types of research questions for Topic Area #3 defined in Phase I

- Are there opportunities to standardize data formats – or deploy standardizing technologies – so that data from different countries and healthcare systems could be aggregated, linked, and shared across populations?
- Which disparate (and potentially conflicting) data sets (e.g., genomics, proteomics, clinical information, clinical imaging) can be linked and combined with harmonized (automated) data aggregators?
- Can/how can personal health libraries be used to combine individuals' health information across multiple different data streams to inform health outcomes?
- How can Privacy Preserving Record Linkage (PPRL; a method for integrating data while maintaining the security of privacy information) be used to combine data on individuals from multiple sources and with different identifiers for precision medicine and public health?
- How should the research context (e.g., clinical, public health) and participants' consent status affect data linkage and aggregation?

Summary of public feedback on Topic Area #3

- In general, participants were somewhat skeptical about the process of data linkage itself but supported linkage to help advance research. Support was dependent on having greater transparency about how data might be used, getting more detailed information during the initial informed consent process, and receiving notification or potentially reconsent prior to future linkage.
- Participants tended to support the use of linkage services specifically to facilitate research recruitment.
- Many participants were wary of the commercial aspect of linkage and allowing companies to have their data.

Main themes of public feedback on Topic Area #3

- **Challenges with informed consent:** Those who supported linkage tended to endorse the idea that initial broad consent gave permission to conduct data linkage, especially for important research. However, participants were generally skeptical of "check the box" consent procedures that do not describe future possible uses in detail, including the possibility of linkage. Some participants noted frustration with feeling as if they were acquiescing as opposed to consenting in many contexts.
- **Transparency**: While some of the benefits of linkage were clear to participants, particularly for addressing issues like COVID-19, many participants perceived the process as lacking transparency and posing too many risks (e.g. re-identification, data breaches), given the amount of data available for linkage and who else might receive the data downstream from research.
- **Privacy and identifiability:** In general, participants were more willing to support linkage between completely anonymized and aggregated datasets. Linking more personalized data, such as from drugstore rewards cards, was more likely to be viewed as inappropriate, though participants recognized that they may have consented to this when signing up (see above note on acquiescing vs. consenting).

Public feedback on Topic Area #3: main points by case study

Case study: *Linkage of data from a dementia study with COVID and other data, along with use of a linkage database to facilitate recruitment for studies related to dementia (all locations)*

Suppose that you decided to participate in a study of dementia looking for early risk factors and treatments. You don't have any health concerns and consider yourself a fairly healthy person. However, you've seen friends and family suffer from dementia and you hope to live well as you age, so you agree to participate. All you are asked to do is answer a basic questionnaire about your daily habits. You also agree to share your personal health data to help this and other research projects. You check a box that says, "I agree that my data will be shared for future biomedical research."

Suppose that completely separate from this study, you provided blood and saliva as part of routine testing for COVID-19. You learned that you had COVID-19, but your symptoms were mild. You had a runny nose and sore throat that went away after about one week. You again check a box that says, "I agree that specimens and associated data from my COVID-19 test will be shared for future biomedical research."

Months after the dementia study and COVID-19 testing have finished, you receive a call from a different research team. They ask whether you want to participate in a study of Long COVID. They are trying to determine how previous COVID-19 infection, which can interfere with brain function (from the loss of smell and taste to "brain fog"), may affect early indicators of dementia. But why are they calling you?

The researchers explain that they received information from a database that links data across studies. The database indicated that you had provided blood and saliva as part of routine testing for COVID-19. Although that was a while ago and not for research, you had consented for those specimens to be preserved for future research use. As it turns out, when you participated in the dementia study, the database automatically linked your dementia survey data to your COVID-19 testing data. Those data are now available for other researchers to request for use in other studies. The Long COVID researchers were able to use the database to see that you had previously participated in research relevant to their new study.

Since dementia might be related to Long COVID, the researchers are very interested in your participation. The researchers have already linked the data from the dementia survey to your data from the COVID-19 testing. They let you know that there is still a lot to learn about the long-term impact of COVID infection and that getting additional personal health data from you will inform new interventions and therapies. Additionally, if the researchers find anything significant in your personal data, you'll have the option to discuss these results with their team physician.

Case study discussion questions

1. *How would you feel if you were recruited for the Long COVID study in this way?*
2. *Are you comfortable with having your data linked from different studies like this? Are you comfortable with researchers linking and using other pieces of your data that you have consented to having collected and stored (e.g., from COVID-19 testing)?*
3. *Is there any information that you think shouldn't be linked in these ways? (a) For example, what if data were linked from a drugstore rewards card purchase? (b) What if data came from Medicare or Medicaid or some other health database that has your personal information? (c) What if data came from the national census or databases on local crime, pollution, housing, and other factors affecting communities?*
4. *In this case, you consented to share your information for future biomedical research but not specifically for data linkage. Would you expect that researchers should have to come to you for permission again before getting any linked data about you?*
5. *What else, if anything, do you think should be done in order to link your data in these ways?*

Case study: main points from community participant responses

- Participants expressed some support in the idea of using linkage to contribute to research but were skeptical about the process of data linkage itself.
- Attendees were more likely to support linkage if there was greater transparency about how data might be used, if they received more detailed information during the informed consent process, and if participants would be notified when linked data was reused.
- There was general support for the use of linkage services specifically to facilitate research recruitment. Some community members, however, thought this was incompatible with the spirit of de-identifying participant data.
- Those who supported linkage tended to endorse the idea that initial broad consent gave permission to conduct data linkage, especially for important research. However, participants were generally skeptical of "check the box" consent procedures that do not describe future possible uses in detail.
- Participants tended to be more supportive of linkage when discussing completely anonymized and aggregated datasets, compared to linkage with personalized data, such as from drugstore rewards cards.
- Participants noted a number of concerns with the involvement of private companies in linkage and how many different entities could be making money off of their data. Some participants felt that linkage illustrated how data sharing is already out of the control of individuals and their communities.
- Some participants were hesitant to agree to linkage because it seems too far removed from the initial research study.
- Attendees generally wanted more options to withdraw their participation from data linkage whenever they wanted