

**All of Us Research Program
Research Priorities Workshop
March 21–23, 2018**

Summary of Plenary Sessions

Welcome

Michelle Hamlet, Ph.D., National Institute of Nursing Research

Dr. Hamlet welcomed the attendees and provided an overview of the plan for the day. The keynote addresses and plenary portions of the meeting were videocast.

The Promise of Precision Medicine

Francis S. Collins, M.D., Ph.D., National Institutes of Health

Dr. Collins said that the Research Program, which aims to enroll a cohort of 1 million people, fulfills the NIH mission to pursue fundamental knowledge and apply that knowledge to extend health and reduce illness and disability. The meeting attendees would be asked to provide ideas on how to use the program cohort to maximize the benefits to human health.

Precision medicine is an emerging approach to disease prevention and treatment that transforms the “one-size-fits-all” approach to medicine by tailoring medical care to the individual. Precision medicine is more than a genome project; it also considers the individual’s environment and lifestyle and requires new insights, technologies, and science.

Now is the time to conduct a prospective study of genes and environment. In 2004, an NIH think tank recommended a study of genes and environment that would include a cohort of 500,000. At the time, the cost of sequencing a human genome was \$22 million, and it took two years to complete. Today it costs less than \$1,000 and can be completed in less than a day. In 2004, fewer than 2% of Americans owned a smartphone; today, more than 90% do. More than 90% of health care providers now use electronic health records (EHRs), compared with about 25% in 2004. Computers are 20 times more powerful now than they were in 2004. In addition, people are interested in taking part in research, EHRs are available as research tools, there is wearable technology that allows investigators to gather data, genomics has progressed to include a variety of omics approaches, and data science has advanced to the big data era.

President Barack Obama announced the Precision Medicine Initiative in January 2015. The Initiative now includes the Research Program and a program to research cancer. Precision medicine research has become an important approach in cancer research as investigations of tumor genomics have helped tailor treatment to the individual’s tumor.

Congress has provided strong bipartisan support to the Research Project, authorizing \$1.5 billion for precision medicine over the course of 10 years through the 21st Century Cures Act.

Precision medicine will benefit patients by targeting treatment to the individual patient, not to the average patient as in the current approach. Professional providers will benefit from having more evidence and options for treatment. The patient's complete medical record will be more readily and centrally available. The program aims to enroll at least 1 million participants. With a database of 1 million people, biomedical researchers will spend less time building their research infrastructure, have greater access to a large and diverse sample size, and have greater incentives for data sharing and collaboration.

The program cohort will contribute its EHR data, will complete multiple surveys, will provide a biospecimen for genomic analysis, and may also have data from wearable sensors. The data will be useful to get a more detailed understanding of diseases and conditions that are grouped by symptoms but that may be different and require different interventions.

A national research cohort of this size will likely include any condition that is reasonably common, and it will provide researchers the opportunity to follow those cases prospectively. The platform will accumulate data for a variety of conditions, and partners may be able to provide additional subsets of information on diseases, including rare diseases.

This national cohort could be a useful way to discover therapeutic targets. Some mutations that are attractive therapeutic targets include the *PCSK9* mutation for low cholesterol and decreased risk of heart disease and *SLC30A8* for type 2 diabetes.

The national cohort could help identify the biological factors associated with resilience by identifying genetic and environmental protective factors. What is the connection between a presenilin mutation and reduced risk of Alzheimer's disease? Why is it that some smokers do not develop cancer or cardiovascular disease?

The cohort will likely inform pharmacogenomics. Early opportunities include work with antidepressants, warfarin, and lipid-lowering agents.

The cohort presents an opportunity to explore what works in the areas of behavioral research (e.g., medication adherence, tobacco cessation), physical activity (e.g., monitoring of activity in real time, ongoing measurement of cardiovascular status), prevention (e.g., prediabetes, birth defects), and chronic disease management (e.g., asthma, obesity). Participants in the cohort are motivated to take part and have consented to be recontacted for follow-up studies. The cohort will present an opportunity to find out what can be learned about outcomes from wearable sensors.

The Research Program also represents an opportunity for global collaborations with at least 50 other studies that are collecting data from 100,000 individuals. Linking these large cohorts could greatly increase the capability for discovery.

Dr. Collins introduced several Research Program participants who were attending the meeting and taking part in the breakout and brainstorming sessions.

Dr. Collins said that he expects that workshop attendees will identify the key research priorities and the data needed to capitalize on the cohort in a way that will help advance precision medicine.

Program Overview

Eric Dishman, All of Us Research Program

Mr. Dishman welcomed the attendees and said that they were invited because of their expertise. There were also 11 participant representatives at the meeting, representing the more than 19,000 people who have enrolled in the Research Program and completed the protocol.

The attendees would be asked to work in groups to develop research priorities for how a cohort of 1 million participants could be used to solve health problems. The first step is to focus on how researchers will want to use the database and what the Research Program must do to ensure that it can fulfill the researchers' requirements.

The program's mission is to accelerate health research and medical breakthroughs to enable individualized prevention, treatment, and care for all. The program will accomplish this by building one of the richest biomedical datasets ever, a resource of clinical, environmental, lifestyle, and genetic data that anybody can access. To do that, it will be necessary to nurture relationships with 1 million people.

The program is described as a 10-year study, but the hope is that it will run between 60 and 80 years. This will require a lot of brainpower to focus on it and to catalyze a robust ecosystem of researchers and funders to use the data. The program also aims to establish relationships with the 27 NIH Institutes and Centers (ICs) to involve them in funding research related to the cohort.

The core values of the program are that it will:

- Be open to interested individuals
- Reflect the country's diversity
- Engage participants as partners
- Engender trust by being fully transparent
- Provide participants access to their own data
- Make the program data broadly accessible to empower return of research information
- Adhere to privacy, trust, and data security principles
- Catalyze innovative research

The program is especially interested in enrolling groups who are underrepresented in biomedical research (UBRs) and engaging a diverse group of researchers. It is also important to enroll healthy people so that researchers can track health problems as they unfold.

The protocol includes the general consent and a separate consent to share EHR data. Participants fill out three surveys (basics, overall health, and lifestyle), physical measurements (blood pressure, heart rate, body mass index, height, hip and waist circumference, and weight), and biosamples (blood or saliva and urine). Future versions of the protocol will include children and new surveys and measurements.

The consortium met for the first time in July 2016. Since then, the program has built a network of more than 100 academic, provider, technology, community partners, and advocacy organizations. Among other initiatives, the program has developed a partnership with the National Library of Medicine (NLM) to help recruit in libraries; developed a bilingual enrollment website, a participant portal, and a call center; and constructed a biobank with the capacity to analyze 35 million vials. The Research Program has been designed with input from participants and patients.

At the time of the meeting, the program was also running its “surge” to test its capacity. The program had completed security and usability testing, obtained its authority to operate (ATO), pilot-tested the method to obtain EHR data, and incorporated wearables into the program.

The protocol will continue to be improved and expanded by identifying new scientific needs and opportunities and implementing them into the protocol. Version 2 of the protocol will be implemented in about three years. The use cases will help set the scientific priorities for Version 2 and beyond. The program will release a new version of the protocol approximately every three years.

The long-term goals are to allow participants to share their EHR data with the press of a button and to make the data easily usable by researchers. Currently, the health care provider organizations (HPOs) send EHR data directly to the Research Program’s Data Research Center (DRC). Sync for Science (S4S) will help direct volunteers (DVs) share their de-identified EHR data with the Research Program and, in a few years, with any other research program. Eventually, the program will begin collecting data from specialty EHRs, such as dental EHRs. The program will also use PicnicHealth to allow participants to more easily gather and centralize their EHR data from all of their providers. The data aggregation effort is expected to take a decade to complete. The program will also do a pilot study in the future to compare the data acquisition capabilities of the EHRs, PicnicHealth, and S4S.

The program is collecting blood and urine specimens, but that could be expanded in future protocols if the scientific priorities demand it. For example, some have suggested collecting hair or fingernail samples and testing them for environmental exposures.

The program plans to do genotyping and whole genome sequencing for all 1 million participants as fast as funding and capacity will allow. The goal is to begin in fall 2018 with either two or three genome centers that the program will fund. The centers will process 100,000 participants per year in 2018 and 200,000 per year by the end of 2019. The goal is to genotype all participants by 2023 and to do whole genome sequencing on all participants by 2025.

The program will pilot the responsible return of genomics information in the fall. The program expects to begin sharing de-identified genomic data with researchers by the first half of 2019. Currently, there is not enough genetic counseling capacity to return the information to all of the participants, but there will be funding to hire and train more genetic counselors.

The program aims to launch the researcher portal in 2019. There will be three access tiers (public, registered, and controlled). Genomic data will be in the controlled access level. Researchers who successfully apply to use the data at the registered and controlled levels will receive a “passport” to access the database. Researchers must publicly state how they will use the data that they access. This system is meant to ensure efficient access while maintaining data privacy and security. The program continues to work on a way to give registered-tier access to citizen scientists who are not affiliated with an institution.

The program aspires to grow beyond 1 million by using a “franchise” model in which other groups use the program’s methods to collect data. This will allow partnering with other cohorts, including some rare disease cohorts that could produce larger sample sizes. The program also plans to share data with other large cohorts, such as the UK Biobank and the Department of Veterans Affairs Million Veteran Program.

Mr. Dishman said that he worked on a project related to Parkinson’s disease that was similar to the Research Program, but on a smaller scale. He and a colleague convened researchers to identify questions about the disease. The researchers made a list of the fundamental questions about Parkinson’s disease, including “What is Parkinson’s disease, really?” The project leaders held a second meeting with the researchers but also included patients, patient advocates, pharmaceutical representatives, physicians, and engineers. The group built an at-home testing device for disease progression research, a research platform that they shared. Bringing together the researchers, patients, and families helped improve cooperation among the researchers, accelerated the research, and diversified the scientific approaches. It also stimulated more funding to research the disease.

The Research Priorities Workshop would use a similar approach, but with larger numbers of participants and researchers and many more health conditions. Attendees were asked to bring their own research ideas, build on those ideas during the workshop, and support other attendees to build on their ideas. These ideas would be developed into “use cases.” Use cases begin with a health condition, identify a key problem in relation to the health condition, identify the research question to address the problem, propose a study to answer the question, and identify the data types to be used, the methods to obtain the data, and the specifications for using the method.

The program had already received more than 800 scientific use cases through IdeaScale before the workshop began. The goal was to generate hundreds of additional use cases that would cover a wide range of scientific areas, diseases, and conditions with specific details about the data needed to make breakthroughs.

The workshop attendees would also be asked to specify who else could benefit from the methods and tools suggested in each use case, whether different stakeholders could agree on common specifications for the tools, and whether the approach would be affordable.

Mr. Dishman provided an example of a use case drawn from Parkinson’s disease. The problem was early detection and monitoring of the disease. The research question looked at voice

changes over time. The researchers debated how to measure voice changes over time, but settled on using a smartphone app as the method to obtain the data, in the form of a 30-second voice sample recording. The study tested voice energy and fundamental frequency variability from periodic voice samples.

When creating use cases, Mr. Dishman asked attendees to consider whether the data they specify for their use case could be used for other research purposes and what the data specifications should be to allow the widest range of use. He also provided tips for brainstorming and for working together on attendees' ideas.

The workshop attendees were asked to develop use cases within nine major areas of health: health and resilience; cardiovascular, respiratory, and blood; mental health and addiction; digestive, renal, and metabolic; musculoskeletal and dental; sensory, pain, and neurologic; human development and aging; and cancer. Attendees were asked to consider seven cross-cutting themes for each condition. The cross-cutting themes were risk factors, prevention, and wellness; health disparities and health care quality and access; genomics and other omics; mobile health; therapeutic and preventive interventions; environmental and other contextual effects; and informatics, methodologies, ethical/legal, and statistical research. Attendees would look for recurring themes, needs, datasets, and tools.

The workshop attendees could also make suggestions for additional participant surveys, EHR data, biospecimens, wearables, assays, and other omics.

Keynote Presentation

Greg Simon, M.D., M.P.H., Kaiser Permanente Washington Health Research Institute

Dr. Simon said that this is a time of “glorious disruption” in biomedical research. Diagnostic boundaries are shifting, hypothesis testing is being replaced by prediction, and barriers to entry into research are rapidly falling.

In the world of big data, it is likely that all null hypotheses will be rejected if the sample size is large enough and the investigator looks at the right subgroup. The findings will be statistically significant but not specific enough to be illuminating.

The Patient-Centered Outcomes Research Institute provides three questions for patients to ask:

- Given my personal characteristics, conditions, and preferences, what should I expect will happen to me?
- What are my options, and what are the potential benefits and harms of those options?
- What can I do to improve the outcomes that are most important to me?

None of these questions can be stated as a null hypothesis, because they are about the future and are personal. However, they can be stated mathematically as a Bayesian question: “Given what I already knew and what I now observe, what can I predict?”

The only thing getting cheaper faster than computing is genetic sequencing. One of the challenges is that as the rate of progress increases, the cultural barriers to who benefits will come into sharper focus.

Rapid technological advances will change the research business model from one that maximizes grant revenues to one that maximizes scientific aims. The current concern of investigators is how well their project funding can support their lab. With large databases, there will be less pressure to support a large team, because so much of the work is done. But this new business model means that new ways to evaluate and reward young investigators must be found.

Research will require investigators to work with a wider array of partners to understand a condition from the level of genes, cells, organs, symptoms, impairment, and handicap.

The health conditions and cross-cutting themes that the workshop attendees will use demonstrate that investigators from different fields and different expertise must work together. For example, investigators working on obsessive-compulsive disorder may partner with those working on acute glomerulonephritis, which is related.

Investigators will thrive if they remember who they work for. Investigators in the Research Program ultimately work for patients and their families, future patients, and the taxpayers. Investigators also answer to a health system (when the health system represents patients who have the condition) and to funders, who answer to the taxpayer.

Investigators will thrive if they handle their mistakes and failures by admitting them quickly and making them public. Sharing, including sharing data, can be hard, but resources are worth more together than separately.

Keynote Presentation: Research Me!

Bray Patrick-Lake, M.F.S., Duke University

Ms. Patrick-Lake is the co-chair of the Advisory Committee to the Director Precision Medicine Initiative Working Group, a member of the Research Program Advisory Panel, and a patient advocacy foundation president. She said that it is time for a major course correction in health care and research.

The first ground rule for the workshop is that participants come first. Precision health is fueled by engaging, impactful research that is person centered and efficient. It can generate evidence-based prevention strategies, reliable and timely diagnoses, and effective treatment options that are responsive to individual needs, characteristics, and preferences.

It is difficult for patients to be healthy when their health records are incomplete and scattered across different providers. Health care must take a holistic view of the person, including the community frameworks within which the person is embedded. Currently, treatments are “our best guess” medicine, missing the holistic view of prevention, diagnosis, and treatment.

It takes 12 years to get treatments to people in need. When drugs reach the market, it is difficult to know whether they are suitable for an individual patient. A drug may benefit some patients but harm others. This wastes resources on those patients who are not at risk, those who will not benefit, and those whose transition to disease could have been prevented with another treatment. The current research model also fails to ascertain interpatient variability. Two patients may live 10 years on a particular treatment, but the effects of the treatment may differ greatly.

There is a need for a more integrated research approach to understanding health and disease, including understanding environmental exposure. Ms. Patrick-Lake's community has an active oil and gas industry within its borders and many complaints of illness among its residents. ZIP codes and credit scores are predictors of mortality, but they are not included in traditional studies.

Family history is important. What genetic or predictive markers should researchers look for in children of addicted parents to help identify early interventions and save lives? Children within the same family can have very different physical and mental health outcomes. Even identical twins who have a similar history can have very different health outcomes.

What data elements and participant-provided information should investigators collect to understand the difference in the health outcomes of identical twins? How could technology capture the data needed to answer this question? How can the Research Program share its findings with participants, health care providers, policy makers, and the community to create value and have the most impact on health?

Migraines are associated with a significantly increased risk of cardiovascular disease (CVD). The Research Program presents an opportunity to study this association longitudinally while also tracking variables such as migraine frequency, use of preventive medications, other CVD risks, and migraine comorbidities. The program can also help develop a natural history of disease and provide a way to launch and complete clinical trials more rapidly.

The workshop plan to look at health conditions and cross-cutting themes is a way for investigators to move away from their silos to make discoveries across the life span and health condition. The program should engage participants by making participation enjoyable (e.g., by sharing facts about the participant's health with them).

It is time to be bold about doing research differently to do justice to people as unique, multidimensional individuals embedded in dynamic family and community frameworks across the life span. Researchers should take advantage of the diversity of people and richness of data of this scientific resource to develop medicines to hit the right target on the right patient with the right dose at the right time. Together, those involved in the program must generate novel ideas and keep innovating and adapting the study to keep participants engaged and meeting the community and scientific needs.

Workshop Goals and Charge

Gina Wei, M.D., M.P.H., National Heart, Lung, and Blood Institute
Scott Wheeler, M.S., Strategy Arts

Dr. Wei said that the goal of the workshop is to tap the brains of all of the attendees to get ideas for future versions of the protocol. Attendees were nominated by the ICs based on their expertise and diversity of thought.

When thinking about the use cases, attendees should go beyond their own areas of expertise to develop use cases that would have the greatest impact. Attendees who have smaller-scale ideas should jot them down and submit them when the process for ancillary studies has been developed.

Mr. Wheeler said that the intent is to gather as much input from the attendees as possible, gather the most important questions about a health condition of interest, find out what data is needed to answer the important questions, and find out how the questions could accelerate the field of research.

Workshop attendees would begin their work in cross-cutting theme breakout groups. Organizers assigned each attendee to a group. The attendees would later move to a health condition breakout. Within these sessions, the attendees would identify the most important and impactful questions that the Research Program could answer.

On Day 2, attendees would discuss what data would be needed to answer the questions they had identified. Attendees would also identify potential protocol elements (PPEs), including data, methodology, and specifications.

On Day 3, participants would identify the most impactful data to develop the use cases. They would also identify which questions could be done in the short, medium, and long terms. At the end of Day 3, participants would return to a general session and talk about what they did.

Each breakout had a facilitator, a person to enter the data, and a protocol expert. Each breakout would also include NIH staff members who helped plan the meeting.

Mr. Wheeler asked attendees to accept that there is uncertainty, to think about what data is needed that researchers do not know how to get, and to encourage everybody to present their views.

Discussion

A participant asked whether he should set aside his expertise in one disease when he is in the breakout discussions. Mr. Dishman said that participants should do both.

An attendee asked how much a use case should encompass. For example, there could be a use case on metabolic disease and sleep, but if exercise is also important, would there be a second use case on metabolic disease and exercise, or could they be combined into one use

case? Mr. Wheeler said that the facilitators will help with such issues. The groups should not become too bogged down in deciding what encompasses a use case. The most important thing is to identify the major questions and the data needed to answer the questions. Mr. Dishman said that the groups should think about the data and tools needed to capture the big questions. Start with the high-level concept (e.g., cancer) and drill down.

An attendee asked whether they should imagine themselves in the research world as it is or in the research world that they want. Mr. Dishman said that it helps to think about what happens now, next, and in the more distant future. Usually, a group will anchor in a particular time and then will move to another timeframe. Mr. Wheeler said that the data experts will collate the data elements at the end of the second day and assess which questions can and cannot be done now. Mr. Dishman said that the attendees do not have to solve the problem or worry about whether a project is doable. Research Program staff can do that later.

An attendee said that some questions, such as the early life determinants of disease, can be answered only by using data from children. Should the groups work on those questions, given that children are not yet part of the study? Mr. Dishman said that the program will begin enrolling children in 2019, so the groups should work on those questions.

An attendee asked about the new approach to research that Dr. Simon described. What replaces the traditional motivators for research (i.e., funding and career advancement)? Mr. Dishman said that it will take time to resolve that question and that it is well beyond the scope of the workshop. However, if groups have ideas about that, they can submit the ideas.

Summary of Days 1 and 2 and Charge for Day 3

Carolyn Hutter, Ph.D., National Human Genome Research Institute

Robert Carter, M.D., National Institute of Arthritis and Musculoskeletal and Skin Diseases

At the end of the second day, after the attendees met in groups during the first and second days, Dr. Carter recapped the progress that attendees had made and gave a preview of the activities for the next day. During these first two days, the attendees had produced more than 500 use cases.

Dr. Hutter displayed network visualizations of the health conditions that the use cases encompassed. Health and resilience was one of the larger areas. Many of the health and resilience use cases came in through IdeaScale. Attendees would be asked to further refine some of the larger areas shown in the visualization on the final day of the meeting. For example, environmental exposures could include air pollution and other factors.

Risk prevention and wellness was the most often mentioned of the cross-cutting themes, making it an important area of research for the program. Risk assessment was the largest of the scientific categories.

On the third day, the attendees would identify the most important protocol elements in each cross-cutting theme, discuss how to collect those elements, and suggest ways to overcome

data collection barriers while minimizing participant burden and the financial burden on the program.

Discussion

An attendee said that she did not see health disparities in the visualization. If that problem has receded to the background, it should be brought to the foreground, because it is an important problem. Dr. Hutter showed that it was among the cross-cutting themes on the visualization labeled “HD HCQ.” The attendees will be asked to discuss that theme on the final day of the meeting.

An attendee recommended that the conference organizers suggest ways in which the attendees can bring this work back to their colleagues and continue to contribute. Mr. Dishman said that the program will continue to do a series of small-group workshops to develop the use cases more fully. Dr. Hutter said that staff from across the ICs helped organize the workshop and will discuss with their ICs how to work with the Research Program on these areas.

An attendee asked whether the participants would receive a draft of the workshop’s finished product. Mr. Dishman said that the program will make public the raw products and the synthesis that comes from the workshop. It will be an iterative process as the program focuses in on priority areas based on health impacts and feasibility. The ICs will fund a lot of the work. This work will continue over the 10-year (or longer) life of the program.

An attendee asked about the plans to partner with big data users such as Amazon and some of the large health companies. Mr. Dishman said that the program would; otherwise, the database could not be used to its maximum potential. Joshua Denny, M.D., said that the DRC is partnering with Verily Life Sciences (formerly Google Life Sciences). The DRC has already done a lot of work on data mining of EHRs and has been working with a variety of nontraditional partners to bring additional expertise in areas like machine learning and natural language processing. David Glazer said that the DRC is currently putting its greatest effort into the data infrastructure to ensure that the program can collect, store, and index the data. Their partners will help ensure that the infrastructure is designed so that users can do the analyses that they want. There are many organizations that the Research Program can partner with in the analysis area. Mr. Dishman said that NIH is working on a broader plan for big data that the Research Program can tap into. Dr. Denny said that the DRC will make different datasets available over time and will release the timeline as it is developed.

An attendee asked about the strategy to enroll UBRs. Dara Richardson-Heron, M.D., said that the program is engaging a diverse group of people by meeting them where they live, ensuring that participation is not burdensome, and making them feel valued. The research literature shows that the reason most often given by UBRs for not participating in research is that they are not invited, not that they are reluctant to participate. The program has a mobile engagement asset that is visiting UBR communities, has formed partnerships with community organizations that have strong ties to UBR communities, has selected HPOs and DV organizations that are located in and around those communities, and is beginning to include UBR participants in the

program's work. Dr. Rutter said that UBRs will be involved in the work of the cohort so that all of the program's work represents the diversity of the population.

An attendee asked whether the homeless population is on the program's list of vulnerable populations. Mr. Dishman said that the program's Special Populations Committee is beginning to plan to include each of the vulnerable populations but will make separate plans for each one. Also, the federally qualified health centers are enrolling vulnerable populations, including people who are homeless. Some of the private partners, such as Walgreens, will further extend the program's reach. The goal is to reach these populations wherever they are.

An attendee asked whether the program will make the data freely available. Mr. Dishman said that the program will do so. Dr. Hutter said that the program may develop a separate consent for participants who want to make subsets of their data fully open. This topic is still under discussion.

An attendee asked whether caretakers would be allowed to consent on behalf of individuals with intellectual disabilities. Dr. Rutter said that the program expects to develop such a plan.

Kathryn Schmitz, Ph.D., M.P.H., asked whether the partnership with private companies extends beyond Walgreens. Mr. Dishman said that the private companies that are involved are those that applied during the first round. The program is currently focusing on the team that it has, but it will be open to new partnerships in the future.

Dr. Hutter adjourned the meeting at 5:16 p.m.

Day 3

Welcome Back

Dianne Babski, M.I.M., NLM

The day began with attendees attending breakout sessions on cross-cutting themes. The goals were to identify the game-changing potential of protocol elements in each cross-cutting theme and to develop ideas to overcome current data collection barriers. Following the breakout session, attendees came together for the final session.

Panel: Takeaways and Building Momentum

Stephanie Devaney, Ph.D., All of Us Research Program

Carolyn Hutter, Ph.D., National Human Genome Research Institute

Dr. Devaney thanked the attendees for the work that they had put into developing high-quality use cases and identifying the data elements needed to address the key questions. Attendees also identified some of the game-changing data types for the program. The workshop has also helped to build a community that is involved in the program.

Following the meeting, the Research Program staff will continue to analyze the information in more depth and to refine the information into the scientific priorities.

Dr. Hutter said that the attendees had produced more than 500 use cases and that there are an additional 800 use cases developed online using IdeaScale. She presented bar graphs showing the number of use cases by health condition and by cross-cutting theme. The number of use cases will shrink as Research Program staff go through them and combine use cases that are essentially the same.

Dr. Devaney said that the attendees identified 100 game-changing data types across the cross-cutting themes, 86 of which were designated the most important research questions.

The facilitators of the cross-cutting themes breakout sessions summarized the discussions that they had had earlier in the day.

Health Disparities, Health Care Quality, and Access

Consuelo Wilkins, M.D., M.S.C.I., Meharry-Vanderbilt Alliance

- Requisites for health disparity research were not discussed often during the conference, in part because health disparity researchers are a minority among the investigators.
- Diversity alone is insufficient to study health disparities. The push for a diverse group of participants is laudable, but the program must develop the contextual information to understand those populations. The contextual information is needed to understand how to eliminate disparities.
- Health disparity research must incorporate a wider range of data elements to understand the effects of discrimination, oppression, and other stressors on health. There are ways to measure them without burdening participants. The process for collecting these data elements could be the game changers.
- Participants should be given choices for which surveys they will complete. Participants could complete the modules that are most relevant to them.
- Capturing and understanding the stories of UBRs is just as important as collecting quantitative measures.
- Some of the most pertinent data about communities are not being captured. For example, individuals who live in a community with a large number of stray dogs may be reluctant to go for a walk in their neighborhood.
- Life course data is important to collect.

Risk Factors, Prevention, and Wellness

Beth Karlson, M.D., M.S., Harvard Medical School

- The group identified the top five risk factors from a list of 22 risk factors.
- The program should measure both cognitive function and physical function as indicators of wellness. Use a game-type format to measure cognitive function. Measure physical function by using a six-minute walk test, a grip strength test, and more.
- Measure all five senses to get a more holistic view of wellness.
- Use smartphone apps and annual surveys to measure diet.

- Measure biomarkers, including metabolites in the blood, as a way to track the metabolome and the “auto-antibiome” that measures antibodies in the blood that predict autoantibody diseases that are present in 10% of the population. Also measure inflammatory factors in the blood.
- Among possible measures, the group ranked hair and nail samples as the highest priority, with stool and vaginal samples rated next highest.
- Track women from pre-pregnancy through birth to help understand pregnancy outcomes.
- Do not do a urinary toxicology screen for opioids.
- Retinal images are a very promising new technology. Currently, the scan requires special equipment, but in the future it may be possible to use a cellphone to scan the retina. Retinal scans provide a lot of information about potential risk factors.
- The group was given a list of 22 risk factors, but the group judged the list incomplete. The group recommended adding mental health and sleep and circadian rhythms to the list.

Therapeutic and Preventative Interventions

Ariana Smith, M.D., University of Pennsylvania

- Measure participants’ health literacy, including their health knowledge and their barriers to good health.
- Obtain information about the participants’ providers, such as their race and ethnicity, where they are educated, and how much time they spend educating their patients.
- Measure nutrition across cultures. Include questions about where the participants learn about diet and nutrition.
- Measure the barriers to healthy eating.
- Use video to capture participants’ environment and food intake.
- Consider using dental chips as a way to track dietary intake.
- Look at RNA, centromeres, and telomeres to monitor health over time.

Mobile Health

Codrin Lungu, M.D.

- The group discussed existing technology and what new technology is needed. In some cases, the technology that the group deemed valuable is already available.
- The group discussed passive versus active data collection.
- Privacy concerns may be more prominent in this theme than in other cross-cutting themes.
- Sleep monitoring is a game-changing element, and some of the technology to measure sleep stage is already available.
- Retinal imaging is a valuable neurological measure. Mobile retinal imaging is already available.
- Facial expression is important to measure.
- Mobile sensors for environmental exposures should be employed.
- Air conditioner filters from participants’ homes should be collected and analyzed.
- Smartphones and smart watches should be used to measure gait, balance, stride length, and total mobility. Some of these sensors could be set up in the homes and could indicate changes in health status over time.
- Sensory experiences, such as olfactory and auditory exposures and taste, should be measured.
- Weather data should be collected.

- Brightness settings on participants' smartphones should be recorded.
- GPS data should be captured.

Informatics, Methodologies, Ethical/Legal, and Statistical Research

Meg Doerr, Strategy Arts

- The potential protocol elements the group received were necessary but not imaginative. The group decided to provide its “big ideas.”
- The statisticians want to build an exposome framework to map all exposomes down to the participant level. This would require standardized ontologies and content.
- The methodologists want to focus on recruiting and retaining a diverse cohort by using natural language processing, gamification, data tools, and more. This would be an effort to maintain diversity in the participant pool.
- The ethical and legal group asked for an ethical, legal, and social issues program that would track the social and policy impacts of the program over time. Self-forming groups on social media could push information from social media to researchers.
- There is a need for a policy databank to monitor policy and policy change over time, such as the legalization of cannabis in Colorado and its effects on cardiovascular health compared with that of the rest of the nation.

Genomics and Other Omics

Michael Snyder, Ph.D., Stanford University

- The group went through the various omics, examining each in terms of costs, throughput, and whether they should be long-term or short-term goals. The omics included the genome, epigenome (DNA methylation and chromatin assays), microbiome, metabolome, proteome, transcriptome, CyTOF (a mass spectrometry–flow cytometry device profiling immune cell types), immunome, exosome, and exposome.
- The group discussed the best way to obtain samples so that they can meet the needs of the omics known today and those needs that are likely to become known in the future.
- Blood would be the primary sample, but skin, fingernails, and breath could also be used.
- High-throughput imaging is important.

Environmental and Other Contextual Factors

Andrea Ramirez, M.D., M.S., Vanderbilt University

- Statistical research should be added as a cross-cutting theme. There is a need for new methodology to handle the scope of the database and to isolate the interaction effects across all health conditions.
- Determine what is motivating participants, and decide how to implement a system that is motivating.
- Use open-space methodology.
- Ask about positive events, not just negative ones.
- Use hair samples to determine dietary intake of substances such as high-fructose corn syrup.
- The microbiome and geocoding are other data elements that should be included.

Impactfully Engaging Participants

Dara Richardson-Heron, M.D., All of Us Research Program

Dr. Richardson-Heron presented the top five tips for impactfully engaging participants:

- Lower the barriers to enrolling in research, including by providing child care, streamlining the consent process, offering transportation, collecting data in a way that is not burdensome to participants, making the research more impactful, and using inclusive language.
- Don't just fill the seats; engage the people who fill them. Make participants part of the team and elevate their voices. Never underestimate what a non-scientist can bring to the table. Participants help set the program's standards, can change the questions that researchers ask, and reorient the research to produce benefits that the participants need. The goal is to retain participants throughout the life of the program.
- Develop unlikely partnerships, which are particularly useful for large, audacious, and complex tasks like the Research Program. The program requires researchers to reach outside their typical networks. The research paradigm is changing. Research is being democratized.
- Get everybody's perspective. The program must represent the diversity of the United States, including a significant number of UBRs, to really be the "All of Us" program. The research team must be culturally diverse and sensitive.
- Impactful engagement occurs only when there is a bidirectional transfer of knowledge. It will be important to focus on resources to ensure that the partners and participants have the knowledge that they need. Success will depend on how well the program engages everyone. Success means authentically and impactfully engaging and retaining all people to leverage their skills, experience, and expertise to accelerate health research and medical breakthroughs to enable individualized treatment.

Reflections and Next Steps

Eric Dishman, All of Us Research Program

Mr. Dishman noted that this meeting used a unique format. Some attendees liked the format, and some did not, but all worked within it. The attendees produced more than 500 use cases, 100 game-changing ideas, and an enormous amount of data. The program must now dig into those use cases, ideas, and data to refine them further. Mr. Dishman highlighted some of the discussions that he heard taking place within the breakout groups.

- Pharmacogenomic data should be collected to predict side effects in the individual.
- Retinal imaging can be used to investigate a number of problems. Conversations about this and other technologies led to further ideas about how technology could be used.
- The program needs to learn how to measure social networks and track how those networks change over time.
- The program could help develop a genomic risk score for hearing loss.
- The topic of resilience came up often, but how should resilience be defined, and how can investigators study it?
- The program should try to further the understanding of itch in conditions such as eczema.

One task is to take "snapshots" of major vectors over the life span. How can the program take those snapshots, how often, and at what ages? What is the most affordable way to do that?

This is the very beginning of this study, which is slated to run for 10 years but could continue for many years beyond that. There will be many opportunities to change and improve the program over that time.

The program staff will post all of the use cases on the website as a searchable resource. The trans-NIH team will begin looking at how the ICs could be involved in further development of the use cases. Further work on the use cases will begin during the summer.

The program will prioritize the use cases. Those that could have a high impact and are relatively easy to do would be given priority over those that would require more work, time, and funding. Upcoming use case workshops will be smaller and focus on only some of the use cases. The criteria used to evaluate the use cases include each case's impact on people's lives, its scope, the current state of the field, how quickly it can be executed, the burden that it places on the participant, whether there are researchers to do the work, what burden it places on the program, and how it fits with the overall program balance.

Adjournment

Mr. Dishman thanked the participants for their hard work and adjourned the meeting at 12:38 p.m.