Fairness in Precision Medicine

February 2018

Kadija Ferryman
Mikaela Pitcan

Data&Society

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The image of "precision medicine" is
a dream of what medical care could become—
driven by data analysis and tailored to individual
patients. Access to data, specifically large
volumes and varieties of health data, could help
health care providers intervene and begin to
address some of our health care ills.
However, this dream should also prompt us to
question how precision medicine might
not develop in the ways that we think, and how
bringing the increasing power of computing
to bear on more and more kinds of health data
could have unintended consequences.

To better understand how bias might impact precision medicine in the sphere of biomedical research, we conducted a qualitative study to identify the tensions and frameworks of diverse medical stakeholders. Our goal was to understand how precision medicine research projects are developing, as these will shape the future directions of clinical precision medicine. Building on what we learned from the research community, we define precision medicine as: the effort to collect, integrate, and analyze multiple sources of data in order to develop individualized insights about health and disease.

Health data is subject to bias in multiple ways, and the increasing quantity and types of data that are available today can make it hard to identify where bias can emerge. In an ecosystem in which research can inform clinical guidelines and treatments, biases can have potentially life-threatening impacts. Our findings identify two main types of bias in precision medicine: 1) bias in the building and analyzing of datasets, and 2) bias as the result of precision medicine research.

> **Bias in Datasets:** datasets can become unintentionally biased through a) a lack of cohort diversity, b) technical processes of data collection and cleaning, and c) the specific incorporation of electronic health record data.

> **Bias in Outcomes:** the outcomes of precision medicine research can be discriminatory in many ways. These include a) too much focus on individual responsibility for health, b) the marginalization of those population groups with lower health literacy or in less resourced areas, and c) the potential to shift the accepted forms of biomedical research.

Challenges remain for achieving the vision of precision medicine. Research is needed to develop a comprehensive understanding of barriers and ensure that the structural determinants of health are acknowledged and used to plan intervention, rather than attributing responsibility solely to the individual. Policy and regulation is needed to ensure that the insights derived will not later be used to surveil and marginalize vulnerable populations. Above all, patients and the public need to be engaged in this endeavor both to guarantee its success and to hold other actors accountable for potential misuses or misunderstandings about when and how their data should be used.

INTRODUCTION

Imagine that in the future, someone seeking medical care meets with a clinician who has a data resource that includes not only her medical history but also her genetic sequence and activity tracker information, as well as data about her housing, water, air quality, and the strength of her social networks. From this data, the clinician would know what diseases she was at risk for before developing any symptoms, and would even know which medications would work best if there was disease onset. For some, this scenario is empowering; for others, it's terrifying. Yet, this is the ideal of precision medicine, an emerging approach that aims to capitalize on the growing availability of health data to both deliver better care to individuals and to improve the efficiency of the health care system as a whole.

RESEARCH QUESTIONS

This dream of a precision medicine future contrasts with the reality of medical care in the United States today. The United States spends more money on health care than any other country, but does not correspondingly have the best health outcomes. Furthermore, health disparities among demographic groups are staggering.[1] Access to data, specifically large volumes and varieties of health data, could help health care providers intervene and begin to address some of our health care ills. The rapid digitization of medical records[2] and the advances in computing and data analysis techniques mean that other kinds of data, in addition to genetics, can also be mined for insights about health and disease.[3] The hope that new forms of data and information processing could be harnessed to lead to better medicine has been an important impetus of medical research for many years. For example, scientists hoped that sequencing the entire genome in the Human Genome Project would not just advance molecular biology, but would be applied to medicine and lead to exciting and unprecedented insights about health.[4] Though we can trace the genealogy of precision medicine from genetic research, another important part of the story is increasing scientific interest and research on other, nongenetic factors for disease.

INTRODUCTION

For example, epigenetics is an emerging field that seeks to understand how non-inherited factors, such as maternal nutrition, can lead to physiological changes and disease risk.[5] Excitement for precision medicine abounds in the U.S. and globally, with companies, academic institutions, and governments investing heavily in precision medicine research and treatment programs.

The dream of precision medicine is a techno-utopia, built on real evidence of how big data analysis can transform other fields. It emphasizes health as determined not just by biology, but on a complex interplay of genetic, social, and economic factors. However, this dream should also prompt us to question how it might not develop in the ways that we think it will, and how bringing the increasing power of computing to bear on more and more kinds of health data could have unintended consequences. The use of data in sectors like criminal justice, welfare, and child services has exacerbated inequalities and caused significant harm to individuals.[6] Additionally, there's a long-standing history of misuse of medical data that disproportionately impacts poor people or regulates their access to services.[7]

Though there is much excitement about precision medicine, there is also skepticism about this promise of a data-driven future of medicine. Nathaniel Comfort, a historian of medicine at Johns Hopkins University, wrote an assessment of the hype surrounding precision medicine as part of a longer history of "inflated medical promises."[8] It is important to recognize that the causes of disease are complex and dependent upon multiple biological and social factors, and precision medicine interventions will need to be able to sufficiently address this complexity. There are also divergent interests among stakeholders in precision medicine—patients' desires may differ from the goals of clinicians or pharmaceutical companies, insurers, or other industry players.[9] Precision medicine will need to face the challenge of bringing together these "misaligned" interests and reducing fear and anxiety among those who are wary that such a move is in their best interest. In addition to addressing these challenges, the excitement for precision medicine must be matched with clear-eyed examinations of the possibility for this emerging field to exacerbate existing risks and harms or create new points of vulnerability to bias and discrimination.

The Fairness in Precision Medicine project is one step in this direction. This project grew out of an imperative to begin to identify potential pitfalls, notably the potential for biased and discriminatory outcomes in precision medicine. The aims of the project are to identify nodes of activity and key stakeholders, connect with these individuals to identify the emergent tensions in precision medicine, and explore possible technical, organizational, and policy-oriented remedies.

"We define precision medicine
as the effort to collect, integrate, and analyze
multiple sources of data in order to develop
individualized insights about health
and disease."

DEFINING PRECISION MEDICINE

We began this study by examining how and where the term "precision medicine" was currently being used in the scientific, medical, and health care fields, including in scholarly, popular, and industry contexts. It soon became clear that precision medicine could refer to activities in the clinical space of medical practice, or in biomedical research. In the clinical space, the term precision medicine most often refers to the use of genetics to tailor medical treatment for cancer. Clinical precision medicine using genetics in cancer treatment includes a number of interventions, including sequencing tumor DNA to identify specific mutations that can provide new information on the specific kind of cancer, as well as what kinds of therapeutics might work best against the tumor. There have been significant successes in clinical precision medicine in cancer care, such as using existing cancer treatments in new ways, based on genotyping results.[10] In clinical precision medicine, there has also been some success in pharmacogenetics, a field that uses an individual's genetic sequence to tailor dosages of medications.[11] These kinds of successes have produced a groundswell of enthusiasm for clinical precision medicine. Many academic institutions have invested in precision medicine institutes, and states have developed consortiums with the hope that genomic analysis can lead to interventions that will greatly improve health care outcomes.[12]

Despite the significant advances in clinical precision medicine, and the potential for growth in this area, we decided to focus on precision medicine in the biomedical research sphere. Our goal is to understand how precision medicine research projects are developing, as these will shape the future directions of clinical precision medicine that will use insights from multiple forms of data. Building on what we learned from the research community, we define precision medicine as the effort to collect, integrate, and analyze multiple sources of data in order to develop individualized insights about health and disease. Precision medicine captures what is traditionally recognized as medical data, such as lab results, as well as other kinds of nonmedical data, such as air and housing quality measures. Our view of precision medicine and our focus on research in this area is in line with what Rothstein calls "big data health research."[13] We have chosen to focus on multiple forms of data because collecting and combining data from different fields poses many unique opportunities and challenges.

This definition of precision medicine also allows us to identify how the challenges posed by big data in other fields may or may not play out in biomedical research and ultimately in medical practice. Despite its name, "big data" is not significant

only because of the volume of data that can now be collected and analyzed by computers. As boyd and Crawford argue, big data is important because "[i]ts value comes from the patterns that can be derived by making connections between pieces of data, about an individual, about individuals in relation to others, about groups of people, or simply about the structure of information itself."[14] Precision medicine research specifically hinges on the ability of big data analysis to reveal relationships that can provide insights on the myriad causes of disease. Therefore, this conception of health data as being analyzable to show relationships between different variables is key for understanding how precision medicine is developing in the research space. This is also why precision medicine can be understood as a new and emerging field, despite a long history of scientists and clinicians using biological and nonbiological information to tailor treatment and care.

In the United States, there are a number of large-scale precision medicine research projects that are just beginning. Much of the work right now focuses on collecting data. Projects in this area include the NIH's All of Us Research Program which aims to collect data from one million volunteers in the U.S., the American Heart Association's My Research Legacy project, Alphabet Inc.'s Baseline study which aspires to create an "atlas" of human data, and New York University's Human project that will collect genetic, health, financial, and other kinds of data from ten thousand New Yorkers.[15] These studies aim to capitalize on the increasing availability of varieties and volumes of digital health data such as genetic sequences, digital health records data, prescription records, and data from digital medical devices and trackers.[16, 17]

BIAS AND DISCRIMINATION IN PRECISION MEDICINE

The goal of using multiple data sources and advances in computation and data storage to tailor medical care to individuals can be understood as part of the increasing emphasis on translating or applying biomedical research to address real-world problems. Precision medicine, at its core, is a positive effort focused on improving health and well-being, rather than limiting or diminishing health. Because precision medicine is suffused with benevolence and good intentions, it is all the more important that researchers, technologists, government actors, and patient advocates be mindful of how this emerging field could unintentionally introduce bias into the process of delivering tailored medical care. There must be an understanding of ways in which gathering and analyzing health data could have discriminatory effects as well as an assessment of the likelihood of these potential risks becoming actual harms.

INTRODUCTION

There is much excitement for the potential for these precision medicine research projects to increase our knowledge of health and disease, as well as improve the health care system, by shifting to prevention, designing better medical interventions, changing how information is delivered, and increasing efficiency. Because the outcomes of data-driven projects in health are universally recognized as positive – better health, a more efficient and effective health care system – these projects may not garner the same attention and concern as big data-focused efforts in other fields, such as predictive policing or data-driven risk profiling in criminal justice. As precision medicine has developed, mitigating risk, addressing bias, and achieving equity has been at the forefront of messaging around precision medicine research projects. Some of these projects have attempted to engage multidisciplinary researchers, practitioners, patient advocates, and patients in order to develop research that accounts for the complexity of tackling health problems and includes the voices of patients in the planning and implementation of these projects.

For example, all of the large-scale precision medicine research projects noted above are attempting to recruit diverse participant pools in order to address the historical lack of representation in medical research. The Privacy and Trust Principles that guide the NIH's All of Us Research Program include, among other tenets, a commitment to participant representation at all levels of the program and evaluation of the potential for the research to lead to stigmatization or social harm.[18] There has been an increased emphasis on involving participants and patients as active participants in medical research, and organizations such as the Patient Centered Outcomes Research Initiative (PCORI) have supported patient data sharing for precision medicine research.[19] Additionally, in light of recent cyberattacks on health data, precision medicine research projects have prioritized data security. Though important, keeping data private and secure will not assure that these data will not be misused.

Health data is subject to bias in multiple ways, and the increasing quantity and types of data that are available today can make it hard to identify where bias can emerge.[20] Bias has been defined generally as systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others.[21] In computer science literature, bias is defined in terms of the dataset—bias in labeling, bias in sample selection, bias in the task, or bias in model structure, favoring certain types of error over others. For example, bias in inputs – the building and analysis of datasets – can be the result of the sources of data, the context in which data is gathered, errors in what aspects of the data are considered important, and methods of analysis.[22] In medical and social science research, bias has been defined as any tendency that prevents unprejudiced consideration of a question or advances

prejudice in favor of or against one group compared with another.[23] The definitions of bias in both computer science and medical/social science research share an acknowledgement that bias implies error resulting in one group being favored over another. They differ in that the former refers to bias in terms of the dataset and the latter refers to bias as a certain outcome.

Because precision medicine brings together multiple data types, it also brings together medical research, social science, and computer science. These varying definitions of bias impact precision medicine, and these different types of bias interact in complicated ways. We define bias in precision medicine with these definitions in mind, acknowledging that bias can be introduced by human assumptions and then made invisible through automation or other technical processes.[24] Bias can also be introduced through different phases of data handling—from data collection, analysis, interpretation, and dissemination. Therefore, we think about bias as the ways that certain errors and outcomes, particularly discriminatory ones, might be favored due to human assumptions, obfuscation due to automation, and data handling. Bias at any point in data handling for precision medicine can lead to the recapitulation of longstanding health disparities. Furthermore, precision medicine is intended to be an iterative process—findings are translated to be implemented in practice, producing more information from which future insights can be derived. Thus, discriminatory actions in the clinical practice could feed biased data back into the system. In an ecosystem in which research can lead to evidence that is used to inform clinical guidelines and treatments, biases can have potentially life-threatening impacts.

The potential for bias and discrimination in precision medicine is a complex issue to tackle. The term was purposefully used vaguely in our interview protocol, allowing participants to use it in whatever sense they felt was appropriate. As such, this document is going to contain multiple competing definitions and understandings of bias. Our goal in the Fairness in Precision Medicine project is to identify the tensions in the emerging data-driven health research space, and the possibility for unintended consequences of the research process and discrimination that could result from implementing data-driven research insights in health. Our hope is that by bringing awareness to these tensions, we may productively shape conversations and interventions in precision medicine research going forward.

METHODS

To better understand how bias might impact precision medicine, we conducted a qualitative empirical study to identify the different tensions and frameworks with which stakeholders were working. The project was an exploratory, one-year study. Thus, we had to balance our desire to conduct a broad examination of how precision medicine research was developing nationally with conducting a more detailed investigation of the budding tensions and potential for unintended bias. To that end, we attended a number of conferences where precision medicine research was discussed, reviewed relevant academic and popular literature, press releases and other documents from big data health research projects, and mapped the stakeholders involved in the development of precision medicine research projects.

Through this exploration, we identified people working with health data and precision medicine research and were able to make personal connections with some at conferences or other professional venues. We built a list of individuals, identified patterns that arose in their roles and activities, and placed them in the following categories: biomedical researchers (in academia, government, and industry), bioethicists, technologists (software engineers and bioinformaticists), and patient advocates. Using these categories as a guide, we then planned to conduct semi-structured qualitative interviews with at least two to three people in each category to delve more deeply into issues of fairness and bias.

Though we knew that this would be a very low-risk study, we submitted our project for review by an Institutional Review Board (IRB), and it was deemed exempt. We designed an interview topic guide that included general questions about the participant's professional background, their current work, what kinds of health data they use or are familiar with, as well as their thoughts and opinions on the development of precision medicine research projects.[25] We specifically asked participants to tell us where they saw promise and opportunity as well as challenges in precision medicine. Our goal was to engage these experts and stakeholders to discuss how efforts to gather, analyze, and draw insights from digital health data could reinforce existing biases in medicine and in the health care system, and/or create new ones.

We conducted twenty-one interviews by phone or in-person between May and September 2017, varying in length from 45 to 90 minutes. In order to recruit participants, we used a snowball sampling of the researchers' professional networks. We spoke to actors in different fields who were formally connected with a precision medicine research study as well as those who were familiar with issues of health data collection, analysis, and implementation, but not formally affiliated with a particular study.[26]

These interviews yielded rich data about the specificities of analyzing different kinds of data, as well as the particularities of institutional and scientific culture and other dynamics that can impact precision medicine research in multiple ways. But our interview data do have limitations. As mentioned above, because of the timeline of the project, our aim was to interview a number of participants in each category, rather than continue to interview participants within the categories until we reached theoretical saturation. Therefore, the findings from our data are not meant to be representative of, for example, academic researchers' or technologists' views on precision medicine. Instead, the interview data present empirical, partial, and situated viewpoints on issues that are important broadly across the emerging field of precision medicine research.

In addition, we were only able to interview two patient advocates, and this perspective is not represented as robustly as we would have liked in our study. In addition, several projects are named in the reporting of our findings. Our goal is not to call out specific projects and pit them against each other, but to show where there are differences as well as common challenges. We asked participants about precision medicine in general, but the timing of the interviews coincided with increased news coverage of the launch of the national All of Us study, so this study was top of mind for many of our respondents, and thus likely more represented in their comments than if the interviews had been done at another time.

Last, but certainly not least, the assessment of potential risk in precision medicine is not an identification of actual harms. Our participants' concerns should not be interpreted as diagnoses of certain harms that will result as precision medicine research goes forward. These were semi-structured interviews, and the reflections are participants' perceptions of the development of precision medicine that are influenced by their relative positions within the precision medicine ecosystem and the amount of information they had access to at the time the interview took place. As these perceptions are discussed, we ask that readers lend a critical eye and keep the following questions in mind:

- Is there evidence that this risk will lead to harmful outcomes?
- What is currently being done to address these concerns?
- Are there actions being taken by researchers to mitigate these risks?
- Does the participant's statement reflect lack of intention or action on the part of precision medicine projects, or a point for which communication can be improved?

FINDINGS

During the interviews, we allowed the participants themselves to define what bias means when we asked them to discuss the promises and challenges facing precision medicine research, including the possibility for bias. Our respondents described where they see current biases in precision medicine, as well as the potential for future biases, and negative and discriminatory outcomes resulting from precision medicine research. Our participants discussed bias in two main ways: as 1) bias that could be part of the building and analyses of precision medicine research datasets, and 2) bias that could emerge as outcomes or results of precision medicine research.

BIAS IN PRECISION MEDICINE RESEARCH DATASETS

Participants discussed the importance of collecting multiple kinds of data for precision medicine research, but described numerous ways that historical processes have impacted the data available for use in precision medicine. We identified and grouped five areas where participants described sources of bias emerging in datasets: (i) genetic data, (ii) electronic health records, (iii) diversity in participants and data types, as well as (iv) historical and (v) analytic bias.

BIAS AND GENETIC DATA

Though our participants were aware that precision medicine research studies aim to collect and analyze multiple forms of data, some felt that talk of bringing multiple data sources to bear in this field was just that – talk – and that these research projects are, and will end up, predominately focusing on genetics. There would be a tendency towards using genetics in precision medicine research either because some studies are influenced by geneticists or because genetic data are more readily available. This would introduce bias because findings would be skewed in favor of locating the cause of disease in genetics rather than other potential factors. As computer scientist James H. Faghmous put it, genetic data are "low hanging fruit," and the methods of collecting and analyzing genetic data are more established than for other kinds of data (such as wearables data), or for analyzing multiple types of data together.

Although genetic data is available and analysis methods may be relatively more known, a respondent pointed out that there are still important differences in how genetic analyses are done. Sabrina Suckiel, a researcher and genetic counselor explained that there are differences embedded in genetic data, and this could impact precision medicine research:

"If you have a patient who has a condition that you don't know what the cause is, then you can order a whole genome sequencing test or a whole exome sequencing test. . . . The company is a clinical testing company like Ambry Genetics or GeneDx. . . . who will do the sequencing and the analysis. They'll give the clinicians back the report. . . . Then you see publications where people are classifying variants very differently amongst different groups. Right? One company might call it a likely pathogenic variant, another company might call it a variant of unknown significance. There's subjectivity to variant classification."

Sabrina Suckiel—Genetic Counselor

Suckiel explains that genetic data that could be in patients' records as clinical data may include differences in interpretation and analysis. Though some precision medicine studies may conduct their own genetic analyses, others may incorporate existing data, if participants have genetic data available. Precision medicine research projects that collect this genetic sequence data could be sweeping up these different genetic analyses without recognizing these embedded differences, and this could lead to bias or error when these data are analyzed en masse.

A number of other participants also stressed that focusing on the impact of genetic variation on disease does not reflect the contribution of biological, environmental, and social factors. Ethicists Arthur Caplan (Head of the Division of Bioethics at New York University School of Medicine) and James Tabery (Associate Professor of Philosophy at the University of Utah) commented on the limitations of focusing on genetics. Caplan argued that "[t]here's a danger of genetic reductionism, getting all whipped up about precision medicine and ignoring obvious environmental triggers and causes." In this vein, Camille Nebeker, Research Ethics Researcher, and others felt that precision medicine "has to be a multidisciplinary approach." Nebeker believed that this was possible "if the people advocating for it are knowledgeable about the value added by the intersection of multidimensional data." Expanding upon this point, she stated "I think if you were only communicating with people interested in genomic data, then that would drive the research question. If you bring in other people who are more interested in environmental or lifestyle data, that's going to drive different research questions. I firmly believe we can advance the concept of precision medicine by including those trained in the behavioral and social sciences." Tabery summed the concerns about focusing on genetic data, saying: "The broader worry here is that by putting it all in the genes, you assume the solution's also only coming from the genes."

As mentioned before, this emphasis on multiple sources of data helped shape how precision medicine is defined in this report. Yet, part of what Tabery and Nebeker highlighted is that, even when researchers know that different factors are important, they are constrained by their expertise and what data they have access to. As a result, their choices of what data to include influence the models that they can build. Bias through invisibility – such as lack of data on certain factors – can trigger discriminatory outcomes just as easily as explicitly problematic data.

BIAS AND ELECTRONIC HEALTH RECORDS

Many participants also asserted that electronic health record (EHR) data, despite its potential for enriching precision medicine research, could prove to be a problematic kind of data to collect and analyze. They pointed out the complexity and variation of EHR data, the mismatch between its intended purpose and the goals of precision medicine, and the challenges of identifying missing data. Jake Marcus, a software engineer who uses health data to build predictive models, argued that electronic health record data is complex and hard to interpret. For example, as Marcus explained:

> "Even looking at one patient's record [from an EHR], it's very complicated and might take some expertise to understand what's going on with the patient. They might have a hundred pages of notes and dozens of visits, diagnoses, and medications."
>
> **Jake Marcus**—Software Engineer

He explained that looking at an electronic health record is not the same as a complete narrative of a patient's clinical history. Though the record may contain a large amount of information, it may not be easy to translate into a story of "what is going on" with the patient. Prabhjot Singh, Director of the Arnhold Institute for Global Health, discussing the challenge of tackling complex datasets, said, "You have to know the problem space, the operating universe, the engineering issues, and then you have a process loop around all of them. You need a multidimensional team to do that, and then you need to be somebody that can at least do the descriptive work to tie them together." Marcus also discussed how individual labels within records, not just the health record as a whole, could be difficult to interpret:

"Bias through invisibility –
such as lack of data on certain factors –
can trigger discriminatory outcomes just as
easily as explicitly problematic data."

> "For mortality, that's a pretty clear label. Other labels are more
>   complicated. Take sepsis as an example. There's a lot of interest in
>   predicting septic shock. . . . My understanding is that that's more
>   complicated, and there's actually doctors that disagree about what
>   the appropriate definition [of sepsis] should be, and conflicting
>   literature. . . . There's a lot of iteration in figuring out and
>   defining the label."

Jake Marcus—Software Engineer

He explained that medical records have multiple fields, as well as multiple users (different clinicians who have added data to the record). Because disease labels, such as sepsis, are not clear cut, individual labels may be used to describe very different clinical realities. He even described how seemingly straightforward data labels such as date of admission in an electronic health record can contain differences because different hospitals start the stopwatch [for when you get admitted] at different times." Christy Collins, a patient advocate and President of the M-CM Network, echoed the sentiment that representation of disease in the EHR is not always reliable as she described the gaps in the record regarding her daughter's condition:

> "I think that there are some things that are not really in the medical
>   record. For genetic syndromes, there's this morphology information
>   or phenotype information. My daughter has characteristics that
>   sort of identify her as having the syndrome, but they're not medical
>   issues and wouldn't really appear in her medical record."

Christy Collins—Patient Advocate

EHRs contain potentially valuable information that can be mined to find patterns and predict health outcomes, as research has shown.[27, 28] As David Page, professor in the Department of Biostatistics and Medical Informatics and Department of Computer Sciences of the School of Medicine and Public Health at University of Wisconsin-Madison, noted, "Generally you can't be 100% certain unless you do some kind [of] controlled, randomized trial, but it turns out that with improved methods you can often get a pretty good prediction of many of these causal relationships just from purely observational data, like the EHR." However, Marcus' and Collins' comments make it clear that EHR data is not a source of clear, objective data on health and

medical care. Instead, these data are biased, in that they reflect the institutional and providers' preferences and differences. These biases themselves are not negative, but they can become problematic if they go unrecognized, especially in the context of large-scale precision medicine research studies.

Another interviewee familiar with health data informatics described that it is often hard or even impossible to combine data from different EHR platforms. This means that if a patient goes to different health care institutions that use different record platforms, it can be difficult to assimilate this data:

> "First of all, the patients have health care interventions in a variety of places. Multiple organizations, clinics, hospitals, emergency rooms, imaging centers, and labs have data on the patient. We want to be able to pull the data from all those sources to create a virtualized patient record. A lot of the electronic health record systems are proprietary and they have their own decision support logic of various types built into them. . . . There's a lot of entrenched vendor control over the marketplace and they are pretty actively resisting other sources of data, either other EHR's or other sources of data from patients being brought into their system. It may be quite a while until we have that change, but I think it's inevitable ultimately."

> **Robert Greenes**—Biomedical Informatics Researcher

These comments point to challenges of working with EHR data, including the differences in software and, importantly, the potential conflict between EHR vendors who want to keep their platforms and products "proprietary" and researchers who want these platforms to be open and able to process data, potentially from competing products. Grappling with these concerns, the NIH's All of Us Research Program will include the Sync for Science program that will allow study participants to bring in their health record data from multiple platforms. Efforts like this may provide impetus for the platforms to shift to increasing interoperability.

A number of participants also mentioned that despite the enthusiasm for collecting data from electronic health records to be used in precision medicine cohort studies, these records were not designed for research, but for billing purposes, which could be a source of systematic error and bias. For example, human-computer interaction scholars Pine and Mazmanian have described how clinicians often complete EHRs after performing care and retroactively fit the sequence of events to match the

options available in the software. The software is designed to build a record of care that can be translated into billing codes, and this sometimes diverges from the realities of clinical care.[29] Several of our respondents voiced the concern that there could be unintended consequences or biased analyses if precision medicine researchers fail to adequately recognize that much of EHR data is billing data, not clinical data.

This caution about making sure that researchers understand the provenance of EHR data also connects with the issue of "data empathy" that was raised by one of our respondents. James Faghmous, a computer scientist who works with large-scale health data, identified what he called a "lack of data empathy" felt by some technologists working with health data. He described this as a distance between these analysts and the data, specifically their lack of knowledge and direct experience of how, why, and where health data were collected. His comments emphasized the challenge of how data scientists, without health backgrounds, could build predictive models using health data that account for the realities of clinical care. This "lack of data empathy" can limit their ability to recognize bias and optimize the analyses because they are too far "from the source."

In addition to being complex sources of data that can have embedded biases, EHR data can be biased not just because of what it includes, but what is missing. That is, EHR data can be biased because it only represents people who have been entered into these records systems. Precision medicine research projects will have to contend not only with the biases within electronic health record data, but with the bias that can result from missing or uneven EHR data. A number of respondents raised the issue of bias via missing data in the EHR. For example, Lisa Parker, Professor of Human Genetics in the Graduate School of Public Health and Director of the University of Pittsburgh, said:

> "There are biases already in existence in terms of what data are available on which people who are already enrolled in say health care. Insofar as we still have a systematically describable group who are not in a health care system with data being collected upon them, from them, then that will be a source of bias. The problems that we've had historically of people lacking access to health care will continue to plague us for some time."

> **Lisa Parker**—Bioethicist

Parker's comments also point to an issue that several other participants mentioned when discussing EHR data: the US health care insurance system. They argued that the structure of the US health care system impacts the structure of EHR data, which can then make data collection for large-scale precision medicine projects more difficult. Collins, a patient advocate who was building a registry of people with her daughter's rare condition bemoaned the fact that EHR data was hard to bring together because of the nature of the US health care system, which she said "is so fragmented and expensive and poorly documented." In other words, EHR data is not just biased, but may be missing data from populations because the technology itself illustrates the characteristics and problems of the health care system. Stefan Zajic, a biomedical researcher, contrasted the US and the UK to explain how the health insurance system impacts EHR data, which can then impact research:

> "A great counterexample is a project called Genomics UK.
> In the United Kingdom they have a national health system, NHS.
> If a patient opts to have genetic sequencing done, that genetic
> sequencing data can be really easily linked to their medical health
> record data. Because of that, they have 100,000 people now
> that they've sequenced, and they have all of their longitudinal
> health record data for decades. It's an amazing dataset. The fact
> that we have such a fragmented health care system in this country
> presents a lot of limitations to research as well. It means that
> it's really difficult to do that kind of work. It's almost impossible
> to put together a single dataset for people that would really
> allow those kinds of questions to be answered."

> **Stefan Zajic**—Biomedical Researcher

Complex records, different meanings behind identified labels, data from multiple vendor platforms, and missing and fragmented data are all important sources of bias that could enter into precision medicine research datasets.

DIVERSITY IN PARTICIPANTS AND DATA TYPES

When asking people about bias in health data, one of the most commonly identified issues concerned representations of, and differences between, people of different sex, gender, socioeconomic status, and racial and ethnic backgrounds. In the United States, biomedical research has long failed to include representative samples of women and minority populations. When the data analyzed comes from narrow populations, findings may not be generalizable to all patient subgroups. Because of this long-standing problem, precision medicine research projects have placed a heavy emphasis on the need to recruit diverse patients to capture genetic and other kinds of diversity in order to avoid bias. Large-scale precision medicine research studies, including the All of Us Research Program, Project Baseline, and NYU's Human project, have made enrolling diverse participants a priority. For example, the All of Us Research Program has a chief engagement officer and the program recently provided funding to community partner organizations who will focus on engaging seniors, African Americans, Latinos, and the LGBTQ community.

Though precision medicine research projects acknowledge the importance of amassing representative and diverse cohorts, our participants added important nuances to the discussion of recruiting diverse cohorts by drawing our attention to historical biases, problematizing notions of racial targets, and emphasizing the importance of engagement. Participants mentioned the importance of enrolling a diverse cohort in precision medicine research studies, because without diversity, these studies could risk replicating the historical biases of medical research. Yet, it is also unclear whether using U.S. Census racial and ethnic categories as recruitment targets for precision medicine research will result in genetic diversity.[30] Those categories include a wide range of populations with different genetic ancestries, and it is possible that only a subset of that genetic variation will ultimately be represented, resulting in a biased sample for certain groups.

For tailoring interventions, census categories may oversimplify the complexities of ancestry, health, and intervention response.[31] As Shawneequa Callier, Associate Professor of Clinical Research and Leadership at George Washington School of Medicine and Health Sciences said, "While some providers may think that race provides some insight into a person's biology or lifestyle or environment, it's not enough. That's how we make mistakes, by thinking that we can judge a person's genetic makeup or likely protein pathways or decisions in life based on their skin color or how they self-identify." However, ignoring these historical categories disregards the reality of social contexts that have disadvantaged groups based on their racial categorization

rather than genomic variation. Our participants also stressed that varying sources of data would need to be used, but that harnessing these different sources would require tackling similar concerns regarding the potential for bias.

HISTORICAL BIAS

Karriem Watson, an academic researcher at the University of Chicago who specializes in recruiting minority and underserved populations for clinical research, discussed how bias can be embedded in health data, specifically noting how medical data can include hidden historical biases:

> "[Here are the] guidelines for lung cancer screening: you must be 55 to 80, and you must have 30 packs per year of smoking. This all came from a study that was done by the National Lung Screening Trial, one of the largest stud[ies] for lung cancer screening to date. 53,000 people. Of those 53,000 people, only four percent were African American. So, you mean to tell me that we now have lung cancer screening guidelines based on four percent of the African American population. . . and because you have such low minority enrollment in the study, you didn't look at the smoking habits of some urban communities. They don't smoke 30 packs a year. . . . They smoke menthols. You don't have to smoke a pack a day if you smoke menthols. . . . So now, when I do lung cancer screening in the community, most of the community members that we screen don't qualify because on the front end, the upstream research that led to this, wasn't precise for the people who carry the greater burden for the disease."
>
> **Karriem Watson**—Biomedical Researcher

This comment speaks to what Anna McCollister-Slipp, patient advocate, described as "consequences of really calcified outcomes measures and decisions and policy decisions based on these predetermined outcomes measures that were based on large, randomized control trials of people, which are inherently biased against outliers or smaller populations." Health data can have embedded historical biases that continue to impact the collection of other medical data as well as clinical care. As Watson explains, the original study that was used to develop the guidelines did not

include a representative sample of African Americans, and the resulting screening guidelines were biased against this group. Thus, data on who receives lung cancer screening is impacted by sample bias: African Americans may not be adequately represented in this data, not because they chose not to get screened, but because some may not have qualified for a screening. This bias has cascading effects, as those who are not screened have potentially important clinical data missing from their medical records which not only impacts their care, but recursively biases medical research since those biased screening data may be used in subsequent medical research on lung cancer.

This is an important example of how historical biases may be hidden in multiple ways, and it points to an important issue in precision medicine research. Because precision medicine research projects aim to collect new health data, as well as gather existing stores of health data like electronic medical records, it is important to recognize the potential limitations within these data today that come from historical legacies of bias and discrimination. Paul Glimcher, Director at NYU's Institute for the Interdisciplinary Study of Decision Making, and professor of Neural Science, Economics and Psychology, at New York University, argued that setting targets for recruitment based on Census race and ethnicity categories is not sufficient and could lead to bias against various subgroups in the dataset. He used Alphabet Inc.'s Project Baseline to illustrate:

> "Because Project Baseline was not designed as a statistically representative sample, it runs the risk of producing biased results for many ethnic, racial, and socioeconomic groups. For example, the project is recruiting in Palo Alto [California]. They're going to have an African American target percentage to assure African American representation in the study, but the African Americans that they are likely to recruit in Palo Alto are not likely to be representative of African American cohorts elsewhere in the nation. So, telling me about that group's health status, may not be telling me anything about typical African American health status. . . . [Project Baseline] runs the risk of doing that. They will get 30% African Americans. The question is, will they wind up with the 30% African Americans who are representative. . . . or will they wind up with this population that's very, very different? That would be okay as long as they acknowledge that. . . . but it would strongly limit the usefulness of their findings."

Paul Glimcher—Life Sciences Researcher

He brings up an important point, and team members at Baseline have been working to address this issue. Project Baseline is multiregional and is also recruiting participants in Los Angeles and North Carolina.[32] Regarding recruitment in Palo Alto, when we spoke with team members at Baseline, they emphasized that they are working with university partners to develop community engagement teams that will help them increase the diversity of their sample. They explicitly noted that although Verily is located in San Francisco, they are recruiting and including people from across the Bay area. They acknowledge that their study sample will not be perfect and that they understand the importance of including a diverse sample of participants. These comments show that people working in the precision medicine research industry understand the importance of diversity, but may think differently about diversity versus representativeness in terms of race, as well as other aspects like geography and socioeconomic status.

Lisa Parker, a bioethicist, voiced similar concerns, noting that oversampling of underrepresented populations will be crucial to the success of precision medicine research, but that a focus on Census race categories might be too crude and could miss important nuances and differences within these populations:

> "I think you have the bias that the early adopters are going to be people who are technology and health interested—patient interest groups, early adopters of technology and wearables, people who are scientifically inclined or interested in participating in this. What we really need to augment the socioeconomic educational background and continental ancestry related data are oversampling of other populations—people who have been omitted. Rural populations, urban poorer populations, people of color, people with continental ancestry that it is not Anglo-White European. It will be interesting to see how this plays out. I know that there supposedly are plans to try to over sample those populations, but we have no clear idea yet of what they're going to do given limited funds to follow a limited number of people and enroll them in this million-genomes, million-person database."

Lisa Parker—Bioethicist

In her comments, where she was specifically thinking of the All of Us Research Program, she made it clear that racial and ethnic inclusion in research is a problem that should be addressed, but she also raised the issue that differences within these

groups, such as "early adopters" or people with a predisposition, curiosity, or willingness to participate in these kinds of studies, could be overrepresented, and thus bias the sample of participants. She also suggested that researchers think about diversity not just in terms of racial groups, but continental ancestry (which does not neatly overlap with racial categories) as well as geographic and socioeconomic diversity.

In addition to citing the potential problems with recruiting participants to build a diverse cohort, and the number of ways that bias – in terms of representativeness – could creep into the cohort sample, participants also discussed the ways that precision medicine research could be broadly inclusive. Watson, for example, discussed engagement as an important, and often-underemphasized component of achieving diversity in medical research:

> "Another thing that I'm adamant about is academic institutions
> and research partners really understand the difference between
> recruitment and engagement. Engagement is where you can have
> those great, honest conversations about medical mistrust, and
> how we can design research to better include those populations
> that carry the greatest burden of disease, that's engagement.
> Recruitment is a study that already has a goal. I need you to
> participate. I need to be brought in every role from X, Y, Z. That's
> recruitment, and that's what we want to have conversations
> about – diverse participation – but we should be doing that much
> earlier. We should even be doing that at multiple touch points.
> Engagement means education and training for communities,
> to participate or to understand what health literacy looks like in
> order to participate."

> **Karriem Watson**—Biomedical Researcher

Because the prominent large-scale precision medicine research studies, such as the aforementioned All of Us Research Program and Project Baseline, use a cohort model, they must not only enroll people, but follow them over a number of years. Thus, Watson argues that it is important to not only recruit participants into precision medicine research studies, but to have them feel engaged and connected so that they will continue to participate longitudinally. He also brought up the importance of having researchers of color involved with precision medicine research. This kind of diversity in recruitment for scientific research is overlooked. He explained that diversity in the research staff is important because having researchers from similar racial and

socioeconomic backgrounds could imbue the studies with biases that come from a limited set of experiences, values, and perspectives. For Watson, a lack of meaningful engagement might not just impact the diversity of the cohort, but could also further alienate groups that might already harbor some earned mistrust of medical research. This means that a lack of engagement could lead to a lack of diversity within the study population, which could then further entrench existing skepticism among some individuals and communities of the value and utility of biomedical research.

ANALYTICAL BIAS

Participants discussed multiple ways that bias could emerge in the analysis of precision medicine data, including the potential for algorithmic bias and the lack of established analysis methods for new kinds of health data. One respondent argued that precision medicine research analyses could be biased by researcher preferences or by what is seen as an important and fundable line of research. This means that although data could be collected from large numbers of people, only particular subgroups or particular conditions may become the focus of analysis and further research.

Our interview with Jake Marcus, software engineer, yielded insights not just about the complexity of EHR data, but also about how problems can emerge in the analysis of health data. When he described his work, he explained that he builds models to predict health outcomes. But in order to build the model, the health data has to be standardized. One way of standardizing the data is to create labels for the health outcomes of interest. As discussed earlier, these labels may actually represent different information. So, in order to use this data to build models, he and other software engineers have worked with doctors to come to a consensus on label definitions. He acknowledged that clinicians working with other teams of computer scientists and engineers could come to different decisions.

This means that the labels and the resulting data analyses generated by these teams and used in the statistical models represent the preferences, experiences, and biases that these teams bring with them. It remains to be seen how the biases affect the resulting predictive models and algorithms, and further if these tools might have differential impacts on individuals or certain groups. However, Marcus' comments draw our attention to the possibility of algorithmic bias making its way into precision medicine research. There is growing attention to how algorithmic bias is present in fields such as policing, criminal justice, education, and child welfare. Much of this work focuses on how bias leads to discriminatory outcomes among minority populations.[33]

Though not a major issue now, the emphasis on collecting and analyzing large volumes of health data spurred on by precision medicine research means that the possibility for algorithmic bias should be seriously considered before it becomes a widespread problem.

However, it is also important to note that algorithms, machine learning, and other data analysis processes may not just be a source of bias in precision medicine, but may be used to detect bias in medical care. Health researchers have analyzed health data to reveal biases in end-of-life care, and the increasing availability of health data, along with advances in data analytics, may lead to increasing ability to identify patterns and predict which groups may receive different care protocols.[34] Work like this points to the opportunities for researchers to use the data gathered by precision medicine research and analytical methods to uncover and identify bias at institutional and systemic levels.

BIAS IN OUTCOMES

As James Tabery commented, "It's the very decisions you make about who you recruit for the study and what data you collect from them in turn that shapes the picture that you get about how the world works." In the previous section, we discussed the myriad ways bias can be produced in data collection and analysis. The following section examines bias in the "picture" that precision medicine produces: the outcomes of precision medicine for individual patients, specific groups, or the health care system overall. We discuss below how, even without biased data, the dissemination and implementation of precision medicine could have discriminatory and potentially negative effects.

DISCRIMINATION AND DISADVANTAGE

Many of our respondents discussed how the application of precision medicine research could disadvantage specific population groups, including individuals with lower health literacy or those in less resourced areas. Lisa Parker, bioethicist, discussed how medical interventions derived from precision medicine research could be used to discriminate against specific groups:

> "I do think if there are particular health risks that are associated
> with traditionally underrepresented, underserved, or discriminated-
> against populations, there's an opportunity or a likelihood that there
> will be exacerbation of that discrimination. If the health risks are
> found in [a] particular population that overlaps with immigrant
> groups that are currently not popular, for example, then we have
> a health-related reason to limit those immigrants coming to this
> country and presenting a health burden for our health care system. .
> . . It may not be realistic, but consider looking at social risks like air
> pollution: Might people who are at increased risk not be allowed
> to live in urban areas anymore if they're at increased risk for asthma

and lung disease due to air pollution? It may be assumed that we just won't let that happen. People would likely say, 'Wait a minute. You can't do that.' Well maybe we can impose such restrictions if these people are also on public funding for their housing. We could imagine making it a condition of their access to social services that we look at what people's health risks are and limit or at least shape people's choices, ostensibly for their own health-related good. Since we have precedent for that now with wellness programs in employer health insurance plans, it seems a short step to having government plans make similar sorts of wellness recommendations and incentives and so on."

Lisa Parker—Bioethicist

In this comment, Parker described how an identified "health risk" that could be developed using precision medicine research could be used to discriminate against persons or groups that are deemed at risk, and that people in already vulnerable social positions may be more vulnerable to this kind of targeting. She uses the case of employee wellness programs, which currently provide incentives, such as insurance discounts for employees to reduce their disease risk, as a potential first step toward penalizing people and groups for their health risks. Scholars such as Ifeoma Ajunwa have also cited the possibility of workplace wellness programs shifting from offering incentives to penalizing and discriminating.[35] It is important to note here that Parker discusses how the application of precision medicine research into a health risk could lead to bias and discrimination against certain groups, even if the data used to develop these insights themselves were unbiased or had been corrected for bias. There are legal statutes, such as the Genetic Nondiscrimination Information Act (GINA), that protect against discrimination based on genetic information and health data, but there are gaps in protection because they do not apply universally and life insurance companies, for example, are not covered by either of these policies. Therefore, our existing laws regarding health data and discrimination may not be sufficient to protect against discrimination that could result from precision medicine research. As articulated by Louise Bier, Director of Genetic Counseling and Clinical Engagement at Columbia University, a way to address concerns of further disadvantaging vulnerable groups would be to ensure that "the risks and the benefits [are] equally accessible and equally shared."

"The danger is that participants with lower health data literacy may not be able to take advantage of precision medicine research in the same ways as those with higher health data literacy."

Others mentioned that the application of precision medicine research and the dissemination of precision medicine interventions could impact vulnerable groups, not because they would be discriminated against, but because they may face limitations in understanding and using these interventions to improve their health outcomes. Bradford Hesse, Chief of the National Cancer Institute's Health Communication and Informatics Research Branch, explained how different levels of health literacy could impact the ultimate implementation of findings and treatments derived from precision medicine research:

> "In a world where quantitative literacy is sparse anyway, you know there are going to be people that if you present numbers the wrong way, they're going to feel left out and they're not going to be able to participate in their own health care. They're going to make bad decisions. They'll regret that later on. It could have bad outcomes later on and we're not going to get truly supported evidence-based medicine. So that's the worst case in my mind and in many cases it could do real, actual serious damage."

> **Bradford Hesse**—Bioinformatics Researcher

Here, the danger is that participants with lower health data literacy may not be able to take advantage of precision medicine research in the same ways as those with higher health data literacy. And, because there will be more emphasis on the individual taking responsibility for their health because they will have access to tailored health recommendations, an inability to understand these data or their implications could lead to negative health outcomes. In a similar vein, scholar and bioethicist Mark Rothstein has argued that precision medicine could widen the gap between the least and most sophisticated health care consumers, only benefiting the "uncommonly tech savvy, highly health literate, self-directed, information seeking, English fluent, health focused, and well insured."[36]

Chelsea Ratcliff, a researcher and doctoral student at the University of Utah, mentioned that the focus on targeting and tailoring to individuals could have unintended and negative impacts. People could resist recommendations or interventions that are too specific to them if they feel their autonomy or privacy is being threatened, or if they fear stigma or bias. Chelsea Ratcliff explained:

> "[This has] already played out in advertising and tailoring, especially when people get tailored messages that they're not expecting, so they

weren't aware that data was gathered and they feel kind of freaked out or spied on. Sometimes people also don't want one tailored recommendation, they want to be able to see the spectrum of options and decide for themselves. Health is also complex and some patients may not want to be singled out for sensitive or stigmatized health conditions or traits. So that's where I'm thinking about the potential for reactance and people feeling that their freedom is being limited in some precision medicine contexts."

Chelsea Ratcliff—Health Sciences Researcher

Ratcliff's comments draw our attention to how precision medicine, though the goal may be to improve health, may for some feel like an intrusion or a limit on their personal freedom and autonomy. Specifically, she brings up the point that participants could feel "spied on," even if they have given their consent for research and there have been no privacy breaches of their data. As Louise Bier noted, precision medicine research projects have heavily emphasized the importance of keeping sensitive health data private and "recognize how valuable and precious the information that people are sharing with us is." However, Ratcliff's comments show, as Helen Nissenbaum and others have argued, that privacy is contextual.[37] This means that although broad privacy principles are important, feelings of breached privacy or trust may still emerge.

INDIVIDUAL RESPONSIBILITY OVER STRUCTURAL INFLUENCES

Several of our interviewees expressed concerns about precision medicine implementation, specifically that interventions designed using precision medicine data could focus too much on the individual, rather than structural forces that shape health outcomes. The ultimate goal of precision medicine is to understand the myriad factors that contribute to an individual's specific disease risk and health outcomes. However, our respondents were concerned that future applications of precision medicine research could narrow the focus too sharply on the individual, which could have negative impacts. For example, Lisa Parker and James Faghmous argued that despite the large body of evidence on how social factors influence health, interventions targeted at individuals can be preferable because tackling social problems is so complex.

Arthur Caplan went further to argue that if precision medicine research results in more individual-level interventions, this will be "worse ethically because it starts to put the blame for disease solely on the individual. So, we look for obesity genes, but

we don't put in the sugary beverage restriction or wonder why we're having gigantic portions served at restaurants." Parker, Caplan, and other interviewees felt that medical interventions derived from precision medicine research could place too much responsibility on the individual for ensuring and managing health, and could downplay the structural and social determinants of health. Participants also expressed concern that health information derived from precision medicine research is not easily amenable to structural interventions because it is delivered and directed at individuals.

On the other hand, if environmental data is taken into account, precision medicine has the potential to illuminate the contextual factors that influence health. Callier pondered: "The more we learn about how an environment contributes to health outcomes, what are we going to do about it as a community, as a society? Precision medicine should really shine a light on the role that environment can play on different treatment outcomes and also risk. If we have a better understanding of the risks, then can't we intervene earlier? Can't we have more preventive measures?"

These comments also point to a concern voiced by several of our participants about a possible shift in accepted forms of biomedical research. An increasing emphasis on gathering large amounts of data, and analyzing these big data to find relationships among a number of factors to produce individual-level, tailored insights and treatments, complicates existing biomedical research methods, such as the randomized-controlled trial. Robert Greenes, a professor of biomedical informatics, explained:

> "Classically, when you were trying a new drug for hypertension or something, you'd take a randomly selected group of people with hypertension, maybe matched by age and gender. The control group would get the classical treatment or no treatment, and the test group would get the new treatment. You would see which did better. Within that, you'll always have some people that didn't do well. . . . With precision medicine, you now have more predictors, so you probably could start with a more refined subgroup. The disadvantage, of course, is as you keep refining, you have a smaller and smaller N [number of people in the study sample]. . . . That's the danger that you get with refined subgroups. . . . The more refining it, the less you're able to do the classical, physical trial."
>
> **Robert Greenes**—Biomedical Informatics Researcher

Arthur Caplan echoed the sentiment and stated that although precision medicine research projects today focus on gathering large numbers of participants, testing insights derived from this research may end up "undercutting what we traditionally used to think of as 'research,' meaning placebo control, double-blinded large number studies." This potential shift in biomedical research was concerning for some, like John Wilbanks, Chief Commons Officer at Sage Bionetworks and a Senior Fellow at FasterCures, who noted the challenges of navigating a "changing environment when the entire superstructure that surrounds you is literally designed to change slowly." Others, such as Hesse, were hopeful that the ability to analyze smaller substratifications of the population would allow for more inclusion of groups formerly excluded from medical research.

Although, as noted above, participants were concerned that an emphasis on individual responsibility would further marginalize, they also expressed hopefulness that patients could benefit from having more access to their own data—if delivered in ways that are interpretable and actionable. There are many positive reasons for giving patients access to their own data. Withholding that data because of a paternalistic view of patients' abilities might have the negative consequence of missing an opportunity to "not do harm" in precision medicine. Bradford Hesse described the story of Dave deBronkart, the founder of the "Gimme My Damn Data" campaign of who is known as an advocate of participatory medicine. deBronkart requested access to his health record and found a nearly incomprehensible mess of data.[38] Hesse explained that this is just one example of why he believes that patient access to data is an important aspect of precision medicine research.

CONCLUSION

The approach we took with our interviews encouraged participants to think broadly about precision medicine research as it exists today, as well as its future promises and risks. Thus, our findings reflect both the current state of precision medicine as well as problems that may arise in the future. Our interviewees were at times doubtful that precision medicine could truly live up to many of its promises. At the same time, they felt that the premise of precision medicine was ripe with potential. They raised many concerns regarding the potential for bias in precision medicine—ranging from the problems of engaging a diverse patient population to risks that emerge because of the power of the health care system. However, these concerns were mostly regarding barriers to achieving the goals of precision medicine rather than criticism of the goals of precision medicine itself. Harnessing big data in health care to achieve more precise medical treatment was not viewed as problematic—rather, the barriers to achieving it in its ideal form are seen as the problem.

The experts we interviewed framed various and complex concerns regarding bias in precision medicine. Data is subject to bias, and the increasing volume, variety, and velocity of data that is available today can make it hard to identify bias. And often this bias can lead to harms that can have uneven impacts on different segments of the population. However, with an understanding of the potential risks, we can begin to identify steps for preventing negative outcomes.

Challenges remain for achieving the vision of precision medicine. The Fairness in Precision Medicine project has identified a number of tensions that are just emerging as the field of digital, data-driven health research and medicine begins to take shape. As Sara Meeder stated "We need to move from that step to, this is what we're going to do to build the bridge to fix this." Most of those who are involved in or set the standards for precision medicine have the best of intentions, but as more stakeholders get involved and the incentives and visions evolve, there is a great potential for undesirable outcomes.

Bridging gaps in knowledge, methods, desired outcomes, goals, and guardrails is a critical but lofty endeavor. In particular, the ambiguity and differences in understanding about what the project of precision medicine is may undermine different stakeholders' ability to speak with one another. In order to achieve the dream of precision medicine, it is necessary to develop and evolve a framework for identifying the risk for bias in precision medicine research and implementation as this field evolves. Many early stakeholders – including those from the government who helped

define and publicize precision medicine as a scientific goal – recognize many of the challenges articulated here, yet many people in the field are struggling to articulate and structure a conversation about how bias should be recognized and how acknowledgment and assessment of these issues should be integrated into every aspect of precision medicine as it develops.

Much work still needs to be done. Research is needed to develop a comprehensive understanding of barriers and ensure that the structural determinants of health are acknowledged and used to plan intervention, rather than attributing responsibility solely within the individual. Policy and regulation is needed to ensure that the insights derived will not later be used to surveil and marginalize vulnerable populations. Above all, patients and the public need to be engaged in this endeavor both to guarantee its success and to hold other actors accountable for potential misuses or misunderstandings about when and how their data should be used.

The preventive medicine framework is prevalent in the rhetoric of precision medicine—if we are able to apply the right intervention at the right time, we can prevent larger health problems in the future. We believe that by engaging with a preventive mindset as we think through the challenges and risks of moving forward with precision medicine, we can intervene at a critical moment to avert negative outcomes. From incorporating genetic information to using data from electronic medical records, precision medicine has the potential to transform health care and medical research for the better.

ACKNOWLEDGMENTS

Designed by Default, Anisa Suthayalai

Data & Society is an independent 501(c)(3) nonprofit research institute focused on the social and cultural issues arising from data-centric technological development. **datasociety.net**

ENDNOTES

1   For example, black women have infant mortality rates that are parallel to women in so-called developing nations. MacDorman, Marian F., and T.J. Mathews. "Recent Trends in Infant Mortality in the United States." Data brief. National Center for Health Statistics, October 9, 2008. https://www.cdc.gov/nchs/products/databriefs/db09.htm.

2   J. Larry Jameson and Dan L. Longo, "Precision Medicine: Personalized, Problematic, and Promising," The New England Journal of Medicine; Boston 372, no. 23 (June 4, 2015): 2229–34; Richard Hodson, "Precision Medicine," Nature 537, no. 7619 (September 8, 2016): S49–S49, https://doi.org/10.1038/537S49a.

3   Partner Collaborations et al., "Precision Medicine: New Paradigms, Risks and Opportunities," Knowledge@Wharton, accessed November 8, 2017, http://knowledge.wharton.upenn.edu/article/precision-medicine-new-paradigms-risks-opportunities/.

4   "All About The Human Genome Project (HGP)," National Human Genome Research Institute (NHGRI), accessed November 29, 2017, https://www.genome.gov/10001772/All-About-The--Human-Genome-Project-HGP.

5   Nancy Krieger, "Stormy Weather: Race, Gene Expression, and the Science of Health Disparities," American Journal of Public Health 95, no. 12 (December 2005): 2155–60, https://doi.org/10.2105/AJPH.2005.067108.

6   Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (St. Martin's Press, 2018).

7   Jean Heller, "Syphilis Victims in U.S. Study Went Untreated for 40 Years," The New York Times, July 26, 1972, sec. Archives, https://www.nytimes.com/1972/07/26/archives/syphilis-victims-in-us-study-went-untreated-for-40-years-syphilis.html; Hideyo Noguchi, "Experimental Research in Syphilis: With Especial Reference to Spiroch[III]Eta Pallida (Treponema Pallidum)," Journal of the American Medical Association LVIII, no. 16 (April 20, 1912): 1163–72, https://doi.org/10.1001/jama.1912.04260040179001; Rebecca Skloot, The Immortal Life of Henrietta Lacks (Movie Tie-In Edition) (Broadway Books, 2017); C. P. Blacker, "'Eugenic' Experiments Conducted by the Nazis on Human Subjects," The Eugenics Review 44, no. 1 (April 1952): 9–19.

8   Nathaniel Comfort, "The Overhyping of Precision Medicine," The Atlantic, December 12, 2016, https://www.theatlantic.com/health/archive/2016/12/the-peril-of-overhyping-precision-medicine/510326/.

9   Jameson and Longo, "Precision Medicine: Personalized, Problematic, and Promising."

10   Seung Ho Shin, Ann M. Bode, and Zigang Dong, "Precision Medicine: The Foundation of Future Cancer Therapeutics," Npj Precision Oncology 1, no. 1 (December 2017): 12, https://doi.org/10.1038/s41698-017-0016-z.

11   Marylyn D. Ritchie, "The Success of Pharmacogenomics in Moving Genetic Association Studies from Bench to Bedside: Study Design and Implementation of Precision Medicine in the Post-GWAS Era," Human Genetics 131, no. 10 (2012): 1615–26, https://doi.org/10.1007/s00439-012-1221-z.

12   Examples include Cornell's Englander Institute for Precision Medicine, Duke University's Center for Applied Genomics and Precision Medicine, and the California Initiative to Advance Precision Medicine.

13   Mark A. Rothstein, "Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics," The Journal of Law, Medicine & Ethics 43, no. 2 (June 1, 2015): 425–29, https://doi.org/10.1111/jlme.12258.

14   danah boyd and Kate Crawford, "Six Provocations for Big Data," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, September 21, 2011), http://papers.ssrn.com/abstract=1926431.

15   The Precision Medicine Initiative (PMI) includes two parts: the Cancer Moonshot and the All of Us Research Program. All of Us is the cohort-building arm of the PMI, and aims to build a cohort of health data from one million people across the US. The American Heart Association's (AHA) Institute for Precision Cardiovascular Medicine is the only organization whose sole focus is precision medicine in cardiovascular care and sponsors the My Research Legacy Study in partnership with the Broad Institute of MIT and Harvard. Currently open, this is a national study that aims to collect health data from individuals who have had a cardiac event and will close in late 2017. Project Baseline describes itself as "the quest to collect comprehensive health data and use it as a map and compass, pointing the way to disease prevention." Verily has partnered with Duke University, Stanford Medicine

and Google to collect health data from 10,000 people for five years. The Baseline study will be longitudinal and will collect multiple forms of health data included self-reported and measured data, as well as sensor data. New York University's Human Project's goal was to recruit 10,000 New Yorkers to share their data as "representatives" of the varying communities of New York City. The project aimed to produce insights about the causes of Alzheimer's disease, the impact of public transit, education, nutrition, and housing on health, and insights about depression and mental health.

16   Collaborations et al., "Precision Medicine."

17   Jameson and Longo, "Precision Medicine: Personalized, Problematic, and Promising."

18   The White House, "Precision Medicine Initiative: Privacy and Trust Principles," November 9, 2015, https://obamawhitehouse.archives.gov/sites/default/files/ microsites/finalpmiprivacyandtrustprinciples.pdf.

19   "Advancing Research by Giving Patients Greater Access to Their Health Data," March 3, 2016. https://www.pcori.org/blog/advancing-research-giving-patients-greater-access-their-health-data.

20   Seeta Peña Gangadharan, danah boyd, and Solon Barocas, "Big Data: A Tool for Inclusion or Exclusion?" (Washington, DC: Federal Trade Commission, August 15, 2014), http://www.ftc.gov/system/files/documents/public_comments/2014/ 08/00023-92391.pdf.

21   All of Us Research Program, NIH VideoCast - Return of Genetic Results in the All Of Us Research Program (Day 1), 2017, https://videocast.nih.gov/summary.asp?live= 21883&bhcp=1.

22   "A Course in Machine Learning," accessed September 15, 2017, http://ciml.info/.

23   Christopher J. Pannucci and Edwin G. Wilkins, "Identifying and Avoiding Bias in Research," Plastic and Reconstructive Surgery 126, no. 2 (August 2010): 619–25, https://doi.org/10.1097/PRS.0b013e3181de24bc.

24   Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," Science 356, no. 6334 (April 14, 2017): 183–86, https://doi.org/10.1126/science.aal4230.

25   See Appendix 1 to view the interview protocol.

26   See Appendix 2 for participant names and affiliations.

27   Peter B. Jensen, Lars J. Jensen, and Søren Brunak, "Mining Electronic Health Records: Towards Better Research Applications and Clinical Care," Nature Reviews Genetics 13, no. 6 (June 2012): 395, https://doi.org/10.1038/nrg3208.

28   Riccardo Miotto et al., "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records," Scientific Reports 6 (May 17, 2016): 26094, https://doi.org/10.1038/srep26094; Maya Rotmensch et al., "Learning a Health Knowledge Graph from Electronic Medical Records," Scientific Reports 7, no. 1 (July 20, 2017): 5994, https://doi.org/10.1038/s41598-017-05778-z.

29   Kathleen Pine and Melissa Mazmanian, "Institutional Logics of the EMR and the Problem of 'perfect' but Inaccurate Accounts" (17th ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2014, Association for Computing Machinery, 2014), https://doi.org/10.1145/2531602.2531652.

30   Elizabeth G. Cohn et al., "Distributive Justice, Diversity, and Inclusion in Precision Medicine: What Will Success Look Like?," Genetics in Medicine, August 4, 2016, https://doi.org/10.1038/gim.2016.92.

31   Pine and Mazmanian, "Institutional Logics of the EMR and the Problem of 'perfect' but Inaccurate Accounts."

32   Verily, "Establishing a New Baseline of Health and Research Participation for Diverse Populations," Verily Blog (blog), December 12, 2017, https://blog.verily.com/2017/12/establishing-new-baseline-of-health-and.html.

33   Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.

34   See Perry, William R.G., Alvin C. Kwok, Christina Kozycki, and Leo A. Celi. "Disparities in End-of-Life Care: A Perspective and Review of Quality." Population Health Management 16, no. 2 (February 13, 2013): 71–73. https://doi.org/10.1089/pop.2012.0061, and Ghassemi, Marzyeh, Leo Anthony Celi, and David J. Stone. "State of the Art Review: The Data Revolution in Critical Care." Critical Care 19 (March 16, 2015): 118. https://doi.org/10.1186/s13054-015-0801-4.

35   Ifeoma Ajunwa, "Workplace Wellness Programs Could Be Putting Your Health Data at Risk," Harvard Business Review, January 19, 2017, https://hbr.org/2017/01/workplace-wellness-programs-could-be-putting-your-health-data-at-risk.

36   Mark A. Rothstein, "Structural Challenges of Precision Medicine: Currents in Contemporary Bioethics," Journal of Law, Medicine & Ethics 45, no. 2 (Summer 2017): 274–79, https://doi.org/10.1177/1073110517720655.

37   Helen Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life (Stanford, Calif: Stanford Law Books, 2009).

38   "Imagine Someone Had Been Managing Your Data, and Then You Looked." Blog: Society for Participatory Medicine, accessed November 10, 2017, https://participatorymedicine.org/epatients/2009/04/imagine-if-someone-had-been-managing-your-data-and-then-you-looked.html.

REFERENCES

"A Course in Machine Learning." Accessed September 15, 2017. http://ciml.info/.

Ajunwa, Ifeoma. "Workplace Wellness Programs Could Be Putting Your Health Data at Risk." Harvard Business Review, January 19, 2017. https://hbr.org/2017/01/workplace-wellness-programs-could-be-putting-your-health-data-at-risk.

"All About The Human Genome Project (HGP)." National Human Genome Research Institute (NHGRI). Accessed November 29, 2017. https://www.genome.gov/10001772/All-About-The--Human-Genome-Project-HGP.

All of Us Research Program. *NIH VideoCast - Return of Genetic Results in the All Of Us Research Program (Day 1),* March 6, 2017. https://videocast.nih.gov/summary.asp?live=21883&bhcp=1.

Blacker, C. P. "'Eugenic' Experiments Conducted by the Nazis on Human Subjects." *The Eugenics Review* 44, no. 1 (April 1952): 9–19.

boyd, danah, and Kate Crawford. "Six Provocations for Big Data." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, September 21, 2011. http://papers.ssrn.com/abstract=1926431.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." Science 356, no. 6334 (April 14, 2017): 183–86. https://doi.org/10.1126/science.aal4230.

Cohn, Elizabeth G., Gail E. Henderson, Paul S. Appelbaum, and for the Working Group on Representation and Inclusion in Precision Medicine Studies. "Distributive Justice, Diversity, and Inclusion in Precision Medicine: What Will Success Look Like?" Genetics in Medicine, August 4, 2016. https://doi.org/10.1038/gim.2016.92.

Collaborations, Partner, Europe, Global Focus, and North America. "Precision Medicine: New Paradigms, Risks and Opportunities." Knowledge@Wharton. Accessed November 8, 2017. http://knowledge.wharton.upenn.edu/article/precision-medicine-new-paradigms-risks-opportunities/.

Comfort, Nathaniel. "The Overhyping of Precision Medicine." *The Atlantic,* December 12, 2016. https://www.theatlantic.com/health/archive/2016/12/the-peril-of-overhyping-precision-medicine/510326/.

Eubanks, Virginia. Automating Inequality: *How High-Tech Tools Profile, Police, and Punish the Poor.* New York: St. Martin's Press, 2018.

Heller, Jean. "Syphilis Victims in U.S. Study Went Untreated for 40 Years." *The New York Times,* July 26, 1972, sec. Archives. https://www.nytimes.com/1972/07/26/archives/syphilis-victims-in-us-study-went-untreated-for-40-years-syphilis.html.

Hodson, Richard. "Precision Medicine." *Nature* 537, no. 7619 (September 8, 2016): S49–S49. https://doi.org/10.1038/537S49a.

"Imagine Someone Had Been Managing Your Data, and Then You Looked" Blog: Society for Participatory Medicine. Accessed November 10, 2017. https://participatorymedicine.org/epatients/2009/04/imagine-if-someone-had-been-managing-your-data-and-then-you-looked.html.

Jameson, J. Larry, and Dan L. Longo. "Precision Medicine: Personalized, Problematic, and Promising." *The New England Journal of Medicine; Boston* 372, no. 23 (June 4, 2015): 2229–34.

Jensen, Peter B., Lars J. Jensen, and Søren Brunak. "Mining Electronic Health Records: Towards Better Research Applications and Clinical Care." *Nature Reviews Genetics* 13, no. 6 (June 2012): 395. https://doi.org/10.1038/nrg3208.

Krieger, Nancy. "Stormy Weather: Race, Gene Expression, and the Science of Health Disparities." *American Journal of Public Health* 95, no. 12 (December 2005): 2155–60. https://doi.org/10.2105/AJPH.2005.067108.

Miotto, Riccardo, Li Li, Brian A. Kidd, and Joel T. Dudley. "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records." *Scientific Reports* 6 (May 17, 2016): 26094. https://doi.org/10.1038/srep26094.

Nissenbaum, Helen. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford: Stanford Law Books, 2009.

Noguchi, Hideyo. "Experimental Research in Syphilis: With Especial Reference to Spiroch[Ill] Eta Pallida (Treponema Pallidum)." *Journal of the American Medical Association* LVIII, no. 16 (April 20, 1912): 1163–72. https://doi.org/10.1001/jama.1912.04260040179001.

Pannucci, Christopher J., and Edwin G. Wilkins. "Identifying and Avoiding Bias in Research." *Plastic and Reconstructive Surgery* 126, no. 2 (August 2010): 619–25. https://doi.org/10.1097/PRS.0b013e3181de24bc.

Pine, Kathleen, and Melissa Mazmanian. "Institutional Logics of the EMR and the Problem of 'Perfect' but Inaccurate Accounts." Association for Computing Machinery, 2014. https://doi.org/10.1145/2531602.2531652.

Ritchie, Marylyn D. "The Success of Pharmacogenomics in Moving Genetic Association Studies from Bench to Bedside: Study Design and Implementation of Precision Medicine in the Post-GWAS Era." *Human Genetics* 131, no. 10 (2012): 1615–26. https://doi.org/10.1007/s00439-012-1221-z.

Rothstein, Mark A. "Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics." *The Journal of Law, Medicine & Ethics* 43, no. 2 (June 1, 2015): 425–29. https://doi.org/10.1111/jlme.12258.

"Structural Challenges of Precision Medicine: Currents in Contemporary Bioethics." *Journal of Law, Medicine & Ethics* 45, no. 2 (Summer 2017): 274–79. https://doi.org/10.1177/1073110517720655.

Rotmensch, Maya, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. "Learning a Health Knowledge Graph from Electronic Medical Records." *Scientific Reports* 7, no. 1 (July 20, 2017): 5994. https://doi.org/10.1038/s41598-017-05778-z.

Seeta Peña Gangadharan, danah boyd, and Solon Barocas. "Big Data: A Tool for Inclusion or Exclusion?" Washington, DC: Federal Trade Commission, August 15, 2014. http://www.ftc.gov/system/files/documents/public_comments/2014/08/00023-92391.pdf.

Shin, Seung Ho, Ann M. Bode, and Zigang Dong. "Precision Medicine: The Foundation of Future Cancer Therapeutics." *Npj Precision Oncology,* no. 1 (December 2017): 12. https://doi.org/10.1038/s41698-017-0016-z.

Skloot, Rebecca. *The Immortal Life of Henrietta Lacks (Movie Tie-In Edition)*. New York: Broadway Books, 2017.

The White House. "Precision Medicine Initiative: Privacy and Trust Principles," November 9, 2015. https://obamawhitehouse.archives.gov/sites/default/files/microsites/finalpmiprivacyandtrustprinciples.pdf.

Verily. "Establishing a New Baseline of Health and Research Participation for Diverse Populations." *Verily,* December 12, 2017. https://blog.verily.com/2017/12/establishing-new-baseline-of-health-and.html.

APPENDIX 1:
INTERVIEW PROTOCOL

- Can you tell us about your research?
    - What are the main issues you work on and what are the questions you're trying to answer through your work?
- (If specific study): What led you to pursue this study? (When did the study begin?)
- Were there other ways that you thought about framing your study?
    - How are you sampling for your study?
- Have you had any challenges reaching your target population?
- Are there projects similar to yours?
- Have you seen similar projects face challenges? What happened? Why did they stumble?
    - Do you think that your project is vulnerable to any of those problems? (probe for technical reasons as well as norms in workplace/field)
    - Do you think that there's a possibility that your project will be used in ways you don't intend? How?
- What are some of the benefits of your research?
- Is there anything that you worry about while doing your research?
- We've heard that people define precision medicine differently. How do you define precision medicine?
- Some people have mentioned that they find the "precision medicine" frame problematic. What are your thoughts on this?
- Who do you think are the major stakeholders in precision medicine? Why are they "major"?
- Who do you think is getting left out of precision medicine?
- Where do you see fractures in the so-called "precision medicine" ecosystem?
    - How do these tensions affect the research outcomes?
    - How do these tensions affect the collaboration possibilities?
    - How do these tensions impact the data collection and analysis?
- Which precision medicine projects are you familiar with?
- Do you think those projects have been/will be successful? Where do you think there are possibilities for those projects to stumble? (probe for difficulties collecting health data, risks to participants, surveilling participants, benefiting certain groups rather than others)
- Do you think that precision medicine can impact health disparities? Why or why not?
- What are the benefits (economic, clinical) to developing precision medicine (refer to field or specific projects mentioned)?
- Is there anything that I haven't asked about that I should have?

APPENDIX 2:
INTERVIEWEES

Louise Bier, MS, CGC: Louise Bier is the Director, Genetic Counseling and Clinical Engagement at Columbia University's Institute for Genomic Medicine. Bier was formerly affiliated with the Icahn School of Medicine at Mount Sinai Hospital in New York City.

Shawneequa Callier, MA, JD: Shawneequa Callier is an Associate Professor of Clinical Research and Leadership at George Washington University. Prior to joining GWU faculty, Professor Shawneequa Callier completed a postdoctoral fellowship at the Center for Genetic Research Ethics and Law, an interdisciplinary center for excellence funded by the National Human Genome Research Institute and located in the Bioethics Department of Case Western Reserve University's School of Medicine. From 2006 to 2009, Ms. Callier practiced health care law in Washington, D.C. Earlier in her career, she also interned at the World Health Organization and the Nuffield Council on Bioethics where she examined international health care ethics policies and human genetics laws and guidelines.

Arthur Caplan, PhD: Arthur Caplan is the Drs. William F. and Virginia Connolly Mitty Professor and founding head of the Division of Bioethics at New York University School of Medicine in New York City. He is also co-founder and Dean of Research of the NYU Sports and Society Program. Dr. Caplan currently serves as the ethics advisor to DOD/DARPA on synthetic biology, a member of the University of Pennsylvania's External Advisory Committee for its Orphan Disease Center, and a member of the Ethics and Ebola Working Group of the World Health Organization. Dr. Caplan also serves as the Chairperson of the Compassionate Use Advisory Committee (CompAC), an independent group of internationally recognized medical experts, bioethicists, and patient representatives which advises Janssen/J&J about requests for compassionate use of some of its investigational medicines. Dr. Caplan is also a regular commentator on bioethics and health care issues for WebMD/Medscape, for WGBH radio in Boston, and WMNF public radio in Tampa.

Christy Collins: Christy Collins is a mother and macrocephaly-capillary malformation (M-CM) patient advocate. Christy Collins founded and is the president of a rare disease advocacy organization called M-CM Network after her daughter was diagnosed with the condition. The founding objective was to create a longitudinal registry of advocacy-owned data and sample repositories so that these research assets can't be siloed in any one institution.

James H. Faghmous, PhD: James H. Faghmous is a visiting assistant professor of medicine at Stanford University. He develops novel machine learning and artificial intelligence methods to measure how social, environmental, and economic factors interact to create health disparities. In 2016, James was selected as an NIH health disparities fellow for his work on the topic. James received a PhD in computer science from the University of Minnesota where his dissertation on applying machine learning to global climate change problems was selected for the "Outstanding Dissertation Award in Physical Sciences and Engineering". James graduated magna cum laude from the City College of New York where he was a Rhodes and Gates scholar nominee. At the time he was interviewed, James was the founding CTO of Arnhold Global Health Institute at the Icahn School of Medicine at Mount Sinai in New York City where he launched the precision global health platform, ATLAS, with generous support from USAID and Gates Foundation.

Paul Glimcher, PhD: Paul Glimcher is Director at NYU's Institute for the Interdisciplinary Study of Decision Making, and professor of Neural Science, Economics and Psychology, at New York University. His postdoctoral training was in oculomotor physiology, researching the brainstem and mesencephalic nuclei that control eye rotations. Paul's laboratory has focused on the identification and characterization of signals that intervene between the neural processes that engage in sensory encoding and the neural processes that engage in movement generation. These are the signals which must, in principle, underlie decisionmaking.

Robert Greenes, MD, PhD: Dr. Greenes is a faculty member at Arizona State University. He joined ASU in September 2007 to lead the new Department of Biomedical Informatics (BMI). This unit, originally in the School of Computing and Informatics, in the Fulton School of Engineering, and for three years reporting directly to the Provost's Office, became part of the new College of Health Solutions in July 2012. After six years leading the Department, Dr. Greenes took a sabbatical for the 2013–14 year, to work on creating a collaboration initiative for interoperable health care apps, and returned in mid-2014 to ASU as Professor in BMI and to continue to pursue this initiative. He is also Professor of BMI at Mayo Clinic.

Bradford Hesse, PhD: Bradford (Brad) Hesse, PhD, was appointed Chief of the National Cancer Institute's (NCI) Health Communication and Informatics Research Branch (HCIRB) in November 2006. He served as the Acting Chief of HCIRB from 2004–06. Dr. Hesse's work focuses on bringing the power of health information technologies to bear on the problem of eliminating death and suffering from cancer. While at NCI,

he has championed several initiatives that evaluate and progress the science of cancer communication and informatics, including the Health Information National Trends Survey (HINTS) and the Centers of Excellence in Cancer Communication (CECCR). As director of NCI's biennial Health Information National Trends Survey (HINTS), Dr. Hesse leads a team of scientists in the development and execution of this nationally representative, general population survey of American adults.

Jake Marcus, MPH: Jake Marcus is a software engineer who works on the Google Brain team (a part of Google AI) applying machine learning to health care. He works on using EHR data to predict clinical outcomes. He builds models as well as the infrastructure to learn from the data.

Sara Meeder, MA: Sara Meeder was the Research Compliance Specialist for ISDM's flagship study, The Human Project, where she is responsible for all things regulatory. Her mission is to ensure that the participants in the project are protected and that the project itself is run within the context of research regulations and ethical standards. Sara has been involved in various areas of research in her career, with an emphasis on human subjects research, infrastructure, and ethics.

Anna McCollister-Slipp, MA: Anna McCollister-Slipp is Chief Advocate for Participatory Research for the Scripps Translational Science Institute (STSI). In addition, she is the founder of VitalCrowd, a Web-based collaborative platform aimed at crowdsourcing the design of health research and is the co-founder of Galileo Analytics, a visual data exploration and data analytics company focused on democratizing access to and understanding of complex health data. Anna seeks to build platforms for better understanding of and engagement with the needs of patients. She speaks frequently about the need for innovation in medical device data and technology, promoting data standards, device interoperability, and user platforms aimed at empowering patients to better manage their health.

Camille Nebeker, MS, EdD: Camille Nebeker is a faculty member at UC San Diego in the Department of Family Medicine and Public Health. Her work there is primarily to do research on research ethics. For the past 15 years, her research has focused on designing instruction to educate the public about science and the scientific method with a particular focus on community health workers who were assisting academic researchers to conduct studies in the Latino community. She, along with her bilingual/bicultural team, has developed and tested a course called "Building Research Integrity and Capacity" (BRIC) that is accessible to people who have little or no

formal scientific research training or academic research training. Nebeker also leads the Connected and Open Research Ethics (CORE) program to develop and crowd-source resources to help investigators design technology-enabled research studies that are ethical and responsible.

David Page, MD: Dr. David Page is a professor at the Department of Biostatistics and Medical Informatics and Department of Computer Sciences of the School of Medicine and Public Health at University of Wisconsin-Madison. Dr. Page works on algorithms for data mining and machine learning and their applications to biomedical data, especially de-identified electronic health records and high-throughput genetic and other molecular data.

Lisa Parker, PhD: Lisa Parker, a philosopher, is Professor of Human Genetics in the Graduate School of Public Health and Director of the University of Pittsburgh's Center for Bioethics & Health Law. Dr. Parker has published extensively on ethical concerns related to the design and conduct of research, particularly genetic research and mental health research, as well as on aesthetic surgery, confidentiality, and informed consent.

Chelsea Ratcliff, MA: Chelsea Ratcliff is a doctoral student at the University of Utah focusing on health, science, and risk communication. Her scholarly interests and current projects pertain to: (1) the use of heuristics in health decision making; (2) precision/personalized medicine communication; (3) psychological reactance and other forms of message resistance; (4) news coverage of health research, public use of, and trust in health journalism; and (5) public understanding of science. She recently published a review in Journal of Health Communication on the potential for patient resistance in precision medicine.

Prabhjot Singh, MD, PhD: Prabhjot Singh, MD, PhD, is Director of the Arnhold Institute for Global Health and Chair of the Department of Health System Design and Global Health at the Icahn School of Medicine at Mount Sinai, as well as Special Advisor for Strategy and Design at the Peterson Center for Health Care. Previously, Prabhjot was a professor of International and Public Affairs at Columbia University and Director of Systems Design at the Earth Institute. He completed a BA and BS at University of Rochester, an MD at Cornell and PhD in Neural & Genetic Systems at Rockefeller University, with a postdoctoral Fellowship in Sustainable Development at Columbia University. He completed residency in Internal Medicine at Mount Sinai Hospital. He is a Robert Wood Johnson Foundation Young Leader, a Truman National Security Fellow, and term member of the Council on Foreign Relations.

Sabrina Suckiel, MS, CGC: Sabrina Suckiel is a genetic counselor at Icahn School of Medicine at Mount Sinai Hospital in New York City. Her research projects have had the underlying rationale that there will be more of a demand for, and use of, genomic sequencing in health care as the cost of sequencing goes down. The first project aimed at the general public, the second at early adopters, and the third at health professionals. Her research projects include developing an educational pamphlet on whole genome sequencing, conducting a study of healthy individuals who undergo whole genome sequencing, and developing a tool to measure health professionals' knowledge of genomics.

James Tabery, PhD: Dr. James Tabery is an Associate Professor of Philosophy at the University of Utah. His research focuses largely on the philosophy of science and applied ethics, as well as the intersection between those domains. On the philosophy of science side, he investigates questions of causation and explanation in biology; while on the applied ethics side, he explores how the answers to those questions have ethical, legal, and social implications.

Karriem Watson, DHSc, MS, MPH: Karriem Watson is a Senior Research Scientist with the University of Illinois Cancer Center and the Director of Community Engaged Research for the UI Cancer Center at UIC and the Mile Square Health Center, a group of Federally Qualified Health Clinics (FQHCs) affiliated with the University of Illinois Hospital and Health Sciences System. Dr. Watson has a Doctorate in Health Science in Global Health, a Master of Science in Basic Medical Research, and a Master's in Public Health in Community Health Sciences. Dr. Watson's work has resulted in the creation of community-based cancer screening, prevention, and navigation programs for breast, lung, colorectal, cervical, and prostate cancer. His work to support community-based breast cancer screening and navigation afforded him recognition by the Metropolitan Chicago Breast Cancer Task Force as a "Community Champion." He is a faculty member at the UIC School of Public Health in the Division of Community Health Sciences and is an Adjunct Faculty at DePaul and Northwestern University. Dr. Watson is also the Core Co-Lead of the Community Engagement Core for a five-year NCI-funded multi-institutional grant with the Robert H. Lurie Comprehensive Cancer Center at Northwestern University, the University of Illinois Cancer Center, and Northeastern Illinois University to address cancer disparities in Chicago. He also serves as Co-Investigator for a four-year NCI-funded grant to develop a partnership with the UI Cancer Center and Governors State University to increase the number of faculty and students from underserved communities who engage in cancer disparities research. Dr. Watson is also a Co-Investigator of the Illinois Precision Medicine Consortium that

was funded by the NIH Precision Medicine Initiative, now called the All of Us Initiative and a newly awarded NIMHD Center of Excellence in Health Disparities at UIC.

John Wilbanks, BA: John Wilbanks is the Chief Commons Officer at Sage Bionetworks and a Senior Fellow at FasterCures. Wilbanks leads the Sage Governance team and serves as Co-Principal Investigator on Sage's award for the All of US Research Program. In conjunction with Academy Health and the Electronic Data Methods Forum, he co-developed novel visual consent processes for mobile clinical health studies that were integrated into Apple's ResearchKit open-source framework.

Stefan Zajic, PhD: Stefan Zajic, PhD, is a research scientist at Coriell focused on the Coriell Personalized Medicine Collaborative (CPMC), a research study examining the clinical utility of genetic information. Prior to joining Coriell, Stefan was a principal scientist in research and development with Merck & Co., contributing to quantitative pharmacology and pharmacometrics analyses and using mathematical modeling and simulation. Dr. Zajic's team has published extensively on the effects of genetic counseling in the CPMC, and continues to be interested in the psychological effects of receiving genetic risk information.