

Digital Health Data in a Million-Person Precision Medicine Initiative Cohort

A Workshop of the Precision Medicine Initiative Working Group
of the Advisory Committee to the NIH Director
May 28-29, 2015
Vanderbilt University, Nashville, Tennessee

EXECUTIVE SUMMARY

Background

On May 28-29, 2015, the [Precision Medicine Initiative \(PMI\) Working Group](#) of the Advisory Committee to the Director (ACD) of the National Institutes of Health (NIH) held a workshop at Vanderbilt University in Nashville, Tennessee, to discuss scientific and methodologic considerations for cohort design and collection of detailed health information in the development of a national research cohort of one million or more Americans. The workshop is part of a series of events and outreach activities that will inform a larger report on the design of the cohort.

The meeting featured panelists and speakers from many disciplines and sectors, including epidemiologic and health care delivery network cohorts, data integration, patient advocacy, and information technology (IT) as well as electronic health record (EHR) systems. There were two special guests: Senator Lamar Alexander (R-TN) and Representative Marsha Blackburn (R-TN). Workshop registration was open to the public. There was active discussion on Twitter using #PMINetwork. The workshop was videocast live and is archived at [the NIH videocast webpage](#).

Workshop Summary

Jeffrey Balser (Vanderbilt University) welcomed everyone to the meeting on behalf of Vanderbilt University. He highlighted how we are in an exciting time for biomedical research, with advances in health IT and genomics paving the way for the promise of precision medicine. Kathy Hudson (NIH) gave a meeting overview describing how this workshop fits in the larger planning steps for the PMI. She stressed that participants are anticipated to be centrally involved in the design and implementation of this initiative, and that the focus of the current workshops is on digital health data and on how to leverage existing cohorts and infrastructure.

Special Panel: The Promise of a National Research Cohort: Francis Collins (NIH) moderated a special session with U.S. Senator Lamar Alexander, Chairman of the Senate Committee on Health, Education, Labor and Pensions. Chairman Alexander noted his long standing interest in precision medicine and in improving EHRs. He stressed that Congress is working to create an environment that will allow for innovation in medicine, and wants input on what is needed for programs like the PMI to succeed. Jay Shendure (University of Washington) shared his enthusiasm that a large national cohort, incorporating next generation sequencing and other emerging technologies, will be transformative in changing the type of biomedical questions we can ask and answer. Bray Patrick-Lake (Duke University) stressed that patients want to be true

participants in the effort, and they want to end fragmentation of data and research efforts. Eric Dishman (Intel) highlighted that the scale and scope of the precision medicine cohort will push forward innovation and intellectual property development. Josh Denny (Vanderbilt University) noted success in the PMI will require multi-disciplinary teams including clinical informaticians and physicians with good understanding of quantitative sciences. Chairman Alexander concluded the discussion by encouraging the group to be visionary and aim for the top (as he noted, “There is more room up there”). He shared that it is important to have a blueprint with specific recommendations for implementation to guide the plan as it moves forward.

Overview of Public Input: Responses to the NIH Request for Information: Gina Wei (NIH National Heart, Lung, and Blood Institute) summarized the 152 public responses to the NIH Request For Information (RFI) on a large U.S. precision medicine cohort (<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-096.html>). The respondents largely agreed with the proposed size of one million participants, with 20% indicating the cohort should be even larger. There was a strong emphasis on the cohort reflecting population diversity with regard to many characteristics including age, ethnicity, socioeconomic status (SES), and gender. Suggested research topics that could be addressed included drug and non-pharmacologic treatment responses, discovery of risk and protective factors for a broad range of diseases (both common and rare), and implementation of precision medicine findings in clinical care. Respondents favored long-term follow-up of at least 5 years; leveraging existing studies and infrastructure, complemented by novel recruitment; and standardized collection of a core data set. The respondents described 58 unique research entities that could contribute to the PMI and provided ideas for research questions and study design, including allowing participants to have access to their own data if desired.

Health Care Delivery Network Cohorts: The first session, moderated by Rory Collins (Oxford University), focused on how existing health care network cohorts could be leveraged to realize the vision of the PMI. A. Keith Stewart (Mayo Clinic) gave an overview of the Mayo biobank, noting the average length of follow-up is 10 years. Marylyn Ritchie (Geisinger Clinic) discussed how Geisinger has established a large-scale research cohort study which includes active engagement of patients and practitioners, mechanisms to return information to study participants, and processes to incorporate genomic and EHR data. Geisinger researchers are beginning to bring in mobile devices and web portals as ways to record patient-reported outcomes. J. Michael Gaziano (Harvard Medical School; Department of Veterans Affairs [VA]) briefly described the VA’s Million Veteran Program, which has already enrolled close to 400,000 people. He noted that VA researchers considered three models: one like the Framingham Heart Study in which the study investigators actively collected all data, which was deemed unworkable at large scale; one like Geisinger with flexible permission to recontact subjects; and one involving purely passive collection of certain types of follow-up data. He recommended a model that collects core elements on all subjects, leverages opportunities for passive follow-up through electronic measures and then recontacts sub-populations to collect additional information as needed. Peggy Peissig (Marshfield Clinic) stated that Marshfield Clinic already has 15-30 years of electronic data on a large number of patients and that they are bringing in insurance data. Lawrence Kushi (Kaiser Permanente) stated that Kaiser has developed a multi-

institution common data model that makes it relatively easy to pool data and is in a strong position to contribute to activities like the PMI because of relatively low turnover among its members. Hakon Hakonarson (Children’s Hospital of Philadelphia [CHOP]) emphasized the value of studying pediatric populations. He described a program at CHOP that has brought in more than 75,000 children from 40 satellite systems, and includes 13 years of EHR data. Rajendu Srivastava (Intermountain Health System) noted that Intermountain has an active biorepository with over 4.3 million samples, along with decades of phenotypic data and linkages to EHRs and detailed pedigree data.

Key discussion points included:

- All agreed that health-system-based cohort studies could be leveraged to study exposures beyond genetics and genomics.
- A very large sample size, exceeding one million, would be necessary to enable cost-effective research on many different diseases and exposures. That is, by leveraging an enormous platform, it would be relatively easy to “find 5000 patients with this disease and 5000 patients with that disease.”
- All panelists said that they were interested in including their cohorts in a national unified effort as long as consent and privacy concerns are addressed.
- All panelists noted that by leveraging digital technologies and integrated health care governance, it is possible to screen, consent, enroll, and follow participants at low cost, on the order of approximately \$100 to \$250 per person.
- To keep the study efficient and cost-effective, the panelists stressed the importance of combining research activities with clinical processes (e.g. one blood draw to obtain clinical and research specimens) and of incorporating scalable electronic consent.

Research Cohorts: The second session, moderated by Sekar Kathiresan (Massachusetts General Hospital) included a discussion of the barriers and potential solutions to expanding the scope of existing research cohorts to include them in the PMI cohort. Jeffrey Olgin (University of California San Francisco) described activities in the Health eHeart Study. He noted that they use an electronic modular consent that allows them to network with other cohorts who use the same consent process. He pointed out that the obstacles to integrating with other electronic systems are more often bureaucratic rather than technical. Susan Gapstur (American Cancer Society) discussed advantages of linking to centralized resources, such as the National Death Index (NDI), as opposed to the state-run tumor registries. She cautioned that there is great variability in biospecimen collection in different studies, but an advantage of existing cohorts is they often have standardized, tight protocols that lead to higher quality. Will Dere (University of Utah) advised that it is essential to name the primary and secondary goals and objectives to guide what is collected and what research is done. These should focus on what we want to see in 1 year, 5 years and 10 years. Eric Boerwinkle (University of Texas, Houston) envisioned the PMI as a federated model with three sources – health care networks, population-based samples, and disease-focused groups. He noted that existing population cohorts have detailed phenotyping data that could seed the PMI and provide early results, and those phenotyping data could be used to validate e-phenotyping. He also pointed out that existing cohorts have ethnic diversity, but he is also worried about socio-economic diversity and cautioned against

designing a study of the “haves” in our have and have-not society. Danny Benjamin (Duke University) stressed the importance of including children and families, since so much adult disease evolves from childhood exposures and experiences. There will be challenges to doing studies that include a pediatric population, but there are also creative approaches to address these challenges.

Key discussion points included:

- All panelists expressed interest in having their cohorts involved, but noted that re-consent would be required and that there were concerns about existing restrictions and privacy.
- For many existing cohorts, the current language states that participants will not have data returned to them, so that would have to be changed for participation in the PMI.
- Many of the current cohorts have a narrow age-structure, with a focus on older adults.
- Legislative action will likely be needed in several areas including access to registries, mechanisms for consent, speed with which information is reported to the NDI and other registries, and the development of a national database where insurers can provide standardized information.
- The cohort should use a contemporary epidemiology approach to recruitment and retention, including social media.
- Stakeholder engagement and trust are important for developing and maintaining cohorts.

Ensuring Inclusion: The third session, moderated by Esteban Burchard (University of California, San Francisco) focused on diversity and inclusion in the PMI cohort. Dr. Burchard started with a brief overview of the scientific advantages of clinical and biomedical research in diverse populations. All the panelists agreed it will be important to be inclusive, and to consider oversampling particular populations since scientific evidence shows that population groups defined by race, ethnicity, socioeconomic status (SES), health literacy and other factors differ in their patterns of morbidity and mortality. William Blot (Vanderbilt University) shared the example that mortality rates differ 2 fold between the lowest vs. highest SES categories even when looking at generally low income populations recruited from southern community health centers. Jennifer DeVoe (OCHIN) noted that you often need to budget more resources to recruit, engage and retain under-represented populations. Herman Taylor (Morehouse School of Medicine) pointed out that if you consider America in 2050 the current minority populations will compose the majority of the US population. Sonia Anand (McMaster University) highlighted that the goals of the project should be informed by the communities included in the project. All panelists agreed that community engagement and trust will be essential for the success of the initiative. The panelists each shared examples of the importance of working with local communities and having community members as valued equal partners in leadership and decision-making positions.

Key discussion points included:

- Specifics on which populations to oversample will depend on the scientific objectives of the PMI. Power calculations and other data could inform the optimal study design and sample strategy to ensure sufficient numbers for specific scientific questions.
- There is large heterogeneity within racial/ethnic categories. For example African Americans and Asian Americans are each diverse groups, with distinct countries of origins, practices and beliefs. Inappropriately combining these subgroups may obscure detection of unique health and disease patterns and clues to optimal treatment. It may also hinder potentially important studies of intragroup differences.
- Keeping people engaged in longitudinal studies is a challenge. Good retention arises from good recruitment, and it is important to set appropriate expectations from the beginning. Offering modest incentives can be useful, but they need to be incentives that are meaningful to the specific population.
- The group stressed the importance of inclusion in terms of SES, with a focus on social determinants of health. The PMI cohort should take into account inclusion on factors like income, health literacy, and access to care.
- Achieving good statistical precision for many subgroups will require an extraordinarily large cohort (more than the proposed one million). This will have serious cost and resource implications. Disease organizations may help with recruitment for rare diseases, but balance is needed between many small versus one, or a few, large cohorts.

Cross-Institutional Digital Health Data Sharing for Research: The fourth session, moderated by Joshua Denny (Vanderbilt University), highlighted issues that need to be considered for successful data sharing. Mark Frisse (Vanderbilt University) pointed out that interoperability is not the only challenge. Barriers to data sharing may also be political, behavioral, or financial. For the PMI to succeed, it must leverage current efforts in health care delivery and create a community of trust with clear, consistent policies. Sachin Kheterpal (University of Michigan) emphasized that the Centers for Medicare & Medicaid Services (CMS) and Centers for Disease Control and Prevention (CDC) need to lead by example in creating a national record linkage infrastructure. He also stated that payer engagement is crucial. For payers and providers, near-term clinical practice improvement is a top priority; these stakeholders need to understand how the PMI can help them achieve their care improvement goals. Abel Kho (Northwestern University) explained that the data in EHRs primarily relates to clinical care given to patients. EHRs don't capture experiences that occur outside the care system or the experiences of people who do not obtain care. Russ Waitman (University of Kansas Medical Center) explained that health care organizations, as highly regulated service providers, have a culture different from that of research organizations. Time and transparency are needed to enable the two types of organizations to work together.

Key discussion points included:

- Data collection might be streamlined by having participants retrieve their own data and then share it with researchers.
- Combining EHR data with geographic data would greatly enhance the value of the PMI as a research resource, but it also increases the risk of making data re-identifiable.

- Even within a single geographic area, people may obtain care from multiple sources. Record linkage is crucial to capture their full clinical experience. However, statistical record linkage methods that are not dependent upon a specific individual identifier have improved sufficiently that the need for an explicit national record linkage infrastructure is uncertain.
- Data is being shared successfully within specialty communities where people know and trust each other, but the PMI may face a very different situation because of its wide variety of stakeholders. Building trust is essential, and so is having policies that facilitate data sharing rather than creating obstacles to it.

White House and Office of the National Coordinator Vision for Future U.S. Health Data Infrastructure: The fifth session was moderated by Pearl O’ Rourke (Partners Health Care System, Inc.), featured presentations by Karen DeSalvo (Office of the National Coordinator for Health Information Technology [ONC]) and Mina Hsiang (The White House). Dr. DeSalvo explained that her office plays a leadership role in the Nation’s efforts to adopt and meaningfully use health information technology (HIT). The ONC’s early work, after its creation in 2004, focused on the uptake of HIT and adoption of electronic health records (EHRs). More recently, interoperability has been a top priority. The ONC is using mechanisms such as its certification program to incentivize payers and providers to behave in ways that enhance interoperability. The world of medicine needs policies and technologies that give consumers the type of health record they want—a longitudinal record that covers all their care experiences and other aspects of their health throughout their lives, no matter where they move. The ONC wants to facilitate the PMI and work with it to help create a learning health care system. Mina Hsiang asked attendees to tell her how the White House can help enable the PMI. She asked for feedback on the pieces of policy and technology infrastructure that are needed to facilitate the PMI and on the types of stories that should be shared with the public to maintain excitement about it. She encouraged attendees to find ways to show interim results and keep stakeholders engaged.

Key discussion points included:

- Achieving interoperability quickly is important. Many patients do not have years to wait.
- Patients, providers, and payers may be more willing to participate if: a) rules are changed to enable community participation in the oversight of Federal databases, b) firewalls are created between Federal health databases and those of other branches of government, and c) effective penalties to disincentivize the misuse of information are in place.
- Over a period of years, the PMI will generate a staggering amount of data—perhaps 20 terabytes per patient. It is important to develop a scalable infrastructure that can cope with the data.
- The new Fast Healthcare Interoperability Resources (FHIR) standard may enable better, faster, and less expensive data sharing.
- The White House can play an important role by helping to create and maintain public awareness of the PMI and encouraging participation in it.

Health IT Developer Solutions for Data Sharing Across Institutions: In the sixth session, moderated by Eric Dishman (Intel), a panel of health IT developers provided their perspectives on challenges in the design and implementation of the PMI. Paul Bleicher (Optum Labs) observed that policy issues, rather than technical ones, are the greatest challenge. The focus should be on how to bring data together and standardize and normalize it to provide useful information, rather than on interoperability between individual organizations. Scott Moss (Epic) said that health IT can bring many things to the table besides application programming interfaces (APIs) and access to data. For example, HIT can provide tools for consenting, recruitment, patient engagement, and collection of patient-reported outcomes. In addition, it provides an infrastructure for guiding precision care via decision support applications. David McCallie, Jr. (Cerner Corporation) explained that ways of getting data into the hands of researchers are constantly evolving. The traditional approach, in which data from EHRs is fed into a data warehouse, where it is normalized for research, may soon be replaced by more efficient methods. Eric Just (Health Catalyst) pointed out that although a major challenge is indeed interoperability, another underlying problem is poor-quality data. Improving EHR data quality involves showing value to the clinicians – that is, to make it worthwhile to them to improve quality of data entry by for example, using the data as quality improvement projects to enable better patient care. Panelists also emphasized that the PMI needs a focused plan; HIT should be employed in ways that align with the Initiative’s goals.

Key discussion points included:

- Because data generated for patient care is not necessarily “research ready” in its native forms, data must be curated.
- The vendor community has rallied around FHIR, which may allow easier interoperability.
- Smartphones should not be rejected as a data collection tool out of concern that low-income people or members of minority groups would not have them. The proportion of the population with smartphones is surprisingly similar among racial and ethnic groups and among people at different income levels.
- The time that health care providers spend entering data into EHRs is a precious resource. Entering structured data adds to provider burden, so it should only be required when it is truly necessary. Providers need to understand why structured data is sometimes needed, and incentives should be put in place to encourage their cooperation.
- The PMI needs to consider how to best present data to clinicians and patients in usable forms.

Summary of Day 1: Kathy Hudson (NIH) used tweets to summarize the agenda and topics covered in day 1. She noted that day 1 saw 970 tweets, 19.1 million Twitter impressions, 281 Twitter participants and 364 people watching via videocast. Day 1 covered the risks and benefits of leveraging existing resources, the challenges and opportunities in data sharing and the importance of inclusion of underrepresented populations and children.

Introduction to Day 2: Kathy Hudson (NIH) reviewed the agenda for day 2. Francis Collins (NIH) again stressed that the PMI will transform our understanding of human health and disease, and will provide a platform for answering questions we would not otherwise be able to answer. He then introduced Representative Marsha Blackburn, highlighting her role as the Vice Chairman of the Energy and Commerce Committee, which has formulated the 21st Century Cures Act. Representative Blackburn shared her excitement about 21st Century Cures, noting it passed unanimously out of House committee and she is optimistic it will be signed into law. In developing the legislation she and Chairman Fred Upton (R-MI) focused on high-cost areas, such as Alzheimer’s disease, and the current barriers to entry for research. The goal is take a “cures” strategy focused on wellness, disease management and quality of life outcomes. The legislation will enable CMS, NIH, the Food and Drug Administration (FDA) and CDC to use novel technologies and delivery systems, and will offset costs to incentivize the process in a revenue neutral manner. She described aspects of bill, including the Software Act, reauthorization of the Reagan-Udall Foundation; and the Children Count Act. Discussion highlighted how to define health information, the value of basic science, and how policy and legislation can foster an environment for innovation for mHealth technology and other biomedical advances.

Core Data Elements and Where They Reside: The seventh panel, moderated by Robert Califf [FDA] focused on the key data elements to collect in a national cohort. Dr. Califf began by noting the key innovations that will facilitate the PMI cohort—participant involvement, revolution in EHRs, and advances in genomic and mobile technologies—and asked each panelist to identify points to bear in mind in designing the study. Herman Taylor (Morehouse School of Medicine) highlighted the importance of contextual information such as family and neighborhood environment in understanding the impact of biologic factors—noting that EHRs alone will miss much in these contextual realms. He also emphasized the value of examining within-group heterogeneity rather than just comparing one group to another. Bradley Malin (Vanderbilt University) noted that if the goal is precision “medicine” alone then focusing on EHRs makes sense, but if the goal is precision *health* more data sources will be needed, many of which are not curated. He also noted the importance of determining the levels of information we want to make available, and to whom, and the need for involvement of data scientists from a variety of disciplines. Decisions will be needed up front about centralized vs. federated data management approaches because standardization alone does not solve harmonization problems; people may use the same terms but mean very different things. Thomas Glass (Johns Hopkins University) noted that the term “cohort” has a very specific meaning to epidemiologists, and describing this effort as a “cohort” could backfire as it did in the National Children’s Study, where a classical cohort design was not feasible. Rory Collins (Oxford University) noted that the UK Biobank has been described as a “resource” that facilitates a variety of types of research, many of them unforeseeable at present, while a “cohort” risks being viewed as all things to all people, and satisfying none.

Key discussion points included:

- Data elements acquired from participants at entry are relatively straightforward, including questions, physical measures, samples, links to other data, and should be targeted to data types that will be most useful to those ultimately using the resource.

- Assessing outcomes is harder as they typically happen outside established data collection streams, and are best acquired from existing information systems such as EHRs.
- There should be a small curated core dataset with a high threshold and specific criteria for adding to it, such as data “perishability” if not collected in real time.
- The PMI’s emphasis on medicine and health should encourage focus on impact of treatments, not just causal inferences. Data related to interventions should have priority.
- Although identifying one over-arching research question for the PMI is not possible, short-term deliverable goals are needed that demonstrate the initiative’s value. One suggestion is a crowd-sourced contest to produce 5 to 10 top research uses.
- Users should be required to enrich the resource by cleaning or refining the data or returning added measures to the resource with adequate documentation for other users.
- Standardized follow-up will be costly, but if web- and IT-enabled participant engagement can be maintained it should be possible to minimize the need for direct contact.
- Risk of re-identification can be quantified based on what is released and to whom; policies should be established to hold users accountable for misuse of the data.
- The four “C’s” of data collection—Collect, Catalog, Clean, Curate—should all have defined processes, timelines, and leaders. “Defer” is a critical fifth concept. It is best to do only what’s needed at the time and defer tasks like sample analyses until needed, as costs come down and methods improve.

Possibilities for Direct-from-Participant Data: The eighth session, moderated by Shiriki Kumanyika (University of Pennsylvania), explored issues related to recruitment, retention, empowerment, and engagement of study participants. Kathy Giusti (Multiple Myeloma Research Foundation) explained that patients, particularly those with rare or challenging diseases, want to help with research. However, they may not understand basic research concepts or even be able to use a password to access a patient portal. Adam Amdur (American Sleep Apnea Association) said that educating patients is often emphasized, but it is also important to educate researchers on how to interact with patients. He explained that study participants with health problems do not necessarily participate in research for altruistic reasons. They often do it for themselves and their families, and they want to see research results that are relevant to them. Christopher Burrow (Humetrix) emphasized the importance of the Blue Button initiative, which enables millions of Americans to have easy access to their own EHR data. With appropriate consents and other protections, patients could obtain their data through Blue Button and contribute it to a research project. John Wilbanks (Sage Bionetworks) discussed the challenges in keeping participants engaged in a long-term research initiative. After the initial excitement dies down, what matters most is that participants know that their data is contributing to scientific knowledge. Sharing results with participants at regular intervals is crucial to retention, but different participants prefer to receive different amounts of information.

Key discussion points included:

- Smartphones establish a longitudinal link with study participants, and they can be used for a wide variety of types of data collection.
- True engagement of study participants includes inviting their feedback and, when appropriate, changing study protocols in response to their comments.
- Enrollment is most successful when it is easy for people to join and participate, when incentives are offered, and when participants are proud to be in the study.
- If study participants are expected to perform tasks that involve electronic devices, it is important to use interface designs that are appropriate for them.
- What upsets study participants the most is not seeing their data used; conversely, communicating with participants that their data has been used is a strong positive reinforcement for continued engagement.

Perspectives on Centralized and Federated Data Models: In the ninth session, moderated by Joshua Denny (Vanderbilt University), panelists discussed the balance between centralized and federated management of data. Both Jeff Brown (Harvard Pilgrim Health Care Institute/Harvard Medical School) and J. Marc Overhage (Siemens Health Services and Indiana statewide Health Information Exchange) urged that the PMI avoid federation as much as possible. Centralized data management is more efficient. A curated core dataset should be created as a centralized resource, but original data could be left at the sites where it originated. John Walsh (COPD Foundation) said that a hybrid model would be best if the PMI plans to take advantage of existing networks and datasets, such as those of Kaiser Permanente, Intermountain Healthcare, and other large health care systems, and those of the VA and the network being created by PCORI. These groups and their participants may not be willing to give up all their data to a central organization, so some degree of federation would be needed.

Key discussion points included:

- Curation of data is essential, but it is important to retain the original data as well because it may need to be reanalyzed in different ways.
- Data should be collected up front; it may not be possible to collect the same data later. However, subsequent steps in the handling of the data—cataloging, cleaning, and curation—can be deferred until necessary, based on research priorities.
- Health Information Exchanges have adopted ‘centralized federated’ architectures where datasets contributed by multiple organizations are kept separated but are maintained in a single central computing resource.
- Biological specimens are finite, so they need to be managed carefully.
- Because precision medicine is being pursued globally, the PMI should consider federation with research initiatives in other countries.

Summary and Next Steps: Kathy Hudson (NIH) noted that a new request for information has been published <http://www.nih.gov/precisionmedicine/rfi-announcement-06022015.htm>, with comments due by June 19, 2015. They NIH will also look into soliciting input on renaming the

cohort and developing research questions. Francis Collins (NIH) wrapped up the meeting, stating that the workshop has moved the needle towards greater specificity. He encouraged people to share suggestions for “early wins” that will help maintain momentum. He noted that the important issues on inclusion raised at this meeting will be elaborated upon at the next PMI workshop on July 1–2, 2015. He thanked the panelists and ACD Working Group for their valuable input, and the staff at Vanderbilt and the NIH for their work in organizing the meeting.

Recurring Points Made at the Workshop

The PMI Needs to Specify Key Research Questions to Guide the Effort: This should include both long-term goals and near-term goals that are meaningful to stakeholders. Short-term successes can help build and maintain the momentum of a long-term project.

The PMI Should Seriously Consider Including Children: Many adult diseases start in childhood or are influenced by childhood experiences. Having children in the PMI poses special challenges, but the scientific value of childhood data is immense.

Ensuring Inclusion Is Important, but Goals Need to Be Clarified: Is the goal to create a cohort that is representative of the U.S. population? Or is the goal to include enough members of various subgroups so that meaningful conclusions can be reached about each group? These approaches require different sampling and recruitment strategies. Special approaches and messaging are needed to recruit and retain underserved populations.

International Collaboration Will Increase the Value of the PMI: Some health systems and cohort studies that may become part of the PMI are already collaborating with their counterparts in other countries, and precision medicine is of worldwide interest.

The PMI Should Take Advantage of Lessons Learned in Other Projects: There is much to be learned from the wisdom—and the misfires—of other studies, cohorts, and networks.

The PMI Should Consider Hybrid Models: Cost efficiencies can be gained by gathering core elements on participants recruited by different mechanisms, and then identifying key subgroups in which to go deep in terms of phenotyping and data collection. However, the PMI cannot be all things for all people.

Challenges to and Opportunities for Data Sharing Should be Addressed Sufficiently: Overcoming barriers to data sharing involves not just providing health IT and other technological solutions, but also building trust and developing policies that facilitate data sharing at multiple levels.