



All of Us National Launch Event. Sunday, May 6th, 2018. Image courtesy of the National Institutes of Health. Photography by Rob Karlic.

*By Makda Zewde*

## From All of Us to Precision Medicine

In his 2015 State of the Union address, former president Barack Obama announced the launch of the \$215 million dollar Precision Medicine Initiative, a national effort to accelerate biomedical discovery that can better tailor healthcare to the needs of individual patients. This emerging approach to medicine has already shown immense potential in the treatment of certain diseases, such as cancer, where genetic tests are frequently used to identify the most effective treatments for a patient. To expand upon this success, the NIH launched the All of Us Research Program, a longitudinal research program that aims to collect health data on one million or more people living in the United States over many years. Using this broad dataset, the program aims to improve our understanding of a wide range of factors – biology, lifestyle, and environment – that contribute to health, and apply this understanding toward more evidence-based, individualized healthcare.

## All of Us at Northwestern

In 2016, Northwestern University became one of over 50 HPOs across the country to receive the All of Us Health Care Provider Organization Award. The designation involves building research protocols as well as enrolling individuals and collecting health data. Over the next five years, the \$60 million award will be distributed across five medical centers in Illinois: Northwestern University, The University of Chicago, Rush University Medical Center, NorthShore University HealthSystem, and the University of Illinois at Chicago, which together make up the Illinois Precision Medicine Consortium.

So far, more than 3,000 full participants – those who have provided general and electronic health record consent and completed in-person visits for biosamples and physical measurements – have enrolled through the consortium. More than 31,000 have enrolled nationally.

**“[The] All of Us Research Program is a longitudinal research program that aims to collect health data on one million or more people living in the United States over many years.”**



## EHRs as Data Source: The Challenges

In addition to its designation as an HPO for the program, the Northwestern center is also involved in developing tools to integrate and curate electronic health record (EHR) data from multiple institutions. Dr. Firas Wehbe, Northwestern Medicine’s Chief Research Informatics Officer, explained that this has been a rate-limiting step for the program. In order to facilitate the kind of large scale discovery anticipated from the All of Us program, data from multiple healthcare systems need to be integrated into a common data model. Even EHR systems within the same institution may require integration – Northwestern University’s various EHR systems, for example, had not been unified into a common data model until March of this year.

According to Dr. Abel Kho, director of the Center for Health Information Partnerships at Northwestern University, another major challenge

involved in health data analysis is integrating scattered health records from the same patient. Patients often change providers and insurance plans, causing their EHR data to be scattered across multiple institutions.

Further complicating the data retrieval process is the fact that EHR data contains unstructured data such as clinical notes, which are not easily organized into searchable databases. Although EHR systems carry a great wealth of patient data, Dr. Wehbe explains, they are designed primarily as transactional systems to manage workflow, and to serve as legal medical documentation, and thus are not directly amenable to analysis. Significant computing resources are therefore required to process the data and extract necessary information.

Another challenge involved in analyzing EHR data stems from the potential inaccuracy of the data.

For instance, Dr. Kho explains that healthcare providers sometimes input diagnosis codes in order to bill certain procedures rather than to document true diagnoses, making these codes somewhat unreliable. Demographic data may also have inaccuracies, as race and ethnicity is not self-reported in EHRs. Indeed, a 2015 study published in the *Journal of General Internal Medicine* found that 3% and 6.6% of Hispanic and African American patients, respectively, were not identified correctly in the EHR.



Photography by Rob Karlic.

*The All of Us program aims to address these problems by crosswalking self-reported, EHR, survey, and genetic data to verify patient information.*

It is clear that data collection at this scale will be challenging, particularly given that EHRs will serve as a major source of data for the project. On the positive side, Dr. Wehbe highlights how harmonious it has been to work with informaticians from all over the country. “Maybe it’s because there’s been enough work to go around that everyone is busy, but I’m surprised it progressed as it has.”



2018 © Nicole Ethen  
(UIC Biomedical Visualization graduate student)

## “Partners” not “Subjects”



The All of Us program has placed a notable emphasis on its recruitment language. Volunteers who enroll in the program are referred to as “participants”, or “partners”, rather than subjects. In addition to providing their health data to the program, which include biosamples, EHRs, physical exams, and surveys, Dr. Wehbe explains that participants will also have a say in the research questions that come out of the program. To facilitate this, the Illinois consortium has created an 11-member Community Participant Advisory Committee (CPAC), which met for the first time this month. Eventually, as the technology evolves and the program builds trust with participants, the program will expand to include data from mobile and wearable technologies.

Another emphasis in recruitment is the enrollment of underrepresented groups in biomedical research (UBR). The program aims to enroll at least 75% from UBR populations, which are broadly defined as people from minority race/ethnicity, low socioeconomic status, or low educational attainment groups. Historically, biomedical studies have failed to reflect the diversity of the US population, thus failing to fully comprehend the factors that affect disease outcomes in all populations. A 2015 study published in PLOS Medicine found that less than 2 percent of clinical cancer trials included enough underrepresented minorities to fulfill the NIH’s own diversity-related criteria. Thus far, however, the Illinois Consortium has successfully oversampled from UBR populations, with over 55% of participants identifying as Black/African American.

*While other biobanks have successfully collected and stored health data in the past, none have attempted to do so at the scale of the All of Us program, making it truly a giant leap forward in the pursuit of precision medicine.*



**Makda Zewde** is a part-time MS in Biostatistics student at Northwestern University and a Research Specialist at the Institute of Molecular Engineering at the University of Chicago, where she studies immunology. She enjoys communicating public health and scientific topics to the public, and is particularly interested in covering the use of big data in improving public health.