

June 24, 2019

ELSI Workshop

All of Us
RESEARCH PROGRAM

The
Future of
Health Begins
With You



National Institutes
of Health

Kelly Gebo, MD MPH
CMSO, All of Us Research Program

Acknowledgements

- Science Committee

- Keisha Bellamy
- Eric Boerwinkle
- Mine Cicek
- Elizabeth Cohn
- Stephanie Devaney
- Pam Factor-Litvak
- Mona Fouad (co-chair)
- Kelly Gebo
- David Goldstein
- Philip Greenland
- Parinda Khatri
- Christopher O'Donnell
- Lucila Ohno-Machado
- Akinlolu Ojo
- Andrea Ramirez
- Steven Reis
- Dan Roden
- Jordan Smoller (co-chair)
- Steven Steinhubl
- John Wilbanks
- Robert Winn

- Methods Co-Chairs

- Assays

- Maria Argos
- Steve Thibodeau

- DHT

- Ed Ramos
- Chris Lunt
- James McClain
- Steve Steinhubl

- EHR

- Abel Kho
- Daniella Meeker
- Andrea Ramirez

- Omics

- Bruce Korf

- PPI

- Rob Cronin
- Brian Ahmedani

- Consortium Members

- LCT

- IC Collaborators

- NIH/Leidos Team

- Keri Althoff
- Brian Florence
- David Galey
- Tram Huyen
- Sara Lioi
- Adele Luta
- Steve Mockrin
- Heather Sansbury
- Sheri Schully
- Carlie Williams

Disclosures

- Research funding: AHRQ, HRSA, NIAID, NIDA, NIA, Robert Wood Johnson Foundation, Johns Hopkins CFAR
- Scientific Consultant: Simon Fraser University

Objectives

- Understand the mission, objectives, and scientific framework of the *All of Us* Research Program
- Appreciate the data currently being collected within the *All of Us* Research Program
- Be able to identify research questions where *All of Us* could serve as a data source



Framingham Heart Study

Factors of Risk in the Development of Coronary Heart Disease— Six-Year Follow-up Experience

The Framingham Study

WILLIAM B. KANNEL, M.D., THOMAS R. DAWBER, M.D., F.A.C.P.,
ABRAHAM KAGAN, M.D., F.A.C.P., NICHOLAS REVOTSKIE, M.D.,
AND JOSEPH STOKES, III, M.D.
Framingham, Massachusetts

INCREASINGLY RELIABLE ESTIMATES of the prevalence and incidence of coronary heart disease (CHD) emphasize the importance of this disease as a contemporary health hazard. Cardiovascular disease is now the leading cause of death, with coronary heart disease accounting for two-thirds of all heart disease deaths. In the diagnosis and treatment of CHD have been made in the last decade, no important morbidity and mortality have been averted. This is apparent from the slight increase in life expectancy which has been achieved in the last decades, while life expectancy has been substantially prolonged.

Because coronary heart disease has manifested as sudden "silent" infarction and mortality in those surviving, it is still distressing. The best therapeutic effort and preventive program

Since it has been established that coronary atherosclerosis is present for many years prior to the development of symptomatic CHD, it seems evident that efforts at prevention must begin many years before the appearance of clinical CHD. A knowledge of the epidemiology of the disease is highly

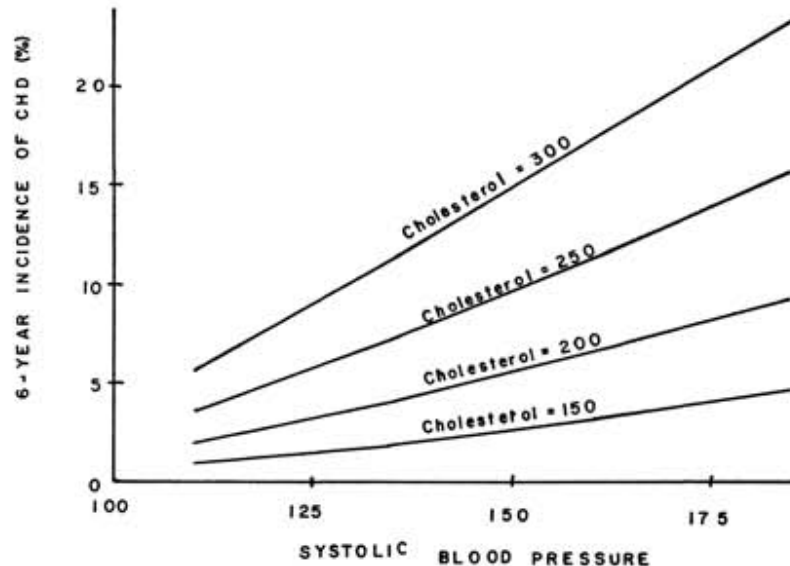


FIGURE 2. Six-year incidence of coronary heart disease according to level of systolic blood pressure at specified serum cholesterol levels (men 45 to 62 years). For explanation, see legends for Figure 1.

Enrolled 5209 men and women in 1948

Some Framingham early discoveries:

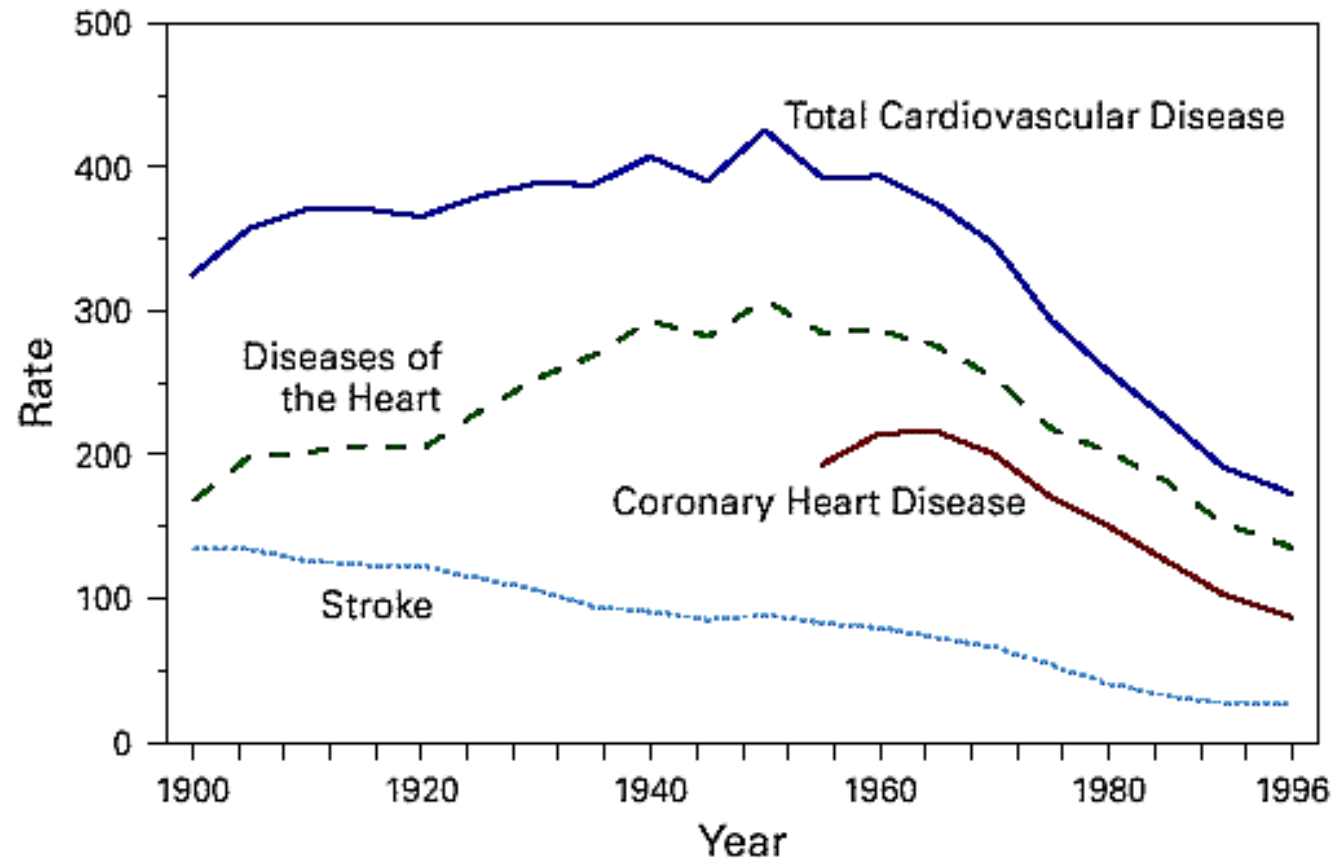
- 1960 – Cigarettes increase heart disease
- 1961 – cholesterol, blood pressure increase heart disease
- 1967 – exercise decreases risk of heart disease; obesity increases it
- 1970 – high blood pressure and atrial fibrillation cause stroke

Received for publication
From the Heart Disease
Framingham, Mass., and the
Institute, National Institutes of
Health, U. S. Department
of Health, Education,
and Welfare, Washington,
D. C.
Presented at the Forty-
fourth Annual Meeting of
The American College of
Cardiology, Miami Beach, Fla.,
November 1961.

Requests for reprints to:
Thomas R. Dawber, M.D.,
Director, Heart Disease Epidemiology
and Prevention Branch,
Framingham, Mass.

The impact of Framingham (and similar cohorts) has been dramatic

FIGURE 1. Age-adjusted death rates* for total cardiovascular disease, diseases of the heart, coronary heart disease, and stroke,† by year — United States, 1900–1996



Our Mission

To accelerate health research
and medical breakthroughs, enabling individualized prevention, treatment,
and care for all of us

Nurture relationships

with one million or more
participant partners, from all
walks of life, for decades

Deliver the largest, richest biomedical dataset ever,

making it as easy, safe, and
free to use as possible

Catalyze a

robust ecosystem

of researchers and funders
hungry to use and support it

All of Us Research Program Core Values

Participation is **open** to all.

Participants reflect the rich **diversity** of the U.S.

Participants are **partners**.

Trust will be earned through **transparency**.

Participants have **access** to their information.

Data will be accessed **broadly** for research purposes.

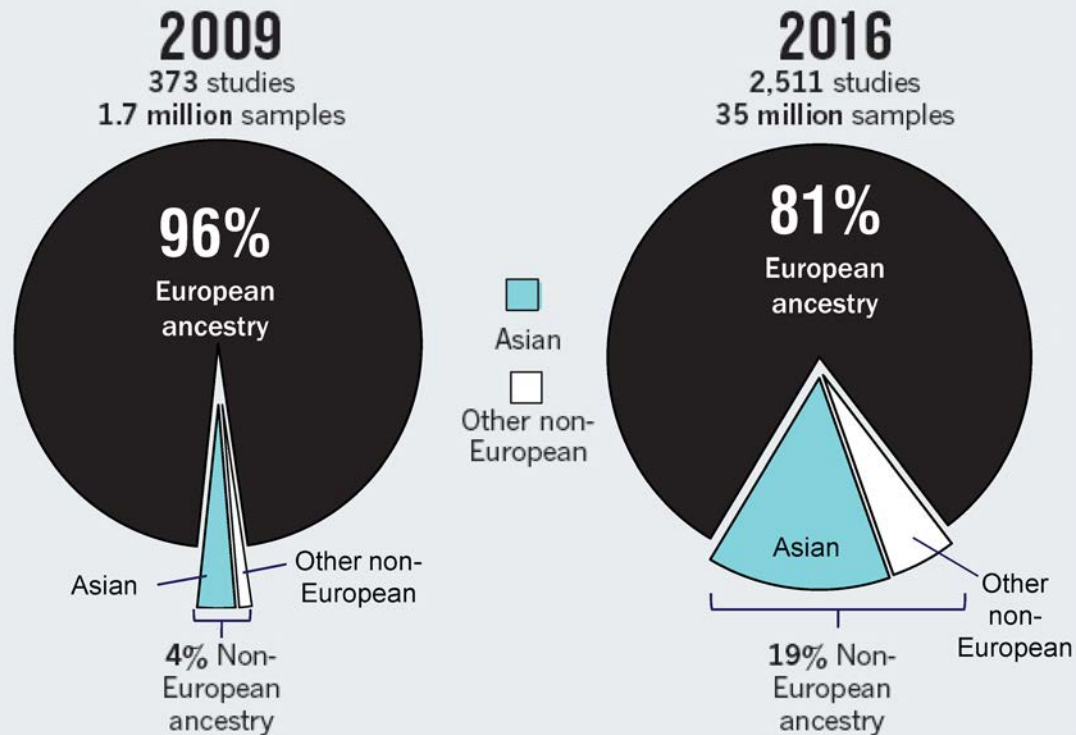
Security and privacy will be of highest importance.

The program will be a catalyst for **positive change** in research.

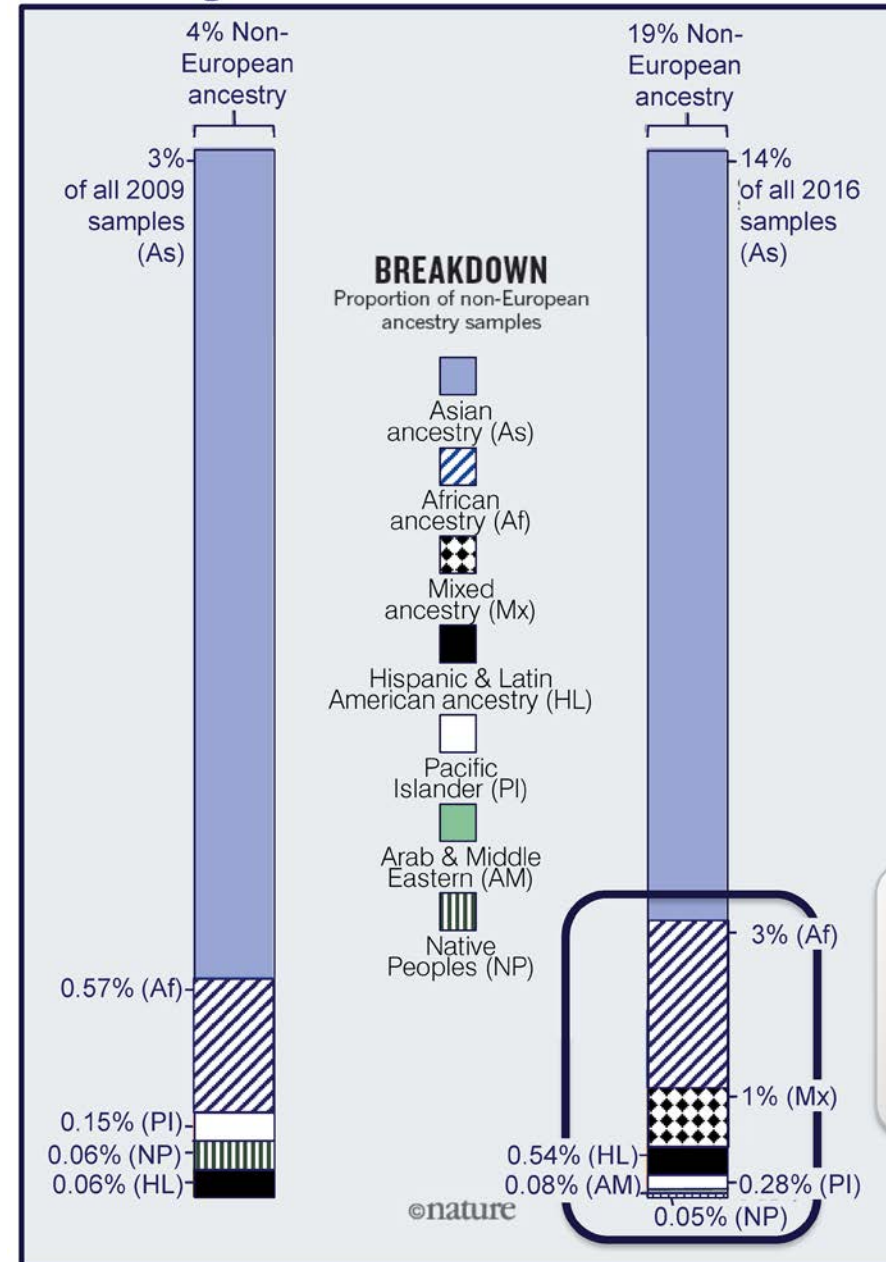
Why Diversity?

PERSISTENT BIAS

Over the past seven years, the proportion of participants in genome-wide association studies (GWAS) that are of Asian ancestry has increased. Groups of other ancestries continue to be very poorly represented.



Popejoy & Fullerton, *Nature* 2016



Terms for ethnicity are those used in the GWAS catalog. Some have changed between 2009 and 2016 as sampling has increased. Samples of European origin have the most specific descriptions of population ancestry.

4% GWAS represents >33% US population

Innovative Aspects of the *All of Us* Research Program

- Diversity at the scale of 1 million people or more
- Longitudinal, able to recontact
- EHR, surveys, baseline physical evaluation and biospecimens—including genomics
- Focus on participants as partners
- **National, open resource for all:** open to all researchers with open source software & tools



Participants as Partners

Involved in every step of program development

- What **data** we collect
- What **lab analyses** we do
- What **research** is conducted
- How data is **returned**
- Partnership with national and local community groups



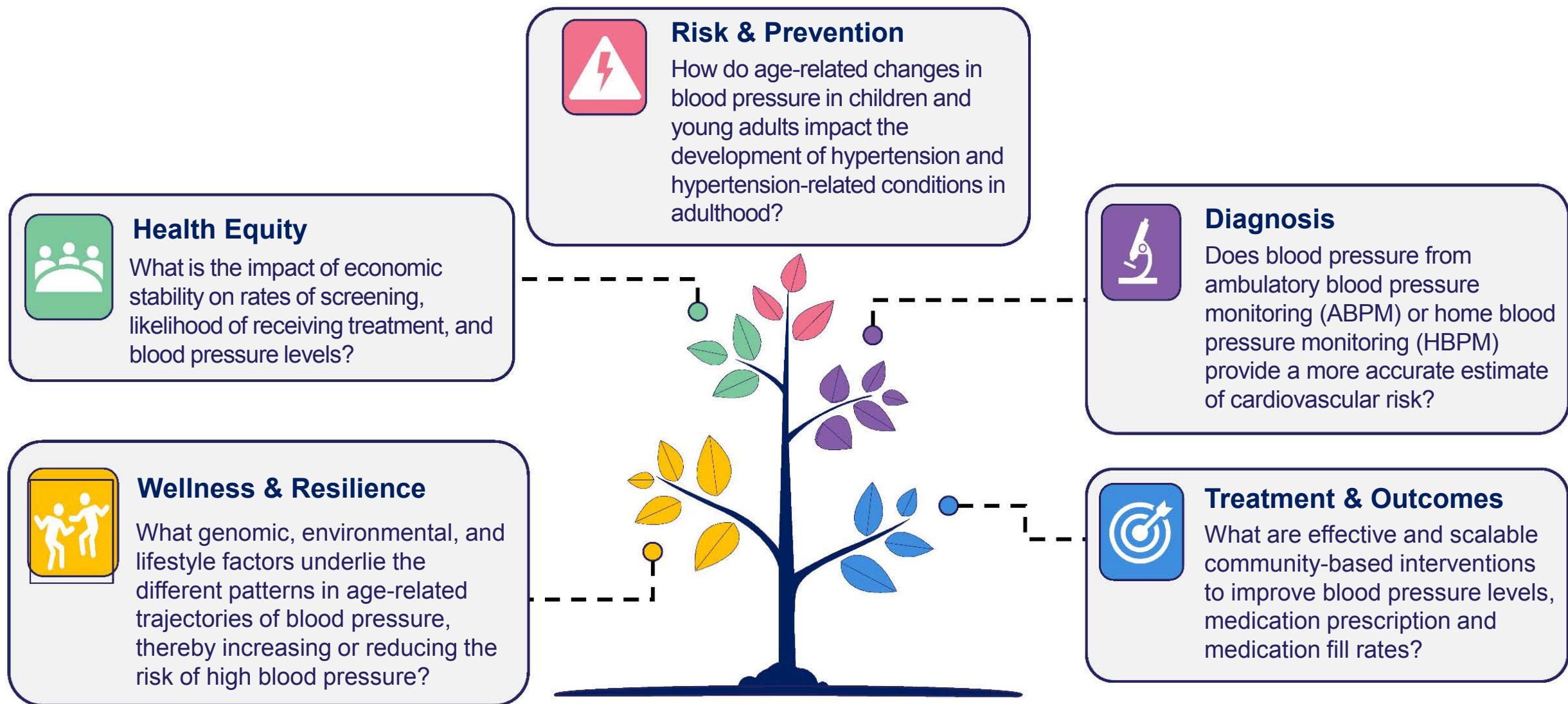
Scientific Framework

Scientific Framework

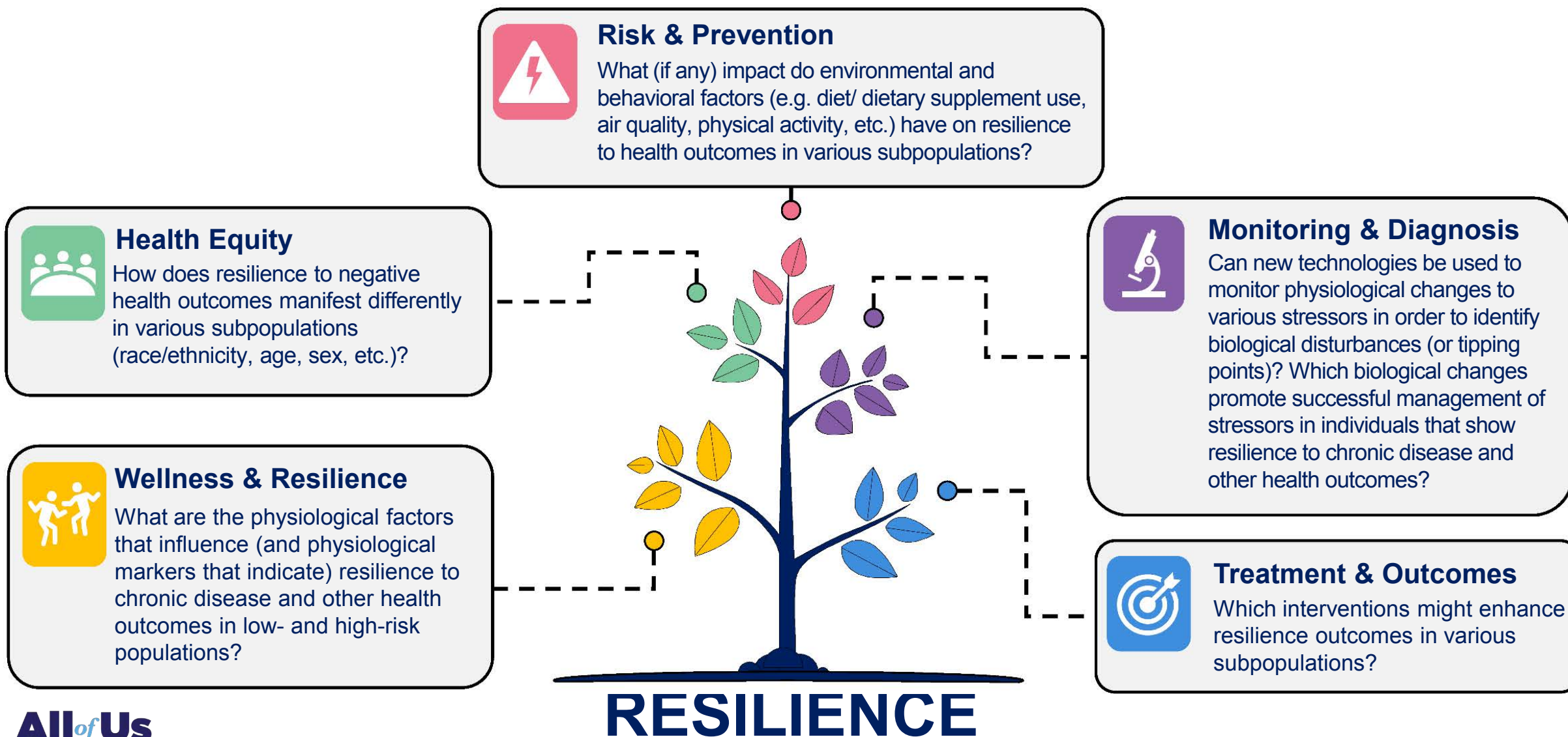
Enable research that will:

- I. Increase wellness and resilience, and promote healthy living
- II. Reduce health disparities and improve health equity in underrepresented in biomedical research (UBR) populations
- III. Develop improved risk assessment and prevention strategies to preempt disease
- IV. Provide earlier and more accurate diagnosis to decrease illness burden
- V. Improve health outcomes and reduce disease impact through improved treatment and development of precision interventions

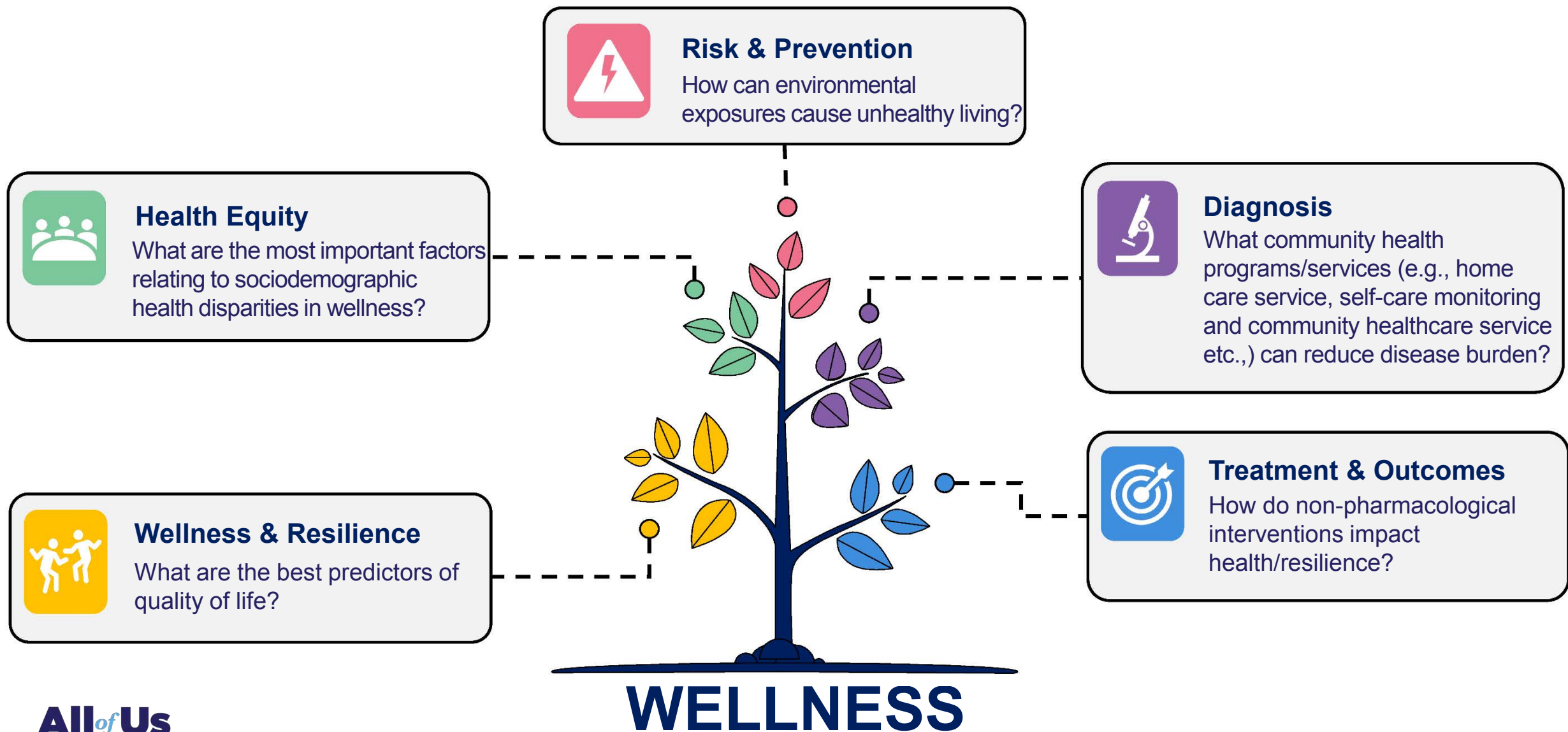
Example Use Case: Blood Pressure



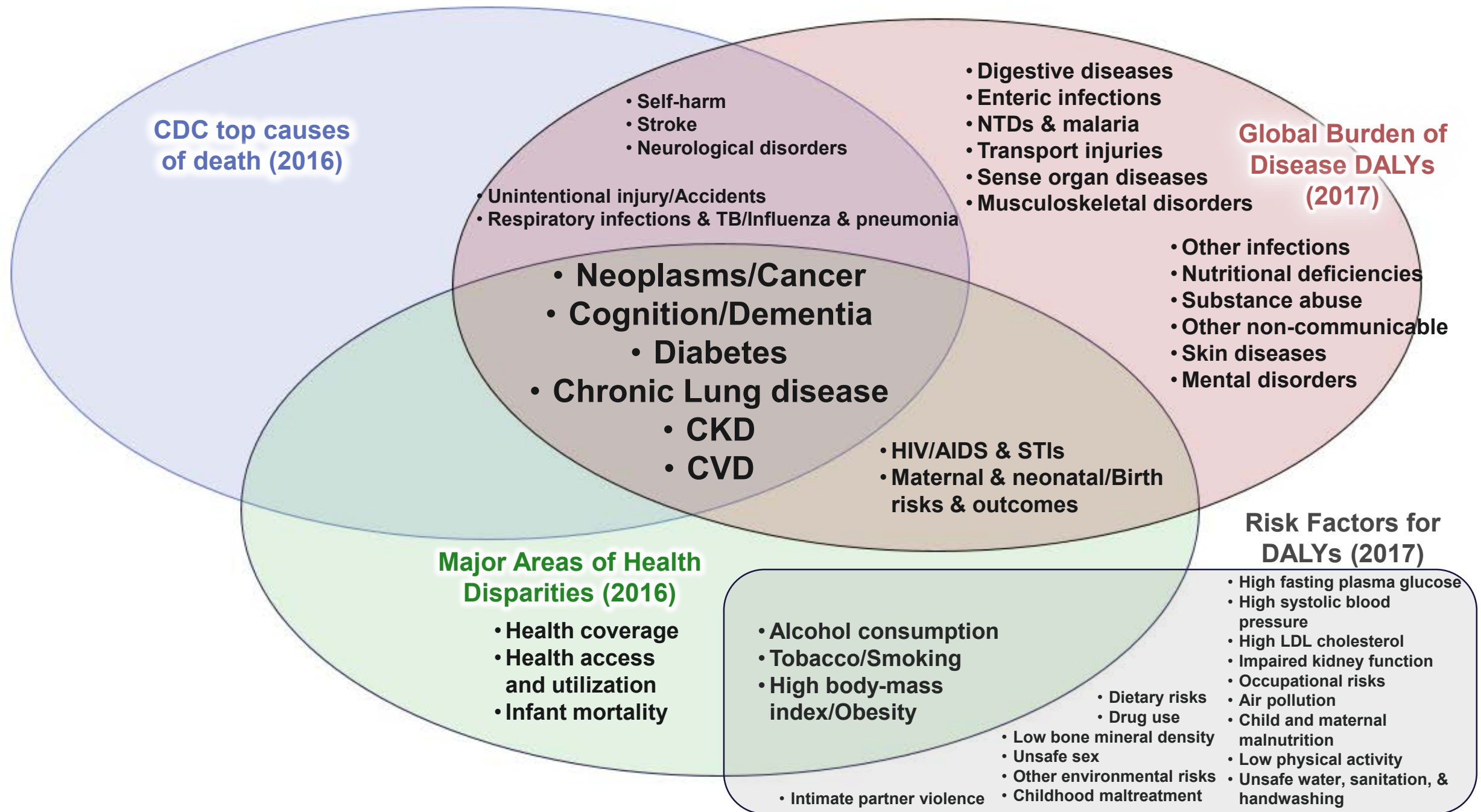
Resilience in Chronic Disease and Health outcomes



Example Use Case: Wellness



Focusing



The “Big 8”

Big 8: Selected Health Areas

1. Cancer
2. Cardiovascular Disease
3. Chronic Kidney Disease
4. Chronic Lung Disease
5. Diabetes/Obesity
6. Mental Health/Cognition
7. Opioid Use/Chronic Pain
8. Wellness

Current Participant Journey

Current protocol



Enroll, Consent & Authorize EHR

- Recruiting 18+ years old initially; plan to include children in 2019
- Online, interactive consent
- Includes authorization to share Electronic Health Record (EHR) data



Answering Surveys

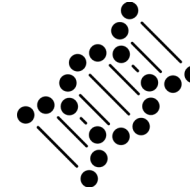
- Initial surveys: The Basics, Overall Health, Personal Habits, Health Care Access & Utilization, Family Medical History
- Additional surveys released on an ongoing basis.



Physical Measurements*

- Blood pressure
- BMI
- Heart rate
- Height
- Hip circumference
- Waist circumference
- Weight

**Based on diverse sampling and capacity*



Provide Biosamples*

- Blood (or saliva, if blood draw is unsuccessful)
- Urine specimen
- Biosamples will be stored at the program's biobank

**Based on diverse sampling and capacity*



Wearables and Digital Apps

- Share data from wearable fitness devices, starting with FitBit
- Share data about mood & cardio-respiratory fitness through integrated apps
- More integrations to come

Coming soon

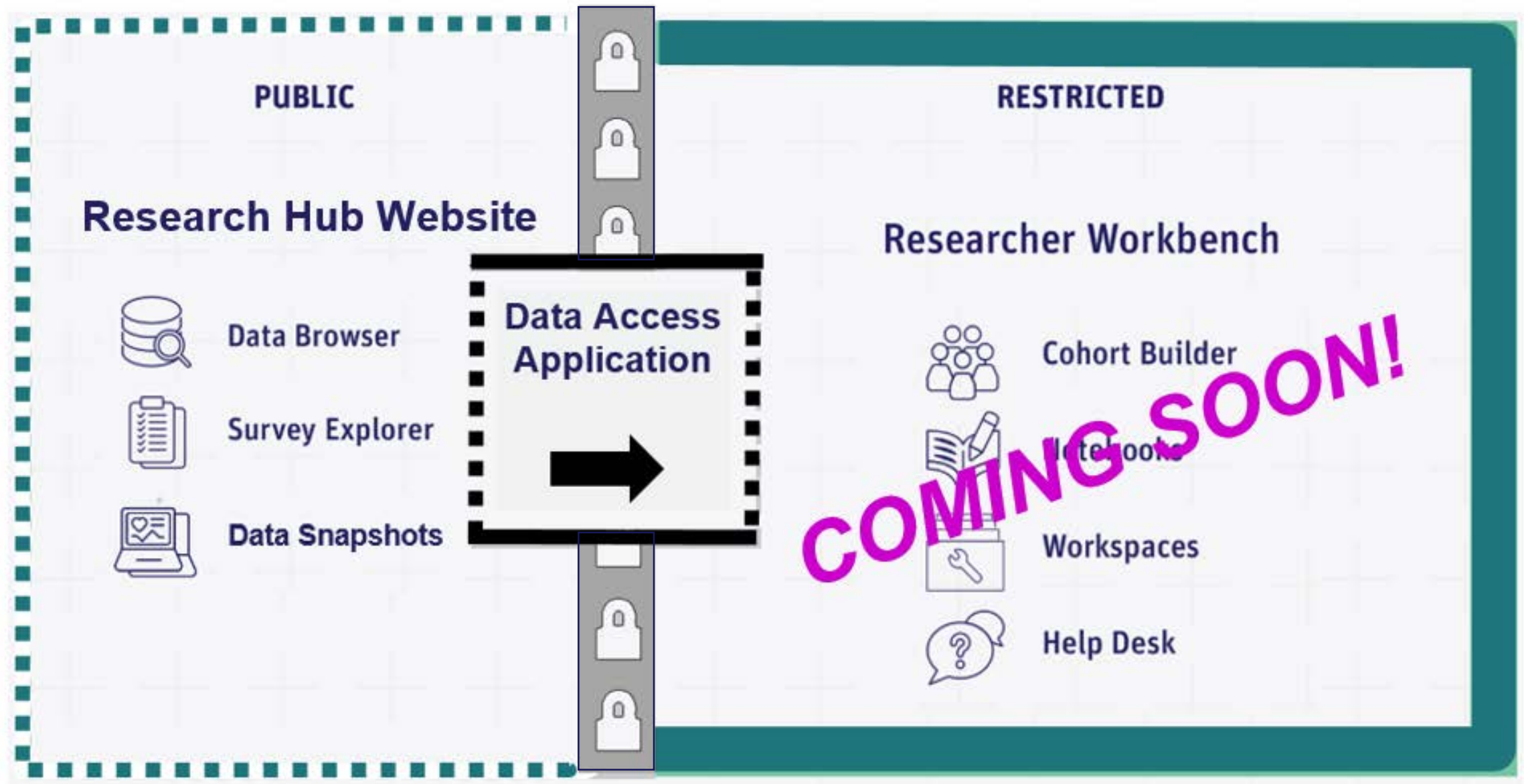
Schedule of Assessments Template: For a Participant Enrolling June 24

Time zero



Assessment	7/19-9/19	10/19-12/19	1/20-3/20	4/20-6/20	Year 2	Year 3 reassessment	Year 4
Physical Measurements: Height, weight, blood pressure, heart rate, hip circumference, waist circumference	✓					TBD	
Biospecimens: Blood (50 mL), urine (30 mL), saliva	✓					TBD	
PPI	Basics, Lifestyle, and Overall Health		Healthcare Access and Utilization, Personal Medical History, and Family History	Mental Health	TBD	Reconfirm Basics, Lifestyle, and Overall Health	TBD
Omics			Genotyping and WGS				
EHR Consent to obtain EHR (day 0)	✓						
DHT • BYOD Fitbit • Project Fitbit (TBD) • Apple (TBD) • Mood App	✓			✓			

All of Us Research Hub



Data Access | *Data Access Principles and Framework*

- Data available to **all types of users**
- Employ a **cloud-based analysis platform**
- Access will be **tiered**
- Users will be granted **data passports**
- Project information will be made **public** and **auditable**
- Developing policies on **access to samples & cohort**





Journey to Protocol Roadmap

Protocol Plan and Timeline

Task	Who	Due					
A. Determine Selected Health Areas*	SciCom	April 5					
B1. Identify candidate variables associated with each Selected Health Area	SciCom	May 1					
B2. Cohort Gap Analysis	Kelly/NIH	May 8					
C. Harmonize candidate variables	NIH/Leidos	May 17					
D. Determine “core” variables	NIH to recommend to SciCom	May 23	Is this variable associated with one or more of the Selected Health Areas?		<div>Selected Health Areas<ul style="list-style-type: none">1. CVD/Cerebrovascular Disease2. Cancer3. Diabetes and Obesity4. Opioids, Marijuana, Alcohol, and Pain5. Mental Health and Cognition6. Chronic Lower Respiratory Diseases7. Chronic Kidney Disease8. Wellness</div>	<div>No</div>	Variable should be considered for an ancillary study or by an NIH IC
			Yes			<div>No</div>	Variable should be considered for an ancillary study or by an NIH IC
			Is the AoU platform the right platform to capture this variable (on 1 million+ people or a significant subset of the population)?				
E. Create list of “core” datatypes from “core” variables	NIH	May 24	Yes				
F. Assign status (in protocol, in development, from RPW) to “core” datatypes	NIH/Leidos	May 24					
G. Filter by “in protocol” and “in development” from “core” datatypes	NIH/Leidos	May 24			<div>Methods Criteria<ul style="list-style-type: none">1. Affordable at scale2. Low participant burden3. Does not pose a safety risk4. Currently available, used, valid & reliable5. Advances innovation in data collection</div>	<div>No</div>	Consider methods pilot project funded by AoU or an NIH IC
H. Recommend methods to collect datatypes in remaining list	Methods co-chairs	Week of May 27	Is there currently an optimized method for collecting this datatype?				
I. Prioritize optimal datatype/method pairs to NIH (NIH chaired WebEx)	SciCom with Methods co-chairs	Week of June 3 Initiated	Yes				
			Does this data type/method leverage and balance expertise across AoU program and can be implemented now?		<div>Yes</div>	Bin Near term	
					<div>No</div>	Bin Medium or Long term	

Protocol Plan and Timeline, cont'd.

Task	Who	Due	Process/Criteria	Outcome
J1. Prepare F2F presentation	NIH/SciCom	June 6-June 11 		
J2. Obtain F2F SC input	SciCom to SC	June 13  we are here		
J4. Continue to prioritize optimal datatype/method pairs	SciCom w/ Methods Co-Chairs	TBD		
J3. Obtain selected LCT member review	NIH			
K. Ensure no populations are alienated	Ambassadors & Community Engagement Partners	TBD		
L. Prioritize PPEs within each time-frame = Draft Road Map	NIH to DCM	TBD		
M. Create Schedule of Assessments & Reassessments with near term PPEs = Protocol V2	NIH to DCM	TBD	<i>*Additional health areas will be considered for future protocol iterations</i>	

Variables Under Consideration

Guiding Principles (GP) for Selecting Methods*

- GP 1: The method(s) chosen should be parsimonious (affordable at scale; low participant burden; available, reliable, and valid; advances innovation in data collection)
- GP 2: The method(s) chosen will collect the datatype from as many participants as possible (e.g. one million participants or a significant subset of them)
- GP 3: If we can't get a datatype from all participants using one method, we use a 2nd method to get the datatype from the remaining participants
(e.g., If we don't get a datatype from all participants using EHR, we use PPI to get the datatype from the remaining participants)
- GP 4: The method(s) should address the heterogeneity and quality of the data from different UBR populations

*Note: numbers are used to facilitate discussion, not to indicate priority

All of Us Research Program Protocol Development

Kidney Disease

In Protocol / In Development

Diagnosis of Kidney Disease (PPI)

	NEAR-TERM					MID TO LONG-TERM				
Variables for consideration	EHR	Exam	PPI	DHT	Assay	EHR	Exam	PPI	Assay	Other
Treatment of Kidney Disease										
Hemodialysis	✓									✓ (Linkage to USRDS)
Peritoneal Dialysis	✓									✓
AV Fistula placement	✓									✓
Complication of Kidney disease (hyper PTH)						✓			✓	

All of Us Research Program Protocol Development

Cancer

In Protocol / In Development

Diagnosis of Cancer, Treatment for Cancer (PPI)

	NEAR-TERM						MID TO LONG-TERM				
Variables for consideration	EHR	Exam	PPI	DHT	Assay	Other	EHR	Exam	PPI	DHT	Other
Diagnosis of Cancer	✓		✓				✓		✓		✓ (Linkage to SEER)
Treatment for cancer							✓		✓		✓
Cancer recurrence							✓		✓		✓ (Linkage to SEER)
Tumor characteristics							✓				
Dx of precancerous condition (FAP)	✓						✓				

Remember

- Proposed variable lists include:
 - *Italicized Blue* are in protocol or in development
 - Black are for consideration
 - Lists are organized by outcome, sociodemographic factor, SDOH, risk factor, exposure, lab tests, genetics and omics
- Remember: This is NOT FINAL
- We are currently seeking feedback

Request

We ask for your feedback to the following questions on the new core datatypes* under consideration:

- What is blatantly missing (variables and methods)?
- What would draw you to this dataset that you don't see represented (variables and methods)?
- Should something be removed from this list?

RESPONSE

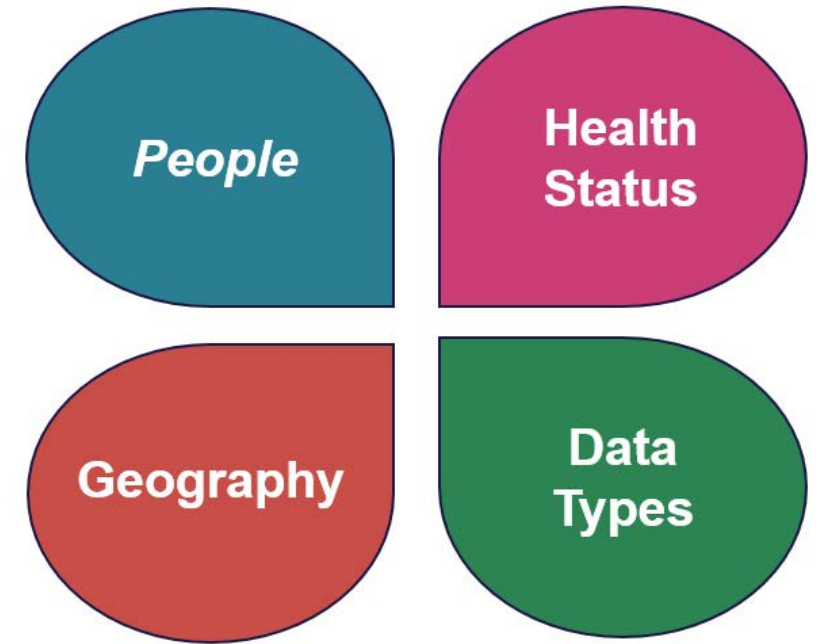
- Please email your response to Kelly Gebo at kelly.gebo@nih.gov by COB **June 28, 2019**

***1M+ participants or a large subset of the population**

Discussion Questions

Innovative Aspects of *All of Us*

- ➔ • **Diversity at the scale of 1 million people:** demographically, geographically, medically, and especially those underrepresented in biomedical research
- ➔ • **Diversity of data types collected longitudinally:** clinical, environmental, genetic, behavioral, socioeconomic
- **Focus on participants as partners:** included in governance, invited to co-invent systems and give input into the science, choice to receive all data and information back
- **National, open resource for all:** open to the public and all researchers, open source software & tools



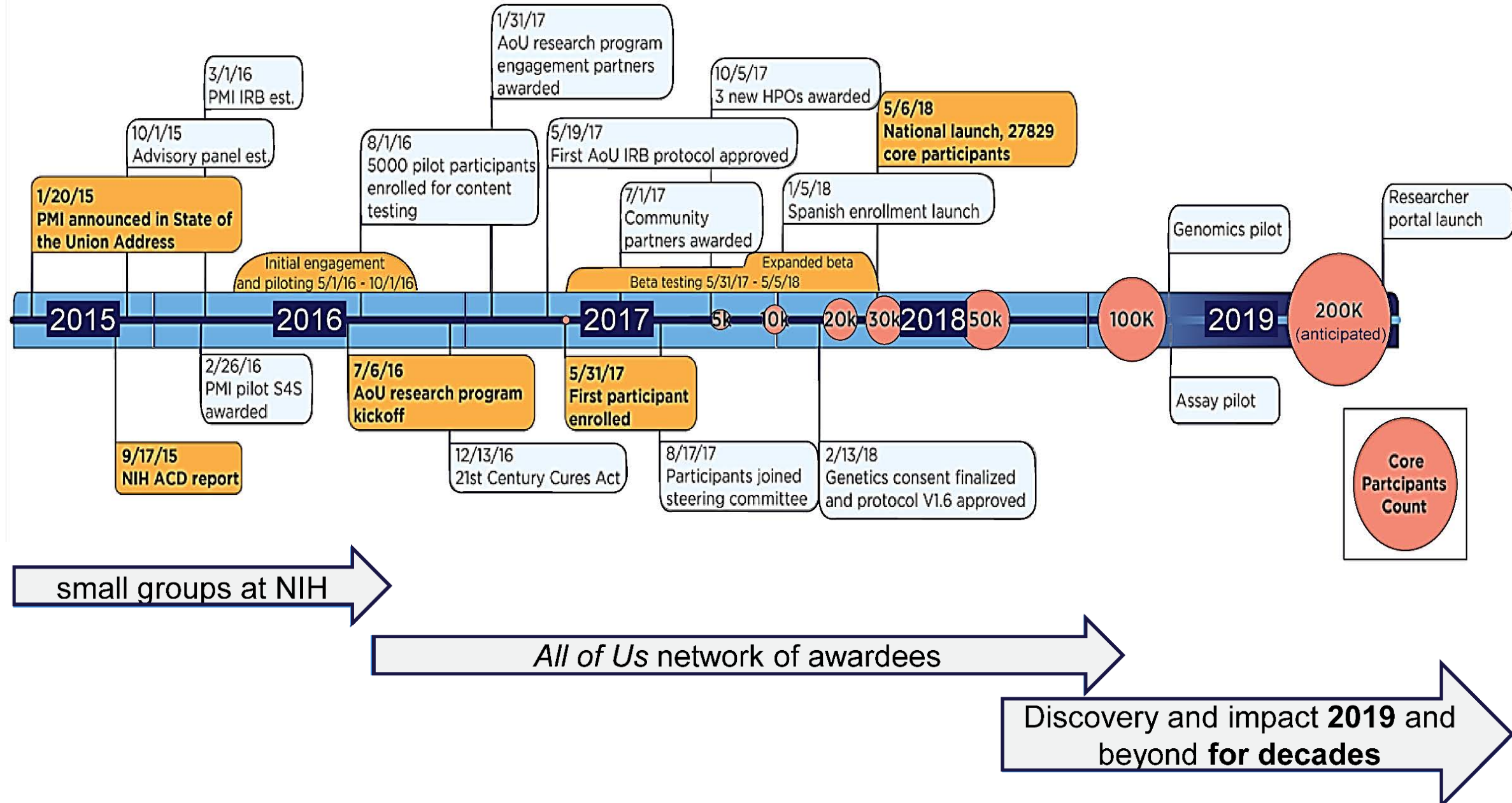
For Discussion

- How do we retain the innovative aspects of *All of Us* while collecting a complete dataset as possible on one million participants?
 - Designing things for a million, but anticipate others will build focused cohorts for deeper phenotypic assessments
 - Propose to collect new variables by EHR initially and after one year assess the completeness of the data
 - What variables are so important that if EHR is not complete, we would employ another method to complete the dataset (e.g., DHT, PPI, Assays)
 - Spirometry
 - Gait analysis
- Should we hold off on doing assays (as technology will become more advanced, cheaper, and efficient) as suggested by Assays Task Force and external stakeholders?

For Discussion

- Realizing we are trying to minimize the burden to participants by maximizing data collection from EHR and passive data collection methods with DHT
- How do we include important ELSI variables/questions into the *All of Us* scientific protocol roadmap and next version of the protocol?
 - What are the most important research questions?
 - What variables are needed to answer these?
 - What methods are used to collect these variables?

We've done a lot in 2.5 years!



All of Us consortium members

DV Network (Direct Volunteers)



Biobank



Communication & Engagement



HPO Network (Health Care Provider Organizations)

RMCs

California Precision Medicine Consortium

UC San Diego Health



Keck Medical Center of USC



Illinois Precision Medicine Consortium



New England Precision Medicine Consortium



Trans-American Consortium for the Health Care Systems Research Network



New York City Precision Medicine Consortium



Southern All of Us Network



SouthEast Enrollment Center



All of Us, Wisconsin



University of Arizona



University of Pittsburgh



FQHCs (Federally Qualified Health Centers)



VA Medical Centers



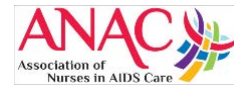
Platform Development



Genomics Infrastructure



All of Us: Current Community Partners Network



It takes *All of Us*....



For more information...



ResearchAllofUs.org



@AllofUsResearch
#JoinAllofUs



National Institutes
of Health

AllofUs.nih.gov

databrowser.researchallofus.org

Questions?
