University of Colorado **Anschutz Medical Campus**

All of Us Research Program

A presentation by the Population Shared Health Resources

Presenter:
Adam Warren

Introduction to the All of Us Research Program (AoURP)

- Introduction (Description, goals, mission and values)
- Limitations & Alternative datasets
- Participation
- Data Structure and Elements
- Researcher Prospective
- Examples

Introduction to the All of Us Research Program (AoURP)

- Created and managed by National Institutes of Health
- DIVERSE database with 668,000+ participants!
- Rooted in Equity
- Designed to be Longitudinal
- Researcher Friendly
- Free-ish ($300 Computational Credits to start)

University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

Introduction to the All of Us Research Program (AoURP)

The All of Us Research Program: Data quality, utility, and diversity

Citation: Ramirez, A. H., Sulieman, L., Schlueter, D. J., Halvorson, A., Qian, J., Ratsimbazafy, F., ... & Wellis, D. (2022). The All of Us Research Program: data quality, utility, and diversity. *Patterns*, 3(8).

Patterns

CellPress
OPEN ACCESS

Descriptor

The *All of Us* Research Program: Data quality, utility, and diversity

Andrea H. Ramirez,[1,2,22,*] Lina Sulieman,[3] David J. Schlueter,[4] Alese Halvorson,[3] Jun Qian,[3] Francis Ratsimbazafy,[5] Roxana Loperena,[5] Kelsey Mayo,[5] Melissa Basford,[6] Nicole Deflaux,[6] Karthik N. Muthuraman,[6] Karthik Natarajan,[7] Abel Kho,[8] Hua Xu,[9] Consuelo Wilkins,[1] Hoda Anton-Culver,[10] Eric Boerwinkle,[11] Mine Cicek,[12] Cheryl R. Clark,[13] Elizabeth Cohn,[14] Lucila Ohno-Machado,[15] Sheri D. Schully,[2] Brian K. Ahmedani,[16] Maria Argos,[17] Robert M. Cronin,[18] Christopher O'Donnell,[19] Mona Fouad,[20] David B. Goldstein,[21] Philip Greenland,[22] Scott J. Hebbring,[23]

(Author list continued on next page)

[1]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
[2]All of Us Research Program, National Institutes of Health, Bethesda, MD, USA
[3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
[4]Center for Precision Health Research, Precision Health Informatics Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
[5]Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA
[6]Verily Life Sciences, San Francisco, CA, USA
[7]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA
[8]Center for Health Information Partnerships, Northwestern University, Chicago, IL, USA
[9]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
[10]Department of Medicine, University of California Irvine, Irvine, CA, USA

(Affiliations continued on next page)

THE BIGGER PICTURE The engagement of participants in the research process and broad availability of data to diverse researchers are essential elements in building precision medicine equitably available for all. The NIH has established the ambitious All of Us Research Program to build one of the most diverse health databases in history with tools to support research to improve human health. Here, we present the initial launch of the Researcher Workbench with data types including surveys, physical measurements, and electronic health record data with validation studies to support researcher use of this novel platform. Broad access for researchers to data like these is a critical step in returning value to participants seeking to support the advancement of precision medicine and improved health for all.

1 2 3 4 5 **Production:** Data science output is validated, understood, and regularly used for multiple domains/platforms

SUMMARY

The *All of Us* Research Program seeks to engage at least one million diverse participants to advance precision medicine and improve human health. We describe here the cloud-based Researcher Workbench that uses a data passport model to democratize access to analytical tools and participant information including survey, physical measurement, and electronic health record (EHR) data. We also present validation study findings for several common complex diseases to demonstrate use of this novel platform in 315,000 participants, 78% of whom are from groups historically underrepresented in biomedical research, including 49% self-reporting non-White races. Replication findings include medication usage pattern differences by race in depression and type 2 diabetes, validation of known cancer associations with smoking, and calculation of cardiovascular risk scores by reported race effects. The cloud-based Researcher Workbench represents an important advance in enabling secure access for a broad range of researchers to this large resource and analytical tools.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

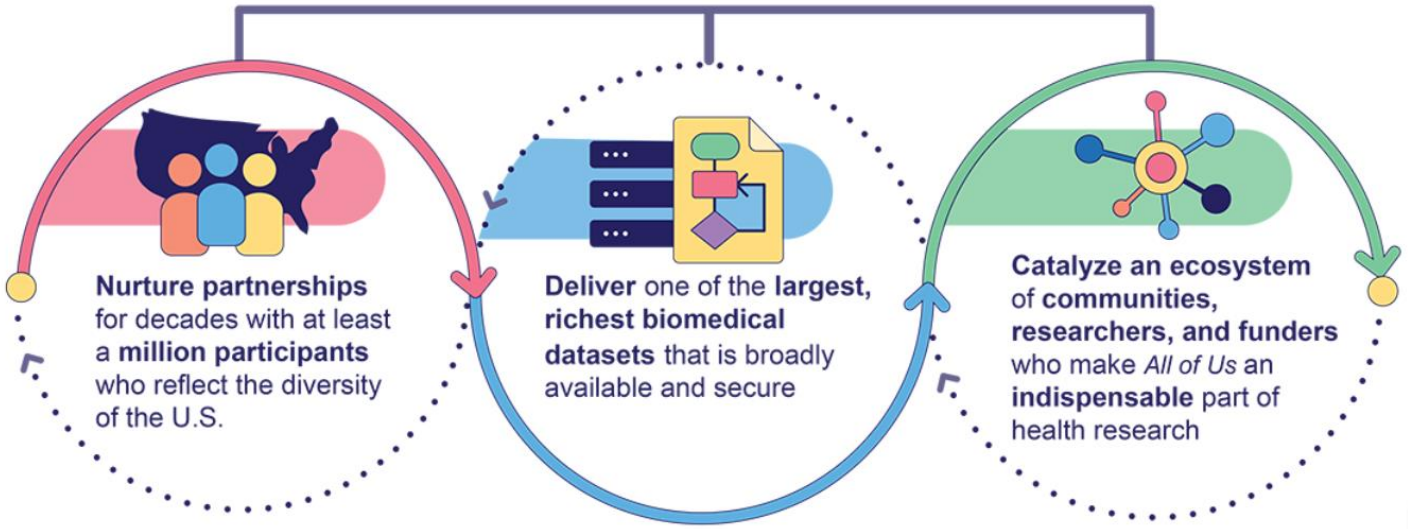
University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

Limitations

- Must use Python or R to conduct analyses
- Robust Sampling methods not used
- Electronic Health Records to have expected missingness
- Biosampling to be a challenge for those living in rural areas/far from recruitment sites.
- Participation Retention
- PHSR Exploration

Mission & Values



Core Values

- Participation open to all
- Participants reflect the rich diversity of the US
- Participants have access to their own Info
- Broadly accessible data for research

Data Structure & Elements

Longitudinal Research Program

University of Colorado
Anschutz Medical Campus

Data Elements:

- Surveys
- Physical Measurements
- Electronic Health Records
- Personal Health Technology
- Genomics
- Biospecimen Collections

All of Us RESEARCH PROGRAM

Who is all of us?

Inclusion/Exclusion Criteria	Data Collection	Demographics

All of Us: Participants

Inclusion Criteria

- Current resident US Adults aged 18 and older w/ capacity to consent
- Insurance not a qualifier

Exclusion Criteria

- Adults' w/o decisional capacity to consent
- Children (<18 years old)
- Incarcerated Individuals

Data Collection

- Health Care Provide
- AoU Survey's
- Biospecimen

Recruitment

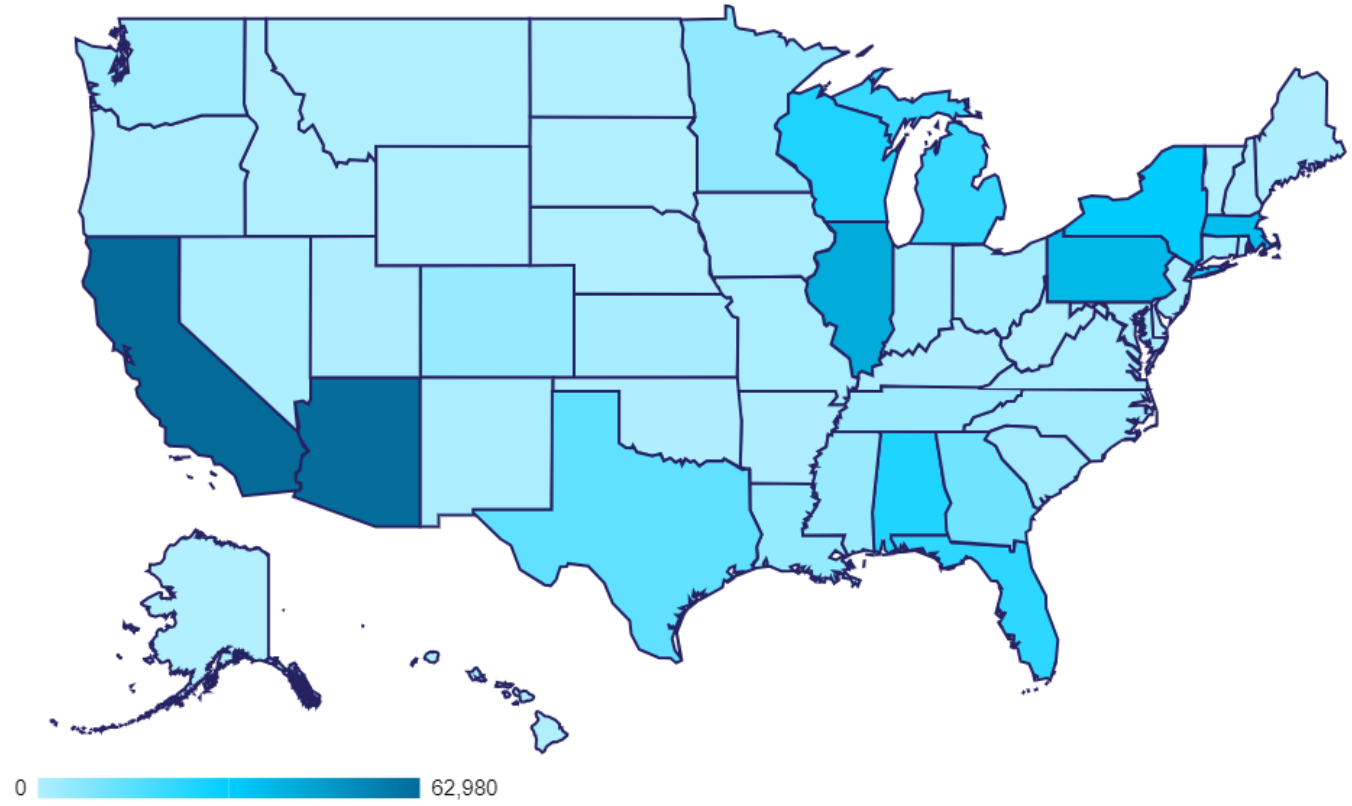
- Active enrollment/recruitment
- Healthcare Provider Organization outreach
- Community groups, seminars, & tabling
- Web-based advertising

University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

All of Us: Demographics

Geographic Distribution

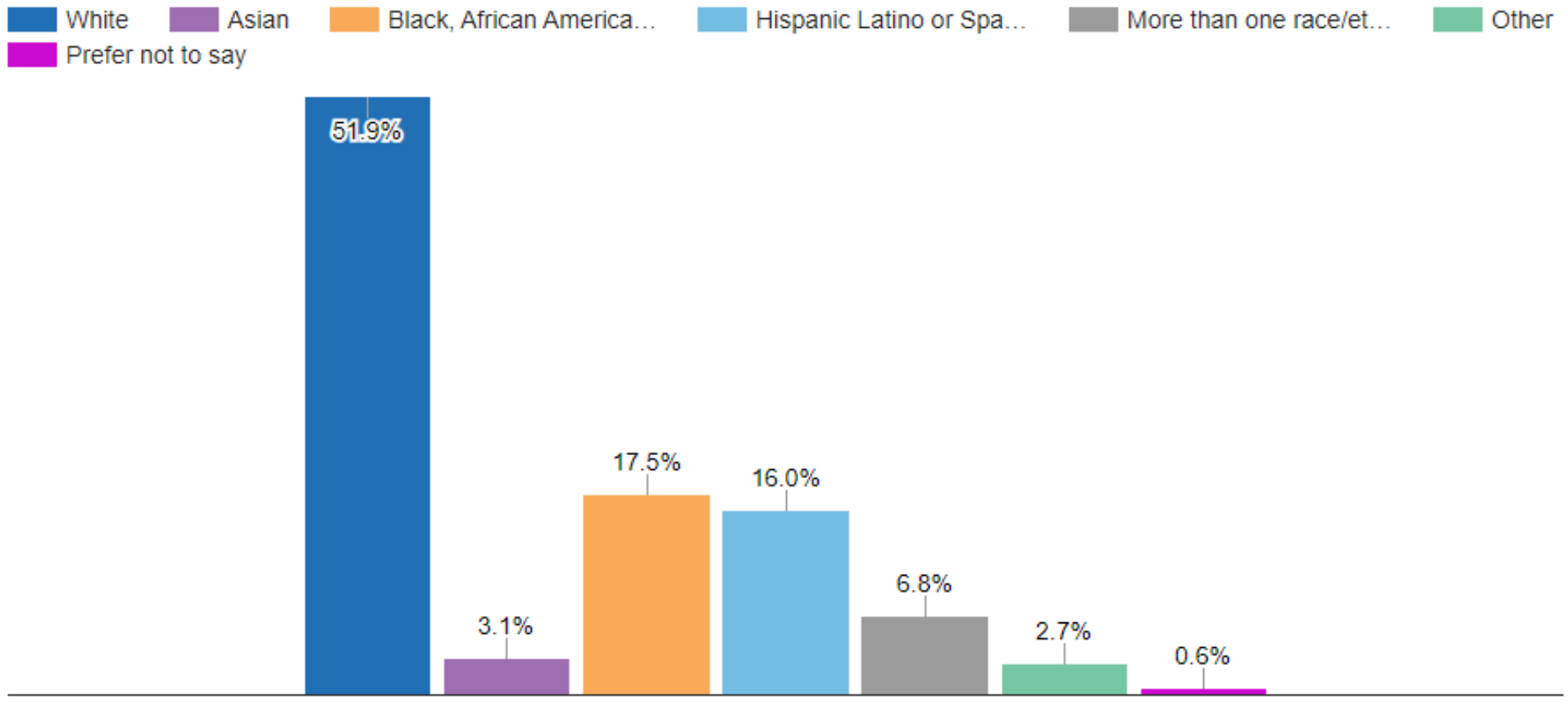


University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

All of Us: Demographics

Race & Ethnicity

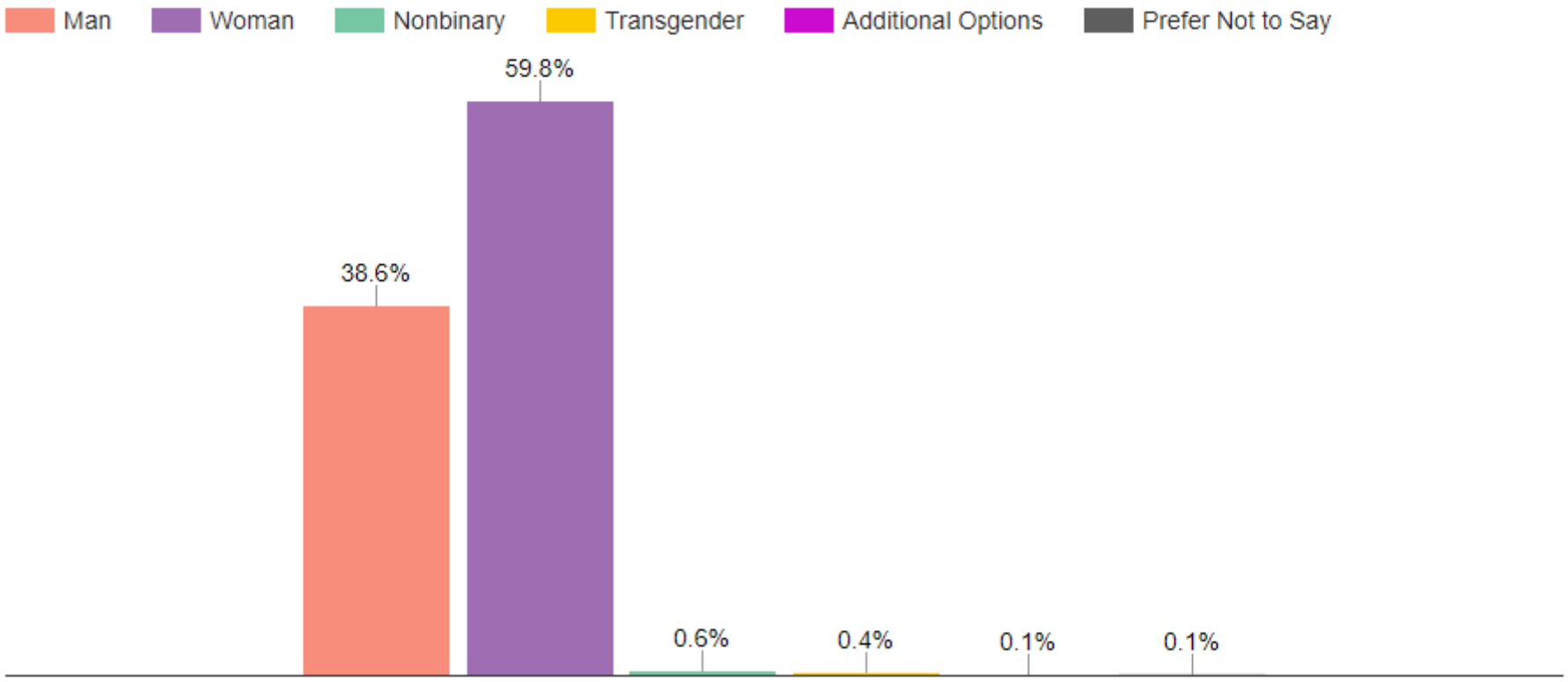


University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

All of Us: Demographics

Gender Identity

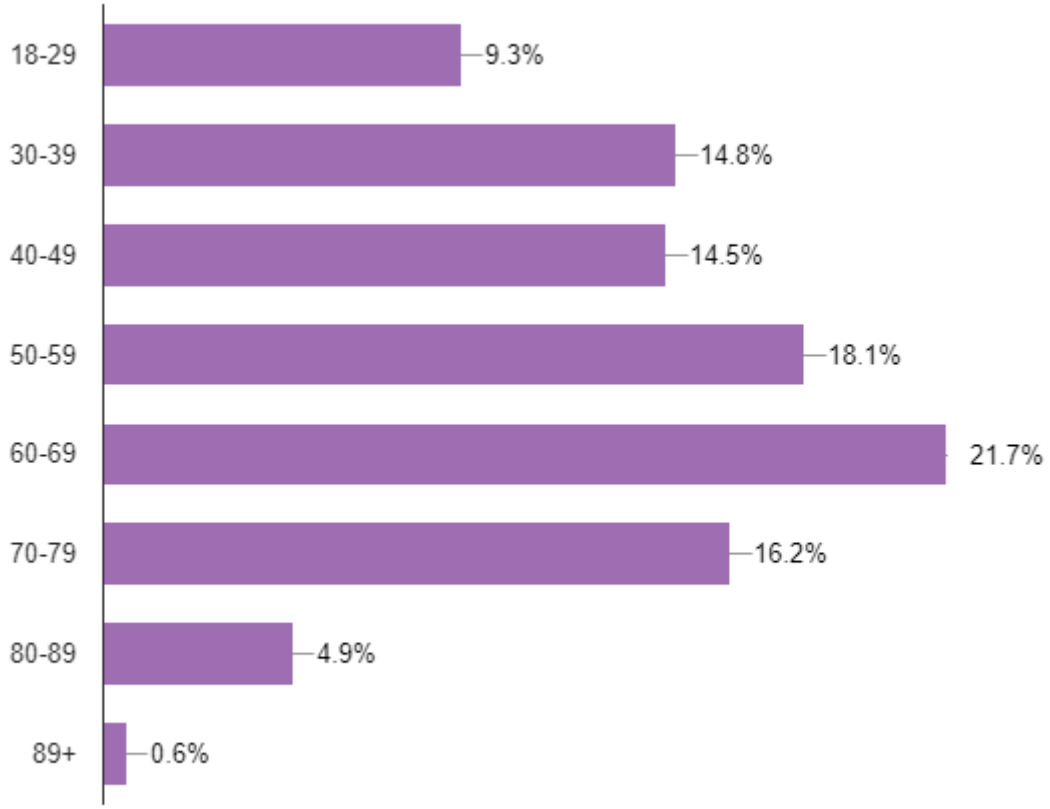


University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

All of Us: Demographics

Age



University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

All of Us: Demographics

Participants included in All of Us research data are diverse.

Underrepresented in Biomedical Research (UBR) Categories	Curated Data (% out of 413,450 participants)
At least one UBR	75%
Non-white race or Hispanic/Latino ethnicity	43%
Age >= 65	24%
Less than GED	9%
Annual Income <=$25k	25%
Sexual and Gender Minorities	10%
Disability	10%

Data Elements - Surveys

- **The Basics* -** basic demographic questions, including questions about a participant's work and home.
- **Lifestyle* -** asks about a participant's use of tobacco, alcohol, and recreational drugs.
- **Overall Health* -** collects information about a participant's overall health including general health, daily activities, and women's health topics.
- **Personal and Family Medical History –** explores past medical history, including medical conditions and approximate age of diagnosis.
- **Social Determinants of Health -** asks about the social determinants of health, including a participant's neighborhood, social life, stress, and feelings about everyday life.
- **Health Care Access and Utilization -** asks questions about a participant's access to and use of health care.
- **COVID-19 Participant Experience -** asks about the impact of COVID-19 on a participant's mental health, well-being, and everyday life.

***Baseline Survey**

University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

Data Elements - Surv

- **The Basics* -** ba
  questions, including
  participant's work an
- **Lifestyle* -** asks a
  of tobacco, alcohol, a
- **Overall Health* -**
  about a participant's
  general health, daily
  health topics.

- **Personal and Family Medical**
  – explores past medical history,
  medical conditions and approximate
  gnosis.

eterminants of Health - asks
social determinants of health,
participant's neighborhood, social
and feelings about everyday life.

Care Access and Utilization -
tions about a participant's access to
health care.

19 Participant Experience -
t the impact of COVID-19 on a
's mental health, well-being, and
ife.

Cancer Sites

Bladder	Head and neck
Blood or soft tissue	Kidney
Bone	Lung
Brain	Ovarian
Breast	Pancreatic
Cervical	Prostate
Colon /Rectal	Skin
Endocrine	Stomach
Endometrial	Thyroid
Esophageal	Other
Eye	

***Baseline Survey**

University of Colorado Anschutz Medical Campus

All of Us RESEARCH PROGRAM

Data Elements – Physical Measurements

Longitudinal Component: Possible if provider takes measurements during each visit

Baseline physical Measurements

- physiologic (e.g., blood pressure, heart rate)
- anthropometric (e.g., height, weight, waist and hip circumference) measurements.

Measurements Collection

- Clinical Setting
- Self-reporting from home
- Home visits when needed

Data Elements- Electronic Health Records

Longitudinal Component:
EHR records updated at least Biannually

Current EHR datatypes collected

- Demographics
- Visits
- Diagnoses
- Procedures
- Medications
- Laboratory Visits
- Vital Signs

EHR Collection

- Direct from Health Care Provider Organization (HPO)
- Outside of HPO's, Secure EHR sharing programs are available (Sync for Science, AuORP piloted program)

Data Elements- Personal Health Technology

Longitudinal Component:
Minute-level observations

Digital Health Data Provided by

- Mobile Phones
- Wellness and Fitness Devices
- Other Sensors
- Mobile Apps

Currently Available (Fitbit)

- Heart Rate by Zones
- Heart Rate (Minute-Level)
- Daily Activity Summary
- Activity Intraday Steps (Minute-Level)
- Sleep Daily Summary
- Sleep Level (Sequence by Level)

Data Elements- Biospecimen Collections

Biospecimens to include collection of:

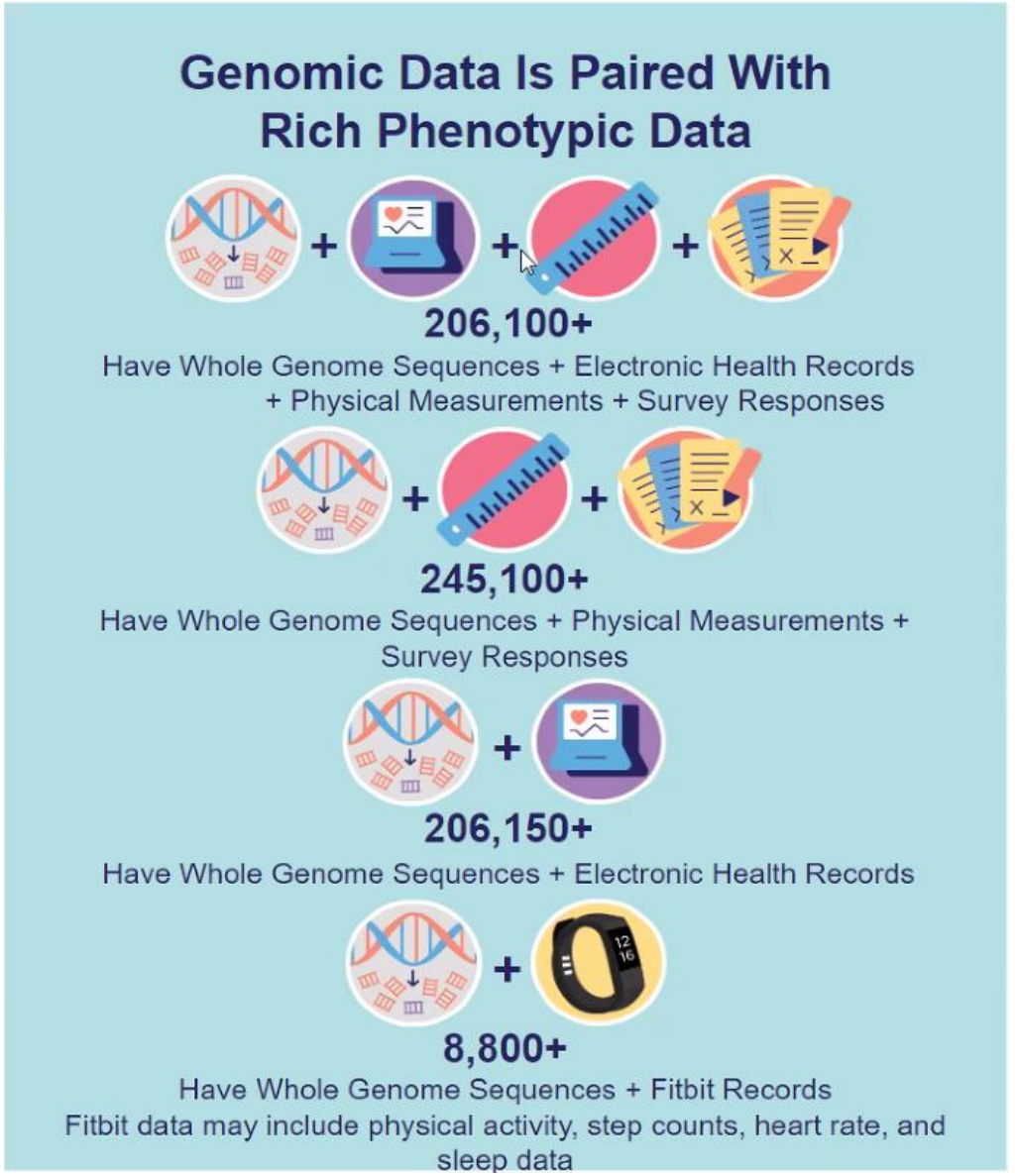
- **Blood**
- **Urine**
- **Saliva**

The objective of the program regarding biospecimens is to collect samples that would allow for the broadest range of clinical and research assays that could be envisioned for the future and to avoid collection, processing, or storage approaches that would inherently preclude such assays

Data Elements- Genomics

Currently, the scope of Genomics data available encompasses 98,500+ whole genome sequencing (WGS) samples and 165,000+ genotyping arrays

Only available at the controlled tier access level



University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

Data Elements- Genomics

All of Us Genomic Data Formats

	srWGS SNP & Indel	srWGS SVs	lrWGS	Array
Raw Data	• CRAM files	• CRAM files	• CRAM files • Graphical Fragment Assembly (GFA) files • FASTA files	• IDAT files
Variant Data	• VariantDataset (VDS) • Variant Call Format (VCF) • Hail MatrixTable • BGEN • PLINK bed files	• Joint Called VCF	• Variant Call Format (VCF) • Hail MatrixTable	• Variant Call Format (VCF) • Hail MatrixTable • PLINK bed files
Auxiliary Files	• Variant Annotation Table • Relatedness • Maximal set of unrelated samples • Ancestry • Limited region callset UCSC BED files • Flagged samples • srWGS Genomic metrics file	Ancestry and relatedness available for srWGS samples based on the srWGS SNP & Indel deliverables	Ancestry and relatedness available for lrWGS samples based on the srWGS SNP & Indel deliverables • lrWGS variant metrics files	Ancestry and relatedness available for array samples that have srWGS data

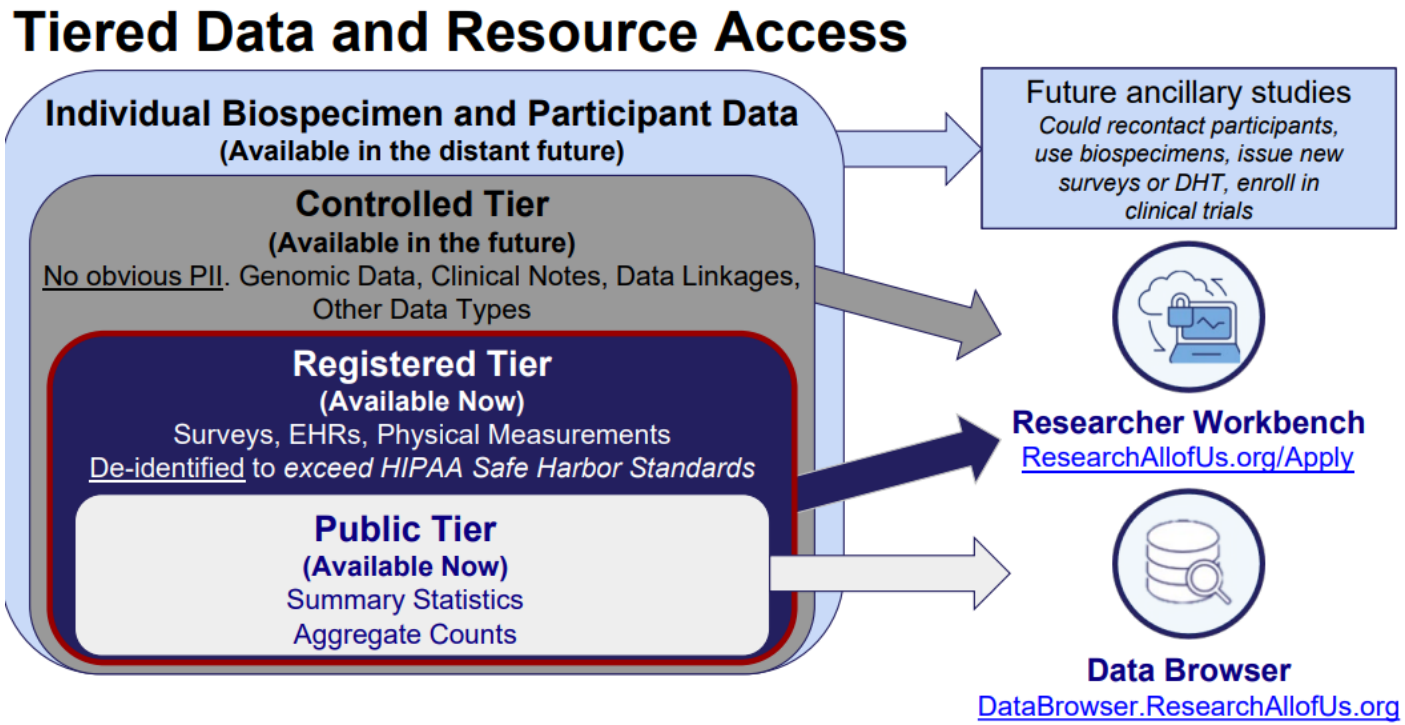
Researchers Prospective

Workbench

Working environment with abilities to:

- **build cohorts & data sets**
- **Perform R/Python data analysis**

Tiers of access



University of Colorado
Anschutz Medical Campus

All of Us RESEARCH PROGRAM

Examples

An Overview of Cancer in the First 315,000 All of Us Participants

Introduction: The NIH All of Us Research Program will have the scale and scope to enable research for a wide range of diseases, including cancer. The program's focus on diversity and inclusion promises a better understanding of the unequal burden of cancer. Preliminary cancer ascertainment in the All of Us cohort from two data sources (self-reported versus electronic health records (EHR)) is considered.

Materials and methods: This work was performed on data collected from the All of Us Research Program's 315,297 enrolled participants to date using the Researcher Workbench, where approved researchers can access and analyze All of Us data on cancer and other diseases. Cancer case ascertainment was performed using data from EHR and self reported surveys across key factors. Distribution of cancer types and concordance of data sources by cancer site and demographics is analyzed.

Results and discussion: Data collected from 315,297 participants resulted in 13,298 cancer cases detected in the survey (in 89,261 participants), 23,520 cancer cases detected in the EHR (in 203,813 participants), and 7,123 cancer cases detected across both sources (in 62,497 participants). Key differences in survey completion by race/ethnicity impacted the makeup of cohorts when compared to cancer in the EHR and national NCI SEER data.

Conclusions: This study provides key insight into cancer detection in the All of Us Research Program and points to the existing strengths and limitations of All of Us as a platform for cancer research now and in the future.

Data Elements used:

Survey- Demographics, Cancer Diagnosis

EHR – Demographics, Cancer Diagnosis

University of Colorado Anschutz Medical Campus

Citation: Aschebrook-Kilfoy, B., Zakin, P., Craver, A., Shah, S., Kibriya, M. G., Stepniak, E., ... & All of Us Research Program Investigators. (2022). An overview of cancer in the first 315,000 All of Us participants. PloS one, 17(9), e0272522.

All of Us RESEARCH PROGRAM

Examples

An Overview of Cancer in the First 315,000 All of Us Participants

Table 2. The relative distribution and prevalence of cancer cases by type in the *All of Us* Research Program from self-reported survey data and electronic health record overall.

	EHR			Survey Data			EHR + Survey		
	N	% dist	prevalence	N	% dist	prevalence	N	% dist	prevalence
Population			203,813			89,261			62,497
Total Cancers	23,520	-	11.54%	13,298	-	14.90%	7,123	-	11.40%
Bladder	983	4.18%	0.48%	483	3.63%	0.54%	301	4.23%	0.48%
Blood	4,841	20.58%	2.38%	1,113	8.37%	1.25%	657	9.22%	1.05%
Bone	350	1.49%	0.17%	181	1.36%	0.20%	107	1.50%	0.17%
Brain	612	2.60%	0.30%	182	1.37%	0.20%	102	1.43%	0.16%
Breast	6,474	27.53%	3.18%	4,062	30.55%	4.55%	2,499	35.08%	4.00%
Cervix	576	2.45%	0.28%	869	6.53%	0.97%	172	2.41%	0.28%
Colon & Rectum	2,601	11.06%	1.28%	722	5.43%	0.81%	385	5.41%	0.62%
Endocrine System	1,887	8.02%	0.93%	129	0.97%	0.14%	63	0.88%	0.10%
Endometrium	1,364	5.80%	0.67%	459	3.45%	0.51%	212	2.98%	0.34%
Esophagus	230	0.98%	0.11%	110	0.83%	0.12%	60	0.84%	0.10%
Eye	123	0.52%	0.06%	66	0.50%	0.07%	28	0.39%	0.04%
Head & Neck	1,698	7.22%	0.83%	333	2.50%	0.37%	155	2.18%	0.25%
Kidney	1,266	5.38%	0.62%	487	3.66%	0.55%	313	4.39%	0.50%
Lung	1,081	4.60%	0.53%	463	3.48%	0.52%	283	3.97%	0.45%
Ovary	786	3.34%	0.39%	348	2.62%	0.39%	207	2.91%	0.33%
Pancreas	548	2.33%	0.27%	119	0.89%	0.13%	77	1.08%	0.12%
Prostate	3,971	16.88%	1.95%	2,165	16.28%	2.43%	1,304	18.31%	2.09%
Stomach	320	1.36%	0.16%	76	0.57%	0.09%	35	0.49%	0.06%
Thyroid	1,648	7.01%	0.81%	924	6.95%	1.04%	573	8.04%	0.92%

*Skin cancer is excluded from the analysis as it is not differentiated as malignant/non-malignant/melanoma in AoU survey.

University of Colorado Anschutz Medical Campus

Citation: Aschebrook-Kilfoy, B., Zakin, P., Craver, A., Shah, S., Kibriya, M. G., Stepniak, E., ... & All of Us Research Program Investigators. (2022). An overview of cancer in the first 315,000 All of Us participants. PloS one, 17(9), e0272522.

All of Us RESEARCH PROGRAM

Examples

An Overview of Cancer in the First 315,000 All of Us Participants

Table 4. Comparison of relative distribution and prevalence of cancer cases by type in the *All of Us* Research Program to SEER's 26-year limited duration prevalence.

	EHR			Survey Data			EHR + Survey			SEER 26-year prevalence		
	N	% dist	prevalence	N	% dist	prevalence	N	% dist	prevalence	N	% dist	prevalence
Population			203,813			89,261			62,497			325,836,757
Total Cancers	23,520	-	11.54%	13,298	-	14.90%	7,123	-	11.40%	14,419,319		4.43%
Bladder	983	4.18%	0.48%	483	3.63%	0.54%	301	4.23%	0.48%	555,999	3.86%	0.17%
Blood	4,841	20.58%	2.38%	1,113	8.37%	1.25%	657	9.22%	1.05%	1,343,512	9.32%	0.41%
Bone	350	1.49%	0.17%	181	1.36%	0.20%	107	1.50%	0.17%	33,086	0.23%	0.01%
Brain	612	2.60%	0.30%	182	1.37%	0.20%	102	1.43%	0.16%	129,633	0.90%	0.04%
Breast	6,474	27.53%	3.18%	4,062	30.55%	4.55%	2,499	35.08%	4.00%	3,096,156	21.47%	0.95%
Cervix	576	2.45%	0.28%	869	6.53%	0.97%	172	2.41%	0.28%	182,868	1.27%	0.06%
Colon & Rectum	2,601	11.06%	1.28%	722	5.43%	0.81%	385	5.41%	0.62%	1,134,250	7.87%	0.35%
Endocrine System	1,887	8.02%	0.93%	129	0.97%	0.14%	63	0.88%	0.10%	70,825	0.49%	0.02%
Endometrium	1,364	5.80%	0.67%	459	3.45%	0.51%	212	2.98%	0.34%	632,326	4.39%	0.19%
Esophagus	230	0.98%	0.11%	110	0.83%	0.12%	60	0.84%	0.10%	21,960	0.15%	0.01%
Eye	123	0.52%	0.06%	66	0.50%	0.07%	28	0.39%	0.04%	~	~	~
Head & Neck	1,698	7.22%	0.83%	333	2.50%	0.37%	155	2.18%	0.25%	396,937	2.75%	0.12%
Kidney	1,266	5.38%	0.62%	487	3.66%	0.55%	313	4.39%	0.50%	451,550	3.13%	0.14%
Lung	1,081	4.60%	0.53%	463	3.48%	0.52%	283	3.97%	0.45%	423,209	2.94%	0.13%
Ovary	786	3.34%	0.39%	348	2.62%	0.39%	207	2.91%	0.33%	167,758	1.16%	0.05%
Pancreas	548	2.33%	0.27%	119	0.89%	0.13%	77	1.08%	0.12%	65,973	0.46%	0.02%
Prostate	3,971	16.88%	1.95%	2,165	16.28%	2.43%	1,304	18.31%	2.09%	3,017,103	20.92%	0.93%
Stomach	320	1.36%	0.16%	76	0.57%	0.09%	35	0.49%	0.06%	96,886	0.67%	0.03%
Thyroid	1,648	7.01%	0.81%	924	6.95%	1.04%	573	8.04%	0.92%	660,323	4.58%	0.20%

*Skin cancer is excluded from the analysis as it is not differentiated as malignant/non-malignant/melanoma in AoU survey.

* SEER data is based on 5-year prevalence frequency counts of 1st invasive tumor.

University of Colorado Anschutz Medical Campus

Citation: Aschebrook-Kilfoy, B., Zakin, P., Craver, A., Shah, S., Kibriya, M. G., Stepniak, E., ... & All of Us Research Program Investigators. (2022). An overview of cancer in the first 315,000 All of Us participants. PloS one, 17(9), e0272522.

All of Us RESEARCH PROGRAM

Examples

An Overview of Cancer in the First 315,000 All of Us Participants

Table 4. Comparison of relative distribution and prevalence of cancer cases by type in the *All of Us* Research Program to SEER's 26-year limited duration prevalence.

	EHR			Survey Data			EHR + Survey			SEER 26-year prevalence		
	N	% dist	prevalence	N	% dist	prevalence	N	% dist	prevalence	N	% dist	prevalence
Population			203,813			89,261			62,497			325,836,757
Total Cancers	23,520	-	11.54%	13,298	-	14.90%	7,123	-	11.40%	14,419,319		4.43%
Bladder	983	4.18%	0.48%	483	3.63%	0.54%	301	4.23%	0.48%	555,999	3.86%	0.17%
Blood	4,841	20.58%	2.38%	1,113	8.37%	1.25%	657	9.22%	1.05%	1,343,512	9.32%	0.41%
Eye	123	0.52%	0.06%	66	0.50%	0.07%	28	0.39%	0.04%	~	~	~
Head & Neck	1,698	7.22%	0.83%	333	2.50%	0.37%	155	2.18%	0.25%	396,937	2.75%	0.12%
Kidney	1,266	5.38%	0.62%	487	3.66%	0.55%	313	4.39%	0.50%	451,550	3.13%	0.14%
Lung	1,081	4.60%	0.53%	463	3.48%	0.52%	283	3.97%	0.45%	423,209	2.94%	0.13%
Ovary	786	3.34%	0.39%	348	2.62%	0.39%	207	2.91%	0.33%	167,758	1.16%	0.05%
Pancreas	548	2.33%	0.27%	119	0.89%	0.13%	77	1.08%	0.12%	65,973	0.46%	0.02%
Prostate	3,971	16.88%	1.95%	2,165	16.28%	2.43%	1,304	18.31%	2.09%	3,017,103	20.92%	0.93%
Stomach	320	1.36%	0.16%	76	0.57%	0.09%	35	0.49%	0.06%	96,886	0.67%	0.03%
Thyroid	1,648	7.01%	0.81%	924	6.95%	1.04%	573	8.04%	0.92%	660,323	4.58%	0.20%

*Skin cancer is excluded from the analysis as it is not differentiated as malignant/non-malignant/melanoma in AoU survey.

* SEER data is based on 5-year prevalence frequency counts of 1st invasive tumor.

	EHR			Survey Data			EHR + Survey			SEER 26-year prevalence		
	N	% dist	prevalence	N	% dist	prevalence	N	% dist	prevalence	N	% dist	prevalence
Population			203,813			89,261			62,497			325,836,757
Total Cancers	23,520	-	11.54%	13,298	-	14.90%	7,123	-	11.40%	14,419,319		4.43%

University of Colorado Anschutz Medical Campus

Citation: Aschebrook-Kilfoy, B., Zakin, P., Craver, A., Shah, S., Kibriya, M. G., Stepniak, E., ... & All of Us Research Program Investigators. (2022). An overview of cancer in the first 315,000 All of Us participants. PloS one, 17(9), e0272522.

All of Us RESEARCH PROGRAM

Examples

Characterizing phenotypic abnormalities associated w/ high-risk individuals developing lung cancer using AoU electronic health records

Data Elements used:

Survey- Demographics, Smoking Status

EHR – Clinical Phenotype Retrieval

Objective: The study sought to test the feasibility of conducting a phenome-wide association study to characterize phenotypic abnormalities associated with individuals at high risk for lung cancer using electronic health records.

Materials and Methods: We used the beta release of the All of Us Researcher Workbench with clinical and survey data from a population of 225 000 subjects. We identified 3 cohorts of individuals at high risk to develop lung cancer based on (1) the 2013 U.S. Preventive Services Task Force criteria, (2) the long-term quitters of cigarette smoking criteria, and (3) the younger age of onset criteria. Logistic regression analysis to identify the associations between individuals' phenotypes and their risk categories. We validated our findings against a lung cancer cohort from the same population and conducted an expert review to understand whether these associations are known or potentially novel.

Results: We found a total of 214 statistically significant associations ($P < .05$ with a Bonferroni correction and odds ratio > 1.5) enriched in the high-risk individuals from 3 cohorts, and 15 enriched in the low-risk individuals. Forty significant associations enriched in the high-risk individuals and 13 enriched in the low-risk individuals were validated in the cancer cohort. Expert review identified 15 potentially new associations enriched in the high-risk individuals.

Conclusions: It is feasible to conduct a phenome-wide association study to characterize phenotypic abnormalities associated in high-risk individuals developing lung cancer using electronic health records. The All of Us Research Workbench is a promising resource for the research studies to evaluate and optimize lung cancer screening criteria.

University of Colorado Anschutz Medical Campus

Citation: Na, J., Zong, N., Wang, C., Midthun, D. E., Luo, Y., Yang, P., & Jiang, G. (2021). Characterizing phenotypic abnormalities associated with high-risk individuals developing lung cancer using electronic health records from the All of Us researcher workbench. Journal of the American Medical Informatics Association, 28(11), 2313-2324.

All of Us RESEARCH PROGRAM

Examples

Characterizing phenotypic abnormalities associated w/ high-risk individuals developing lung cancer using AoU electronic health records

Data Elements used:
Survey- Demographics, Smoking Status
EHR – Clinical Phenotype Retrieval

Risk Group	Case	Control
‘13 USPSTF	2,594	5024
Long-term Quitters of Smoking	990	1,951
Younger age	538	1006
Cancer Cohort	445	507

Risk Group	Significant Associations	Validated
‘13 USPSTF	153	39
Long-term Quitters of Smoking	141	34
Younger age	53	19

Conclusions: It is feasible to conduct a phenome-wide association study to characterize phenotypic abnormalities associated in high-risk individuals developing lung cancer using electronic health records. The All of Us Research Workbench is a promising resource for the research studies to evaluate and optimize lung cancer screening criteria.

University of Colorado Anschutz Medical Campus

Citation: Na, J., Zong, N., Wang, C., Midthun, D. E., Luo, Y., Yang, P., & Jiang, G. (2021). Characterizing phenotypic abnormalities associated with high-risk individuals developing lung cancer using electronic health records from the All of Us researcher workbench. Journal of the American Medical Informatics Association, 28(11), 2313-2324.

All of Us RESEARCH PROGRAM

Examples

Characterizing phenotypic abnormalities associated w/ high-risk individuals developing lung cancer using AoU electronic health records

Table 6. Expert review results for the validated phenotypes

Phecode	Phenotype	Category	Review Results
433.1	Occlusion and stenosis of precerebral arteries	Circulatory system	2
433.11	Occlusion of cerebral arteries, with cerebral infarction	Circulatory system	2
440.9	Atherosclerosis of aorta	Circulatory system	2
443.8	Other specified peripheral vascular diseases	Circulatory system	2
443.9	Peripheral vascular disease, unspecified	Circulatory system	2
681	Superficial cellulitis and abscess	Dermatologic	2
681.3	Cellulitis and abscess of arm/hand	Dermatologic	2
288.2	Elevated white blood cell count	Hematopoietic	2
70.3	Viral hepatitis C	Infectious diseases	2
90	Sexually transmitted infections (not HIV or hepatitis)	Infectious diseases	2
851	Complications of transplants and reattached limbs	Injuries and poisonings	1/2
969	Poisoning by psychotropic agents	Injuries and poisonings	2
318	Tobacco use disorder	Mental disorders	1
296.1	Bipolar	Mental disorders	2
317	Alcohol-related disorders	Mental disorders	2
317.1	Alcoholism	Mental disorders	2
480	Pneumonia	Respiratory	1
480.11	Pneumococcal pneumonia	Respiratory	1
496	Chronic airway obstruction	Respiratory	1
496.21	Obstructive chronic bronchitis	Respiratory	1
506	Empyema and pneumothorax	Respiratory	1
514.2	Solitary pulmonary nodule	Respiratory	1
480.1	Bacterial pneumonia	Respiratory	1/2
480.3	Pneumonia due to fungus (mycoses)	Respiratory	2

1 indicates known association; 2 indicates potentially new association; 1/2 indicates disagreement between reviewers.

Examples

Pharmacogenomic testing & prescribing patterns for patients with cancer in a large national precision medicine cohort

Population databases could help patients with cancer and providers better understand current pharmacogenomic prescribing and testing practices. This retrospective observational study analysed patients with cancer, drugs with pharmacogenomic evidence and related genetic testing in the National Institutes of Health All of Us database. Most patients with cancer (19,633 (88.3%) vs 2,590 (11.7%)) received ≥1 drug and 36 (0.2%) received genetic testing, with a significant association between receiving ≥1 drug and age group ($p<0.001$), but not sex ($p=0.612$), race ($p=0.232$) or ethnicity ($p=0.971$). Drugs with pharmacogenomic evidence—but not genetic testing—were common for patients with cancer, reflecting key gaps preventing precision medicine from becoming standard of care

Data Elements used:

Survey- Demographics

EHR – Cancer Diagnosis, Medications, & Genomic Testing

University of Colorado Anschutz Medical Campus

Citation: Ronquillo, J. G., & Lester, W. T. (2023). Pharmacogenomic testing and prescribing patterns for patients with cancer in a large national precision medicine cohort. Journal of Medical Genetics, 60(1), 81-83.

All of Us RESEARCH PROGRAM

Examples

Socioeconomic and Racial/Ethnic Disparities in Perception of Health Status and Literacy in Spine Oncological Patients

Data Elements used:

Survey- Demographics, Health Status

EHR – Spinal Tumor Identification

OBJECTIVE: The aim of this study was to assess socioeconomic and racial disparities in the perception of personal health, health literacy, and healthcare access among spine oncology patients.

BACKGROUND: Racial, ethnic, and socioeconomic disparities in health literacy and perception of health status have been described for many disease processes. However, few studies have assessed the prevalence of these disparities among spine oncology patients.

METHODS: Adult spine oncology patients, identified using ICD-9/10-CM codes, were categorized by race/ethnicity: White/Caucasian (WC), Black/African-American (BAA), and Non-White Hispanic (NWH). Demographics and socioeconomic status were assessed. Questionnaire responses regarding baseline health status, perception of health status, health literacy, and barriers to healthcare were compared.

RESULTS: Of the 1,175 patients identified, 207 (17.6%) were BAA, 267 (22.7%) NWH, & 701 (59.7%) WC. Socioeconomic status varied among cohorts, with WC patients reporting higher levels of education ($p<0.001$), annual income greater than $50K ($p<0.001$), and home ownership ($p<0.001$). BAA and NWH patients reported greater rates of 7-day “Severe fatigue” ($p<0.001$) and “10/10 pain” ($p<0.001$) and lower rates of “Completely” able to perform everyday activities ($p<0.001$). WC patients had a higher response rate for “Excellent/Very Good” regarding their own general health ($p<0.001$) and quality ($p<0.001$). The WC cohort had a significantly higher proportion of patients responding “Never” when assessing difficulty understanding ($p<0.001$) and needing assistance with health materials ($p<0.001$). BAA and NWH were significantly less likely to report feeling “Extremely” confident with medical forms ($p<0.001$). BAA and NWH had significantly higher response rates to feeling “Somewhat Worried” about healthcare costs ($p<0.001$) and with delaying medical care given “Can’t Afford Co-pay” ($p<0.001$).

CONCLUSION: We identified disparities in perception of health status, literacy, and access among spine oncology patients.

University of Colorado Anschutz Medical Campus

Citation: Elsamadicy, A. A., Wang, C., Reeves, B. C., Sherman, J. J., Craft, S., Rajjoub, R., ... & Sciubba, D. M. Socioeconomic and Racial/Ethnic Disparities in Perception of Health Status and Literacy in Spine Oncological Patients: Insights from the All of Us Research Program. Spine, 10-1097.

All of Us RESEARCH PROGRAM

All of Us: Questions?

Get in touch with the Population Health Shared Resources Team!

Adam.warren@cuanschutz.edu

University of Colorado
Anschutz Medical Campus

All of Us
RESEARCH PROGRAM

All of Us: References

National Institutes of Health *All of Us Research Program Protocol.* 2021.

(2023, June). All of Us Research Hub. https://www.researchallofus.org/