



INDIANA UNIVERSITY
SCHOOL OF NURSING



THE ALL OF US DATASET WITH HEALTHCARE ACCESS/UTILIZATION

Asa B. Smith, PhD, RN | Assistant Professor, Indiana University School of Nursing



#AHA25

DISCLOSURES

- I am not affiliated with nor sponsored by the *All of Us Research Program (All of Us)* or its funding body, the National Institute of Health.
- I take full responsibility for the content and accuracy of this presentation.
- I am an independent researcher sharing my own experiences with using *All of Us*, and your individual experiences/results may vary.
- I have no other conflicts of interest.

OUTLINE

- Overview of *All of Us*
- Accessing and using the data basics
- Healthcare access and utilization
- Advantages and disadvantages



HELPFUL LINKS

The All of Us User Support Hub

- New user orientation
- Using the data browser, cohort builder, and dataset builder
- Understanding the survey, phenotypic and genomic data
- Prerecorded office hours
- Other real-world application presentations
- Links to data dictionary and additional resources

ALL OF US: QUICK FACT SHEET

- Large biomedical data resource.
- *All of Us* aims to enroll one million or more participants with different backgrounds and come from different parts of the country.
- Data types available: surveys, physical measurements, electronic health records, wearables, genomics.
- Currently over 867,000 participants enrolled.

ACCESSING THE DATASET – INITIAL STEPS

- Institution must have a Data Use and Registration Agreement.
 - Search for “All of Us institutional agreements”
 - If not, submit a request to *All of Us*.
- Mandatory training, set up profile in research workbench.
- Controlled tier needed to access sensitive information (e.g., genomic data).

Indiana University



Brenda Lee and Kosali Simon

Key:  Registered Tier

 Controlled Tier

 Individual agreements required by institution

COSTS

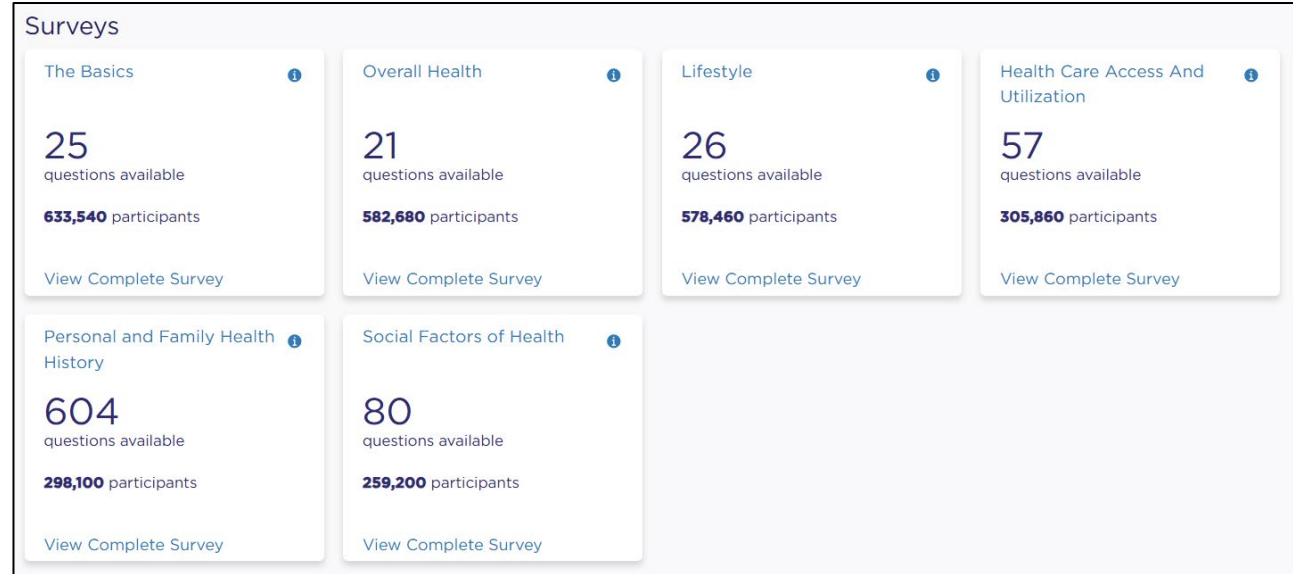
- \$300 of credits – must be used within one year of completing training.
- Costs are depleted based on the cloud computing environment and are different for active analysis vs. persistent disk storage.
- In SAS, example costs for 4 CPUs, 15 GB RAM (non-adjustable), and a 150 GB persistent disk space (adjustable).
 - Active analysis running: \$0.40 per hour
 - Persistent disk storage: \$6 per month
 - Your analysis space will be deleted by default after 8 hours of inactivity, but your disk will stay until you manually delete it.

ANALYSIS BASICS

- All analysis is run using a cloud environment – not local statistical software.
- Processing time for commands can vary substantially.
- A 9/10/2025 initialization of SAS at approx. 9AM EST took 13 minutes.
- Working with datasets can take several minutes per action.
 - Example: Conditions dataset for adults with chronic pain and HF consisted of over **2.8 million** rows of data.
- Do NOT record/write down/store data from the datasets, or even leave the dataset open for a long period of time.
- Do NOT report sample sizes < 20 participants.

AVAILABLE SURVEYS

- Blank surveys can be downloaded so you can view questions.
- Check survey question structure before using.
- Note that surveys have varying sample sizes.



WHICH SURVEYS FOR HEALTHCARE ACCESS/UTILIZATION?

Social Factors

- Education access and quality
 - Economic stability
 - Social and community
 - Neighborhood quality
 - Drug use and crime
 - Companionship/isolation
 - Discrimination and harassment
 - Nervousness and stress
 - Spirituality
- With?

Healthcare Access and Utilization

- Insurance status/acceptance
- Coverage stability over 12 months
- Visits to providers

Important to provide operationalization of measures in AOURP for your research question for transparency/replicability

EXAMPLE

- Keep a record of dataset preparation activities:
 - Transparency
 - You can recreate your datasets and analyses
 - Keep track of how sample sizes change

- Example table identifying how constructs are linked to *All of Us* surveys

Construct	Operationalization	Question (Concept codes from AOURP) Number of items, response scale, response options
Healthcare access	Insurance status	Are you covered by health insurance or some other kind of health care plan? (1585386) ²¹ 1 item, dichotomous (yes/no)
	Insurance acceptance	During the past 12 months, were you told by a health care provider or doctor's office that they did not accept your health care coverage? (43530418) ²⁵ 1 item, dichotomous (yes/no)
	Stability of coverage	In regard to your health insurance or health care coverage, how does it compare to a year ago? Is it better, worse, or about the same? (43530559) ²⁵ 1 item, ordinal. Response options: <ul style="list-style-type: none"> • Worse • About the same • Better
	Self-reported access barriers	There are many reasons people delay getting medical care. Have you delayed getting care for any of the following reasons in the past 12 months? (43529906, 43530268, 43530594, 43529905, 43529903, 43529904, 43530583, 43530585, 43530584) ²⁵ 1 item, ordinal † Response options: 0-9 barriers including <ul style="list-style-type: none"> • Didn't have transportation, • Lived in a rural area where distance was too far • Was nervous about seeing a provider • Couldn't get time off work • Couldn't get childcare • Couldn't get adult care • Couldn't afford copay • Deductible was too high, or could not afford the deductible • Had to pay out of pocket

EXAMPLE

- Flowsheet of dataset preparation documenting:
 - Data sources
 - Inclusion/exclusion criteria
 - Changing sample sizes

AOURP dataset v.8 with “The Basics”, “Overall Health” and “Healthcare Access/Utilization” surveys available AND medical records available
N = 195,261

- Survey data (Date of birth, natal sex, race, pain severity, fatigue, healthcare-related SDOH)
- Medical record data (chronic HF, chronic pain, comorbidity, depression)

Inclusion criteria applied to specify cohort

- Diagnosis of chronic heart failure in medical record
AND
- Diagnosis of chronic pain in medical record

AOURP reduced dataset

N = 7,644

Exclusion criteria applied during dataset preparation (N = 3,711)

- No historical medical records (n = 836)
- Skipped/missing survey questions (n = 2,875)

AOURP final dataset

N = 3,933

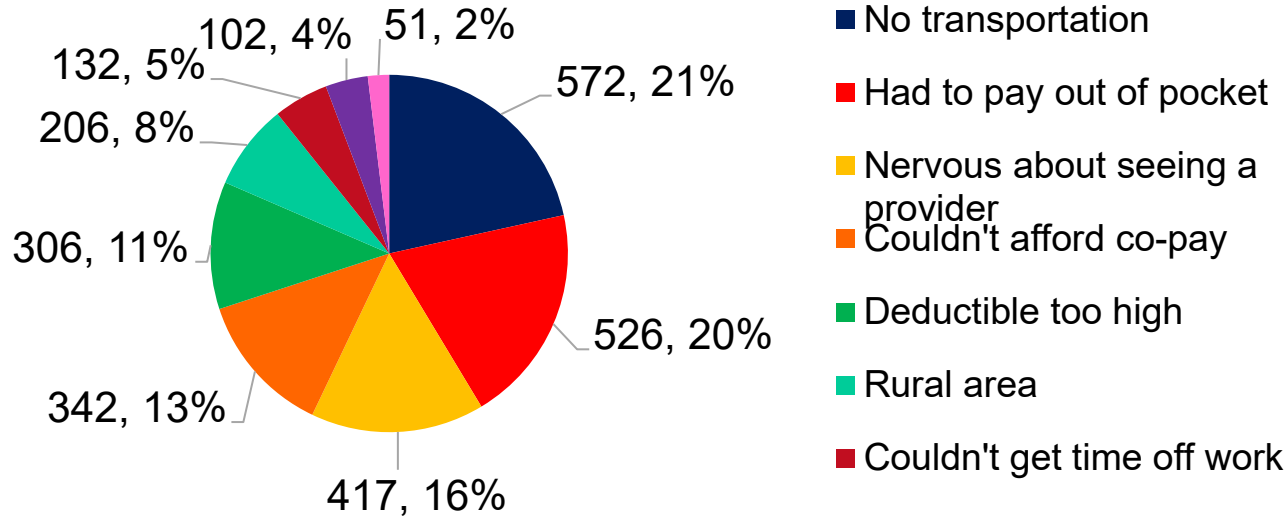
MISSING DATA

- Responses that may need to be treated as missing
 - “None indicated”
 - “PMI: Prefer Not to Answer”
 - ”PMI: Don’t Know”
 - “PMI: Skip”
 - “no matching concept”
- Generate missing data report
- Discussing with your statistician

	Race					Natal sex		TOTAL
	Asian	Black	> 1 race	Other	White	Female	Male	
Pain severity	5.32%	7.93%	3.17%	5.51%	4.00%	4.87%	4.64%	4.98%
Insurance status	0.00%	1.75%	1.27%	2.68%	1.00%	1.32%	1.30%	1.47%
Insurance acceptance	1.06%	2.40%	1.59%	2.83%	1.53%	1.88%	1.72%	1.87%
Stability of coverage	0.00%	4.24%	3.81%	5.80%	2.00%	3.46%	1.93%	2.94%
Self-reported barriers	1.06%	3.32%	2.54%	2.98%	1.58%	2.42%	1.65%	2.20%
Unaffordable services	1.06%	3.14%	1.90%	3.57%	1.53%	2.22%	1.83%	2.14%
Money-saving activities	2.13%	2.68%	2.22%	2.53%	0.73%	1.37%	1.26%	1.44%
General provider visits	5.32%	6.18%	3.17%	5.51%	4.29%	5.05%	4.21%	4.83%
Advance practice visits	9.57%	17.34%	13.97%	15.92%	13.35%	14.98%	13.63%	14.51%
Specialist visits	11.70%	23.62%	22.22%	22.62%	22.38%	22.07%	23.36%	22.91%
Traditional healer visits	10.64%	21.68%	22.22%	19.64%	22.62%	21.14%	23.53%	22.46%

EXAMPLE RESULTS

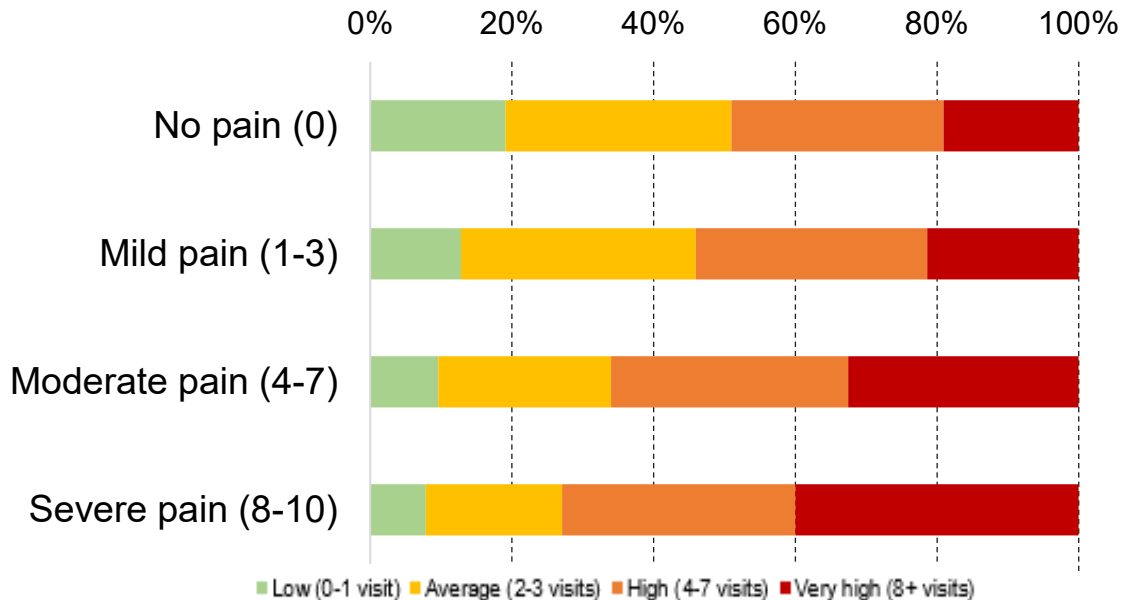
Reasons for delaying care by patients with heart failure and chronic pain



Note. This variable was recoded to “number of self-reported access barriers”

EXAMPLE RESULTS

General Provider Visits over previous 12 months by pain severity among patients with heart failure and chronic pain



- Provider visits are ordinal categories consisting of more than one visit frequency
 - 0 visits
 - 1 visit
 - 2-3 visits
 - 4-5 visits
 - 6-7 visits
- Provider visits are general, limited ability to determine cause for visit

ADVANTAGES AND DISADVANTAGES

Advantages

- Wide range of available data for “big picture” questions.
- Variety of statistical and other resources for using *All of Us*.
- Data are available for free using credits.
- Variables/questions are easily trackable.

Disadvantages

- Participants will be lost if combining multiple surveys.
- Strain on the cloud software can complicate analyses.
- Some variables need to be calculated in statistical software (e.g. age), or recoded (e.g., chronic conditions).

THANK YOU



Scientific
Sessions

#AHA25

Asa B. Smith, PhD, RN
Assistant Professor, IU School of Nursing
asasmit@iu.edu