



Published in final edited form as:

Nat Med. 2023 February ; 29(2): 283–286. doi:10.1038/s41591-022-02171-w.

American Life in Realtime: a benchmark registry of health data for equitable precision health

Ritika R. Chaturvedi^{1,‡}, Marco Angrisani², Wendy M. Troxel³, Tania Gutsche², Eva Ortega², Monika Jain⁴, Adrien Boch⁴, Arie Kapteyn²

¹Leonard D. Schaeffer Center for Health Policy & Economics, University of Southern California, 635 Downey Way, Verna & Peter Dauterive Hall (VPD), Los Angeles, CA

²Center for Economic and Social Research, University of Southern California, 635 Downey Way, Verna & Peter Dauterive Hall (VPD), Los Angeles, CA

³RAND Corporation, Division of Social and Economic Wellbeing, 4570 Fifth Avenue, Pittsburgh, PA 15213

⁴Evidation Health, Inc., 63 Bovet Road #146, San Mateo, CA 94402

Social, structural, and environmental determinants of health, like food or housing insecurity, systemic racism, or chronic stress, account for 60-80% of the modifiable risk of disparities in marginalized populations.^{1,2} Such determinants have been difficult to systematically address due to their complexity, multidimensionality, and heterogeneity (Fig 1). Emerging precision health methods use large-scale person-generated health data from smartphones and wearables to better characterize and ultimately improve health and wellbeing through strategies customized to individual context and need.^{3,4} Applying artificial intelligence and machine learning to person-generated health data allows for unprecedented assessment of recursive, networked, and latent associations between everyday life and health, including social, structural, and environmental exposures, behaviors, biometrics, and health outcomes. As such, precision health provides an important opportunity to reduce health disparities amongst minoritized racial or ethnic groups, or those who are under-resourced.

Despite the potential to improve health equity, the research community lacks benchmark training datasets of person-generated health data, limiting the ability to develop precision health models that are equally effective across diverse populations. Both validity and generalizability of an artificial intelligence or machine learning system are intrinsically tied to the underlying training data. The ideal benchmark dataset should feature high quality, well-characterized data that comprehensively represent the target population in order to instill the highest standards of scientific transparency and rigor to model development, validation, and evaluation.

[‡]Corresponding Author, requests for reprints; ritikac@usc.edu (213-740-7715).

COMPETING INTERESTS:

The authors have no financial, personal, or professional competing interests that have influenced the content of the work described herein.

Person-generated health data cohorts within the National Institutes of Health's All of Us, UK Biobank, the Framingham Heart Study, and the majority of commercial studies rely on convenience sampling and/or bring-your-own-device designs. Consequently, those who lack access to digital technologies, who tend to be older, Black, Latino, Indigenous, poorer, and sicker, are systematically underrepresented.^{5,6} The National Health and Nutrition Examination Survey is representative, but employs a cross-sectional design and a 1-week accelerometer measurement period, which limits the ability to assess temporal effects or account for seasonality. The absence of a benchmark dataset risks introducing systemic bias, exacerbating health disparities, and causing additional patient harm in already marginalized groups.^{7,8}

To bridge this critical gap, we created American Life in Realtime (ALiR), a publicly available benchmark registry and research infrastructure for person-generated health data. ALiR has four primary objectives that advance equitable precision health: promoting inclusive representation; encouraging methodological rigor in artificial intelligence and machine learning; fostering interdisciplinary collaboration and transparency; and, facilitating comprehensive exploration of the dynamic interplay between everyday life and health. We highlight several design choices to achieve these objectives, as well as precision health uses for the data and infrastructure.

We employed several strategies to ensure that the enrolled ALiR cohort (n=1,038) would achieve inclusive representation of the adult U.S. population across demographic, socioeconomic, and health factors.⁹ Participants were invited from the Understanding America Study,¹⁰ an established, probability-based survey panel whose members are randomly sampled from all U.S. addresses. To reduce digital inclusion barriers, we provided a Fitbit Inspire 2 to all participants as a study incentive and a 4G Samsung Galaxy Tablet to those who would not otherwise have internet access. We designed the participant study app to be compatible with a wide range of mobile devices and operating systems and maintained a helpdesk for technical support. As factors beyond digital inclusion, like mistrust or privacy concerns, could lower participation among historically marginalized groups, we also oversampled individuals who were Black, American Indian, Alaska Native, Hawaiian, Pacific Islander, mixed race, Hispanic or Latino, and people whose education was lower than a bachelor's degree.

These ALiR features offer several advantages. Probability sampling improves accuracy and validity of population-level inference such as generalizable predictions of health outcomes in response to population-level stressors like current events, systemic racism, natural disasters, or surges in SARS-CoV-2 cases. Hardware provisioning, in our experience, eliminated sociodemographic disparities in participation rates. Oversampling resulted in a proportionally larger sample of historically underrepresented and marginalized populations,* providing statistical power to detect subgroup-specific differences, such as heterogeneity in outcomes experienced by Black and Latino individuals.

*Weights are provided to rebalance the sample to match the demographic composition of the general U.S. adult population.

To encourage artificial intelligence and machine learning methodological rigor, we designed a comprehensive data collection strategy with validated, longitudinal measures (labels) of participant exposures, behaviors, and outcomes over long time-scales (Table 1). Measures are derived from high-quality consensus instruments such as University of Michigan's Health and Retirement Study, Patient Reported Outcomes Measurement Information System, and the National Institutes of Health's Consensus Measures on Phenotypes and eXposures. A custom mobile app facilitates Fitbit integration, fields electronic surveys, deploys push notification announcements and reminders, and incentivizes long-term engagement through earned points, which are redeemable as monetary compensation.

We created a flexible and sustainable infrastructure to foster growth, agility, collaboration, and transparency. ALiR is both a data resource and research platform, which can encourage agile and cost-efficient community collaboration. Data collected through ALiR will be available to registered users through the Understanding America Study website upon curation of each cohort-year of data, with year 1 anticipated in mid-2023. With appropriate privacy and data security safeguards, freely available data encourages transparency, reproducibility, and explainability of outputs from statistical, artificial intelligence, and machine learning analyses.

The research platform can accommodate additional features, including: participants; application programming interfaces from wearables, medical devices, Internet of Things, genomics and biomarkers; survey designs including preloaded information, randomization, experiments, and ecological momentary assessments; and interactive communications like notifications and visual dashboards. ALiR seeks to achieve a truly large-scale sample through leveraging the ongoing expansion of the Understanding America Study, which is expected to reach 20,000 participants by 2025, and by incorporating special populations like those with specific diseases, which are housed at academic and industry partners. Accordingly, the code will be open-sourced to encourage harmonizable data collection efforts by others.

The ultimate aim of ALiR is to facilitate broad, multidisciplinary, and equitable investigation of precision health. Engineers may leverage the platform to test performance of new hardware or sensors in diverse populations. Methodologists may characterize factors that drive or amplify selection biases, such as social and structural patterning of study participation and attrition, data quality and missingness as well as develop and test solutions to minimize their impact, such as incentive designs and imputation techniques. Social scientists may investigate social determinant clustering and importance in different populations to prioritize public health investments. Behavioral researchers may develop just-in-time interventions where deviations from individual-specific baselines trigger automated nudges and/or suggestions, such as passive detection of influenza-like symptoms via Fitbit data, which triggers a SARS-CoV-2 testing recommendation. Operations researchers may evaluate utility of case worker or health system integrations.

Collectively, ALiR is a model for achieving diversity, equity, inclusion, transparency, and multi-disciplinary collaboration in precision health.

ACKNOWLEDGEMENTS:

We are grateful to the participants who agreed to contribute their time and digital data for research. We thank Evidation Health, Inc. for study app development. This work was funded by a grant from the National Library of Medicine (NLM) at the NIH awarded to Dr. Chaturvedi (R01LM013237). ALiR relies on data from survey(s) administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Hood CM, et al. *American Journal of Preventive Medicine* 50(2):129–135. [PubMed: 26526164]
2. Bradley EH, et al. 2011. *BMJ Quality and Safety in Health Care* 20(10):826–831.
3. Collins Francis S., et al. *New England Journal of Medicine* 372.9 (2015): 793–795. [PubMed: 25635347]
4. Khoury MJ, et al. *JAMA* (2016) 316(13):1357–8. 10.1001/jama.2016.12260 [PubMed: 27541310]
5. Cho Peter Jaeho, et al. *JMIR mHealth and uHealth* 10.4 (2022): e29510. [PubMed: 34913871]
6. Pratap Abhishek, et al. *NPJ Digital Medicine* 3.1 (2020): 1–10. [PubMed: 31934645]
7. Zou J, Schiebinger L, *Nature* 559, 324–326 (2018). [PubMed: 30018439]
8. Obermeyer Ziad, et al. *Science* 366.6464 (2019): 447–453. [PubMed: 31649194]
9. USC Schaeffer Center, 2022. [Online] USC Schaeffer. Available at: <<https://healthpolicy.usc.edu/article/biomedical-expert-designs-health-dataset-representing-all-demographic-groups>> [Accessed 29 August 2022].
10. Alattar Laith, et al. *Social Security Bulletin*, Vol. 78 No. 2, 2018.



Figure 1.
A framework for social, structural, and environmental determinants of health.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

ALiR's Individual-level Data.

	Electronic surveys (repeated every 1-2 years)	Electronic surveys (repeated every 13 months)	Fitbit (continuous)	Public Datasets (continuous)	
Physical Environment	Location	Zip Code / Census Region		Weather	
	Neighborhood	Urbanicity		Air / Water Quality	
		Features / Built Environment	Perceived Safety		Crime
		Transportation / Walkability	Police Presence		
Social Environment	Housing	Type / Characteristics			
	Network	Social Network	Household		
			Marital Status / Family Structure		
			Social Support (PROMIS)		
Context	Culture / Religiosity / Politics	Hope (PhenX)			
	Stressors	Racism and Discrimination	Loneliness		
			Childhood Trauma (ACEs)		
			Stress (PSS-4)		
Economic Environment	Education and Competencies	Education		Fitbit Engagement	
		Financial Literacy		Fitbit Proficiency	
		Digital Literacy			
		Health Literacy			
	Work and Finance	Wealth / Debt (HRS)	Labor Force Status / Occupation		
		Financial Wellbeing / Readiness	Income		
		Housing Stability	Food Security (USDA)		
	Healthcare	Insurance	Access		
			Quality / Satisfaction		
			Age		
Intrinsic Factors	Demographics and Personality	Personality (Big-5)			
		Decision Making	Race / Ethnicity		
		Risk Tolerance	Sex / Gender Identity		
		Mood / Affect	Sexual Orientation		
		Life Satisfaction	Citizenship		
			Immigration Generation		

		Electronic surveys (repeated every 1-2 years)	Electronic surveys (repeated every 13 months)	Fitbit (continuous)	Public Datasets (continuous)
Health Behaviors			Birthplace		
			Primary Language		
			Health Motivations		
	Exercise	Physical Activity (HRS)		Active / Sedentary Time	
				Step Count	
				Distance Covered	
	Diet		Diet (BRFSS)	Walking Speed	
			Sugar Consumption		
	Sleep		Sleep Hygiene (PROMIS)	Time in Bed	
				Sleep / Wake Onset	
Health Outcomes				Napping	
	Substance Use		Drugs / Alcohol (ASSIST)		
	Healthcare	Prevention / Screening (HRS)	Utilization (HRS)		
		Medical Mistrust (PhenX)			
		Vaccine Hesitancy (PhenX)			
	Health Status	Self-reported Health (HRS)	General Health (BRFSS)		
	Physical Health	Chronic Condition Diagnoses (HRS)	Functional Ability (iADL / ADL)	Physical Health	
		Pregnancy (PhenX)	Body Mass Index	Resting / Active Heart Rate	
	Sleep Health	Sleep Disorders (HRS)	Influenza-like Illness (PhenX)	Circadian Rhythm	
			Sleep Disturbance (PROMIS)	Sleep Duration	
				Sleep Efficiency / Latency	
	Mental Health	Mental Illness Diagnoses (HRS)	Depression (PHQ-8)	Sleep Stages	
	Cognitive Health		Anxiety (GAD-7)		
		Substance Abuse (ASSIST)			
	Cognitive Status / Ability				
	Episodic Memory Test Scores				
	Mental Function Test Scores				

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Note that this list is non-exhaustive. Core Understanding America Study survey measures, repeated quarterly for sociodemographics and at least biennially for other variables, are available from 2014. For ALiR, survey measures are repeated monthly and minute-level data from wrist-worn Fitbits are collected continuously. ALiR activities occur for at least 1 year, though participants may subsequently choose to contribute indefinitely. Contextual data from public datasets are matched geospatially and temporally to participant timelines. Specific instruments are in parentheses. HRS = University of Michigan's Health and Retirement Study; PROMIS = Patient Reported Outcomes Measurement Information System; PhenX = National Institutes of Health's Consensus Measures for Phenotypes and Exposures. ACEs = Adverse Childhood Events Scale; PSS-4 = Perceived Stress Scale-4; USDA = U.S. Department of Agriculture; (i)ADL = (instrumental) Activities of Daily Living; ASSIST = World Health Organization Alcohol, Smoking, and Substance Involvement Screening Test; BRFSS = Behavioral Risk Factor Surveillance System; GAD-7 = General Anxiety Disorder-7; PHQ-8 = Patient Health Questionnaire-8.