Pediatric data from the All of Us research program: demonstration of pediatric obesity over time

Nicholas P. Giangreco 1,2, Sulieman Lina3, Jun Qian3, Aymone Kuoame4, Vignesh Subbian 5,6, Eric Boerwinkle7, Mine Cicek8, Cheryl R. Clark 9, Elizabeth Cohen10, Kelly A. Gebo11, Roxana Loperena-Cortes12, Kelsey Mayo12, Stephen Mockrin13,14, Lucia Ohno-Machado15, Sheri D. Schully13, Nicholas P. Tatonetti1,2, and Andrea H. Ramirez13,16

1Department of Biomedical Informatics, Columbia University, New York, New York, USA, 2Department of Systems Biology, Columbia University, New York, New York, USA, 3Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA, 4Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA, 5Department of Biomedical Engineering, The University of Arizona, Tucson, Arizona, USA, 6Department of Systems & Industrial Engineering, The University of Arizona, Tucson, Arizona, USA, 7School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA, 8Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA, 9Department of Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, USA, 10Hunter-Bellevue School of Nursing, Hunter College City University of New York, New York, New York, USA, 11Bloomberg School of Public Health, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA, 12Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA, 13All of Us Research Program, National Institutes of Health, Bethesda, Maryland, USA, 14Leidos, Inc, Frederick, Maryland, USA, 15Department of Biomedical Informatics, UCSD Health, La Jolla, California, USA, and 16Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Corresponding Author: Andrea H. Ramirez, All of Us Research Program, National Institutes of Health, 6710B Rockledge Dr, 4th Floor, Bethesda, MD 20817, USA; andrea.ramirez@nih.gov

Received 24 August 2021; Revised 17 November 2021; Editorial Decision 3 December 2021; Accepted 15 December 2021

ABSTRACT

Objective: To describe and demonstrate use of pediatric data collected by the All of Us Research Program.

Materials and Methods: All of Us participant physical measurements and electronic health record (EHR) data were analyzed including investigation of trends in childhood obesity and correlation with adult body mass index (BMI).

Results: We identified 19,729 participants with legacy pediatric EHR data including diagnoses, prescriptions, visits, procedures, and measurements gathered since 1980. We found an increase in pediatric obesity diagnosis over time that correlates with BMI measurements recorded in participants’ adult EHRs and those physical measurements taken at enrollment in the research program.

Discussion: We highlight the availability of retrospective pediatric EHR data for nearly 20,000 All of Us participants. These data are relevant to current issues such as the rise in pediatric obesity.

Conclusion: All of Us contains a rich resource of retrospective pediatric EHR data to accelerate pediatric research studies.

Key words: electronic health records, retrospective studies, pediatrics, public health informatics, medical informatics

Published by Oxford University Press on behalf of the American Medical Informatics Association 2021. This work is written by a US Government employee and is in the public domain in the US.
BACKGROUND AND SIGNIFICANCE

The All of Us Research Program is developing one of the largest and most diverse patient cohorts for precision medicine research. The program includes prospective new data gathered through surveys and digital health technology as well as retrospective data from electronic health records (EHRs). As of December 2020, the All of Us Researcher Workbench includes multiple data types from nearly 316,000 participants, of whom 55% have any EHR data, including physical measurements (PMs) at enrollment, made available from 41 unique hospital sites. Observational analyses of such retrospective datasets promote evidence-based medicine.

All of Us participants may consent to share EHR data at enrollment. When participants consent, they are invited to provide biospecimens. If participants do not consent, they will not be invited to provide biospecimens but can still provide data via health surveys. In some cases, these data may extend into childhood. These data, collected over multiple decades across the United States, provide an opportunity for pediatric researchers to perform retrospective studies including lifestyle and later-life traits. Moreover, the wealth and diversity of data addresses limitations in child health studies, such as limited enrollment of children in clinical trials and inadequate phenotyping of pediatric diseases. The large amount of data from participant EHRs over 3 decades may also improve our understanding of pediatric disease and treatment where evidence cannot be obtained from randomized clinical trials alone. Already, large and diverse sources of pediatric data, such as from PEDSnet, supports pediatric research to phenotype pediatric diseases and assess treatment outcomes. The All of Us observational pediatric cohort data may further accelerate detecting serious pediatric outcomes and elucidating the mechanisms of pediatric disease.

OBJECTIVES

We report the collection of pediatric data from 19,729 participants across the country over 3 decades. We identified characteristics of participants who had pediatric data by age, race, sex, and ethnicity. We evaluated the data through reproducing the known rise in pediatric obesity over time, which has become more important during the COVID-19 pandemic. We evaluated the relationship between pediatric obesity diagnosis and other EHR and health data, including participant body mass index (BMI) and health survey responses at adult enrollment. Multimodal pediatric obesity evaluation was only possible through the large-scale, comprehensive health measures provided by participants in All of Us.

MATERIALS AND METHODS

All of Us participants

The goals, recruitment methods and sites, and scientific rationale for the All of Us Research Program have been described previously. Demonstration projects were designed to describe the cohort, replicate previous findings for validation, and avoid novel discovery in line with the program value to ensure equal access by researchers to the data. The work described here was proposed by Consortium members, reviewed and overseen by the program’s Science Committee, and was confirmed as meeting criteria for nonhuman subjects research by the All of Us Institutional Review Board. The initial release of data and tools used in this work was published recently. Results reported are in compliance with the All of Us Data and Statistics Dissemination Policy and are only displayed in groups of 20 individuals or more.

All of Us data

This work was performed on data collected by the previously described All of Us Research Program using the All of Us Researcher Workbench (fourth dataset version), a cloud-based platform where approved researchers can access and analyze All of Us data. The All of Us data currently includes surveys, EHR, and PMs. The EHR data from each site were provided from both inpatient and outpatient settings, including primary and acute care. Participants consent to the study and may authorize sharing of the EHRs through an online portal or smartphone application after which they can answer health surveys, share digital health data (such as FitBit and Apple HealthKit), and can view their study information. Through in-person visits, participants can be invited to contribute biospecimens and undergo PMs. The details of the surveys are available in the Survey Explorer found in the Research Hub, a website designed to support researchers. Each survey includes branching logic and all questions are optional and may be skipped by the participant. PM recorded at enrollment include systolic and diastolic blood pressure, height, weight, heart rate, waist and hip measurement, wheelchair use, and current pregnancy status.

EHR data for participants who consented to share their EHR was extracted from their healthcare organization, mapped, and de-identified. All 3 datatypes (survey, PM, and EHR) are mapped to the Observational Health and Medicines Outcomes Partnership common data model v 5.2 maintained by the Observational Health and Data Sciences Initiative collaborative. To protect participant privacy, a series of de-identification algorithms were applied. These included data suppression of codes with a high risk of re-identification such as military status; generalization of categories, including age, sex at birth, gender identity, sexual orientation, and race; and date shifting by a random (less than 1 year) number of days, implemented consistently across each participant record. Documentation on privacy implementation and creation of the data is available in the All of Us Registered Tier Data Dictionary. All of Us participants represent a diverse patient population, demographically and geographically. The Researcher Workbench currently
offers tools with a user interface built for selecting groups of participants (Cohort Builder), creating datasets for analysis (Dataset Builder), and Workspaces with Jupyter Notebooks (Notebooks) to analyze data. The Notebooks enable use of saved datasets and direct query using R (version 4.0.2) and Python 3 programming languages.

Analysis methods
All participant data analyzed were consented per the above protocols. We identified clinical events (also called concepts) within EHR data based on standardized domains or types in the common data model including conditions, drug exposures, visits, procedures, and measurements. We considered pediatric data as any data type recorded after birth (age 0) and before or at 21 years of age (age defined as the difference between the participant’s date of birth from the starting date or the date recorded in the EHR for the concept) within child developmental stages defined by the Eunice Kennedy Shriver National Institute of Child and Human Development.17 We considered adult data as any data type recorded after 21 years of age.

We defined trends for clinical concepts per EHR datatype per year across participants. We first calculated a standardized z-score:

\[ z = \frac{x - \mu}{\sigma} \]

where \( x \) was the percent of a concept or the number of participants with a recorded concept within a year and \( \mu \) was the average percent across concepts for a year, and \( \sigma \) was the sample standard deviation of the percent across concepts. We then calculated a linear model to estimate the association of concept z-scores across time. We quantified the slope and \( R^2 \) squared coefficient between the z-score and a year, across all years where EHR data was provided. This generated a beta coefficient for each year representing a trend in clinical concepts, relative to other EHR concepts, recorded in participant’s EHRs. Significant trends were defined as the linear model beta coefficient greater than 0, beta coefficient confidence interval not containing the null association, and \( R^2 \) squared coefficient between the date and z scores greater than 0.8. We performed summation, visualization, and statistical analyses using R packages including tidyverse and data.table and Python 3 libraries Numpy, Matplotlib, Pandas, Sklearn, and Seaborn.

Obesity in pediatrics and adulthood
To identify BMI data, we used EHR direct recorded measurements of BMI and BMI derived from the formula kg/m² (Table S1). We also calculated BMI from height and weight measurements taken at participant enrollment. We filtered out implausible height, weight, and BMI pediatric measurements according to the age and sex stratified thresholds from the Youth Risk Behavior Surveillance System epidemiological survey conducted by the Center for Disease Control (CDC).18

We identified childhood obesity diagnoses using condition concepts Obesity and Morbid obesity (Table S1) and then classified participants into 3 groups: obesity not diagnosed (obesity diagnosis code not present in the data), obesity diagnosed, and morbid obesity diagnosed, where participants who did not have one of these obesity codes were placed in the obesity not diagnosed group. We evaluated associations between these childhood obesity diagnosis categories with the program-obtained PM BMI using Student’s t test.

We classified each BMI value from the pediatric EHR into 4 categories: underweight, normal weight, overweight, and obese, according to the age- and sex-matched CDC definition.19 We assigned each participant one obesity category based on their most frequent category during childhood, which then persisted through that participant’s adult BMI measurements obesity category. We filtered out implausible BMI adult measurements using the lowest plausible pediatric measurement and an upper threshold of 80 and adult obesity was defined as BMI above 30 according to the CDC definition.20

We made available all data extraction, visualization, and analysis code for replication and reuse by approved researchers as notebooks in the All of Us Researcher Workbench Featured Workspaces.

RESULTS

Pediatric patient EHR characteristics in All of Us
In the fourth release of the Researcher Workbench, 19 729 of the 315 297 total All of Us participants have legacy or pediatric EHR data (Table 1). These patients were significantly younger (Age at enrollment 27.96 [21.67, 50.08]), as expected, when comparing to total All of Us participants (50.55 [21.17, 79.42]; 2-sample Kolmogorov-Smirnov test P-value < .001). We found participants with pediatric data were comprised of more underrepresented groups in biomedical research compared to the overall All of Us population (Table S2).

We examined the diversity of EHR datatypes available from this cohort such as drugs, conditions, measurements, visits, and procedures (Figure 1A). The participant data was drawn from across all 41 sites and the date of participant’s pediatric data within the EHR spans up to 3 decades (Figure 1B). The participant data represented all the different child development stages with participants often containing data over multiple child development stages (Table 2).

Increase in pediatric obesity diagnosis over time
The longitudinal nature of the data motivated temporal trend identification in the pediatric EHR datatypes. Over the past 3 decades

### Table 1. Demographics of All of Us participants with pediatric data.

| All of Us participants with pediatric data | Number of participants | Age at enrollment | Sex at birth | Race | Ethnicity |
|-------------------------------------------|------------------------|-------------------|-------------|------|----------|
|                                           | 19 729                 | 29.46 (7.96)      | Female      | White | Hispanic |
|                                           |                        |                   | Male        | Black | Other    |
|                                           |                        |                   | Other⁵      | Asian | Other⁶   |
|                                           |                        |                   |             | Other⁷ | Other⁸   |
| Female                                   | 11 766 (72.4%)         |                   | 4104 (25.2%)| 388  (2.4%)| 7389 (59.2%)|
| Male                                     | 4104 (25.2%)           |                   | 3641 (29.2%)| 404  (3.2%)| 264  (8.4%)|
| Other⁵                                   | 388  (2.4%)            |                   |             |       |          |
| White                                    | 7389 (59.2%)           |                   | 3641 (29.2%)| 404  (3.2%)| 264  (8.4%)|
| Black                                    | 3641 (29.2%)           |                   | 404  (3.2%)|       |          |
| Asian                                    | 404  (3.2%)            |                   |             |       |          |
| Other⁶                                   | 264  (8.4%)            |                   |             |       |          |
| Other⁷                                   | 264  (8.4%)            |                   |             |       |          |
| Other⁸                                   | 11 780 (73.1%)         |                   |             |       |          |

Note: Number of participants with pediatric data in their electronic health record. The demographic makeup of participants (number of participants and percent of total) including the sex at birth, race, and ethnicity. The age at enrollment is specified as the average (standard deviation) across participants.

1 Comprised of participants with the concept name “GeneralizedDiffSex.”
2 Comprised of participants with concept names “No matching concept,” “skip,” “PreferNotToAnswer,” “GeneralizedPopulation,” “RaceEthnicityNoneOfThese,” and “GeneralizedMultiPopulations.”
3 Comprised of participants with concept names “No matching concept,” “skip,” “PreferNotToAnswer,” and “RaceEthnicityNoneOfThese.”
Figure 1. Description of All of Us EHR pediatric data. A, Venn diagram of participants with pediatric EHR data of different datatypes. Number of pediatric patients with different and overlapping EHR data-types. B, Length of care (ie, participants’ recorded date of the EHR concept from the participants’ date of birth), where vertical minor axes represent 5-year increments, for participants’ pediatric data per hospital site contributing data. Abbreviation: EHR: electronic health record.
(1980–2018), we identified, out of 16,460 EHR concepts across datatypes, 2348 significantly trending concepts, specifically 1299 lab test measurements, 245 diagnoses, 460 drugs, 339 procedures, and 5 visit types, with a significant positive relationship over 3 decades (linear model beta coefficient > 0, confidence interval not containing the null association, \( R^2 > 0.8 \); see supplemental material).

There were 15,431 participants with any EHR diagnosis during childhood. Out of all diagnosis codes, pediatric obesity increased to the third most frequent diagnosis, only behind abdominal pain and pregnancy findings, over time (1980–2018 linear model beta 2.56 [1.59, 2.71]; \( R^2 = 0.88 \); rank 96 out of 245). Also, morbid obesity significantly increased in more recent years (1998–2018 linear model beta 4.42 [3.38, 4.96]; \( R^2 = 0.75 \); Figure 2). Moreover, the obesity and morbid obesity diagnoses codes were the most significantly represented among other obesity-related diagnosis codes (1690 or 93.2% of participants with either obesity or morbid obesity diagnoses codes).

**Table 2. Summary statistics for EHR datatypes in the All of Us participant’s pediatric EHR.**

|          | Conditions | Drug exposures | Procedures | Visits | Measurements |
|----------|------------|----------------|------------|--------|--------------|
| Number of participants | 15,431 | 10,977 | 13,986 | 18,195 | 12,876 |
| Unique concepts | 17.19 (22.90) | 18.36 (25.98) | 12.37 (15.86) | 2.09 (1.14) | 50.01 (43.49) |
| NICHD stages |NICHD stages |NICHD stages |NICHD stages |NICHD stages |NICHD stages |
| Term neonatal (birth to 27 days) | 0.02 (0.12) | 0.02 (0.15) | 0.03 (0.18) | 0.04 (0.20) | 0.02 (0.14) |
| Infancy (28 days to 1 year) | 0.06 (0.23) | 0.12 (0.32) | 0.05 (0.22) | 0.07 (0.26) | 0.02 (0.15) |
| Toddler (1 to 2 years) | 0.06 (0.24) | 0.12 (0.33) | 0.05 (0.22) | 0.07 (0.26) | 0.03 (0.16) |
| Early childhood (2 to 5 years) | 0.09 (0.29) | 0.10 (0.30) | 0.09 (0.28) | 0.10 (0.30) | 0.05 (0.21) |
| Middle childhood (6 to 11 years) | 0.18 (0.39) | 0.17 (0.37) | 0.18 (0.39) | 0.19 (0.39) | 0.11 (0.31) |
| Early adolescence (12 to 18 years) | 0.53 (0.50) | 0.49 (0.50) | 0.54 (0.50) | 0.53 (0.50) | 0.48 (0.50) |
| Late adolescence (19 to 21 years) | 0.88 (0.33) | 0.88 (0.33) | 0.87 (0.34) | 0.89 (0.32) | 0.90 (0.30) |
| Span of NICHD stages | 1.82 (1.24) | 1.90 (1.44) | 1.81 (1.23) | 1.90 (1.37) | 1.60 (0.95) |

Abbreviations: EHR: electronic health record; NICHD: Eunice Kennedy Shriver National Institute of Child and Human Development.

*Mean (SD) unique concepts per datatype for participants.

1Proportion (SD) of participants within NICHD-defined child developmental stages.

2Mean (SD) of developmental stages with concepts per datatype for participants.

**Figure 2.** Increasing pediatric diagnoses over time. Pediatric diagnoses that are significantly trending across the 41 hospital sites over 3 decades of participant’s electronic health records. The significantly trending Obesity (blue) and Morbid obesity (red) diagnoses are highlighted. Diagnoses from the top 10 trending over the 3 decades are shown where “Pregnancy findings” include “Findings related to pregnancy,” “Complications related to pregnancy,” “First trimester pregnancy,” and “Third trimester pregnancy.” The diagnosis trend score is the linear association between the years and concept percentage standardized z score (see Methods).
Table 3. *All of Us* participant EHR diagnoses, EHR BMI measurements, and Program BMI measurements as children and adults.

| Number of participants with any diagnosis as a child (N = 15 431) | Pediatric obesity diagnosis categories | Obesity | Morbid obesity | NA (Absence of obesity diagnosis) |
|---------------------------------------------------------------|--------------------------------------|---------|----------------|----------------------------------|
|                                                               |                                     | 1553 (10.1%) | 433 (2.8%) | 13 741 (89.0%) |
| Number of participants with any diagnosis and at least one BMI measurement as a child (N = 2768) |                                     | 447 (16.2%) | 125 (4.5%) | 2180 (79.3%) |
| Number of participants with any diagnosis as child and at least one Program Physical BMI measurements (N = 15 374) |                                     | 1058 (6.9%) | 268 (1.7%) | 10 636 (91.4%) |

Note: Number of participants and the percent of total diagnosed in the pediatric obesity diagnosis categories as a child. Abbreviation: BMI: body mass index.

within the context of pediatric obesity diagnoses, namely obesity, morbid obesity, and the absence of an obesity diagnosis. Participants varied in the number of recorded BMI measurements in their EHRs, which contained at least 1 but up to 543 measurements (7.4 on average and 95% confidence interval, 1–26). Of the 15 431 participants with any EHR diagnosis during childhood, there were 1553 participants diagnosed with obesity, 433 diagnosed with morbid obesity, and 13 741 without an obesity diagnosis (Table 3). There were 296 participants (17.5%) who had both an obesity (N = 1553) and morbid obesity (N = 433) diagnosis during childhood.

We found 2768 participants had 20 509 BMI measurements in their EHR during childhood. Of these, there were 447 (16.2%) participants who had an obesity diagnosis, 125 (4.5%) participants who had a morbid obesity diagnosis, and 2180 (79.3%) participants who had no obesity diagnosis. We found those diagnosed with obesity had on average higher BMIs than those not diagnosed with obesity (t-test statistic = 80.3, P-value = 0). Furthermore, those diagnosed with morbid obesity had higher BMIs, on average, than those diagnosed with obesity (statistic = 10.3, P-value = 6.97E-25) and without an obesity diagnosis (statistic = 71.6, P-value = 0).

Of the 15 374 *All of Us* participants with both pediatric EHR data and recorded BMI measurements at program enrollment, we found 10 636 participants (69.2%) had no obesity diagnosis, 1058 participants (6.9%) had an obesity diagnosis, and 268 participants (1.7%) had a morbid obesity diagnosis in childhood, respectively. We found those diagnosed with pediatric obesity had on average higher BMI at enrollment than those not diagnosed with obesity during childhood (t-test statistic = 38.6, P-value = 9.77E-307; Figure 3). Furthermore, those diagnosed with morbid obesity had higher BMIs at enrollment than those diagnosed with obesity (t-test statistic = 7.98, P-value = 3.12E-15) and without an obesity diagnosis (t-test statistic = 31.0, P-value = 2.21E-202) during childhood.

**Childhood obesity persists through adulthood in *All of Us* participants**

We next investigated the BMI measurements recorded for *All of Us* participants in their EHRs through adulthood. We found 2735 participants with BMI measurements taken during childhood and adulthood. We stratified participants by sex- and age-matched obesity categories during childhood and viewed its relationship with BMI measurements during adulthood (see Methods). There were 223 underweight, 1034 normal, 380 overweight, and 1098 obese participants according to child-defined categories. We found participant’s survey response to “How old were you when diagnosed with Obesity?” upon enrollment aligned with this categorization: 93% who responded adolescence and 92% who responded child also had an obese BMI measurement during childhood. Interestingly, 73% who responded adult had an obese BMI during adulthood.

Strikingly, obese BMI measurements were the largest proportion of single category BMI measurements for 93.3 or 48.5% of the participants with BMI measurements, followed by all normal, overweight, and underweight BMI measurements (Figure 4A). Considering the longitudinal nature of the data, BMI measurements were found within participant’s EHRs as early as 11 years old and as late as 39 years old. We found 86% of participants with majority obese BMI measurements during childhood were also obese in adulthood, compared to 49%, 16%, and 2% of participants with majority overweight, normal, and underweight BMI measurements (chi-squared test P-values = 3.69E-26, 1.60E-111, and 4.86E-80). This closely aligned with participant surveys where 80% who responded they were diagnosed with obesity as a child and adult also had an obese BMI measurement as an adult. Children in underweight, normal, and overweight classes did not show a trend for obese BMI measurements in adulthood unlike children in the obese class (Figure 4B).

**DISCUSSION**

In this study, we highlight the availability of retrospective pediatric EHR data for nearly 20 000 *All of Us* participants. These participants are from a highly diverse cohort representing those tradition-
Figure 4. *All of Us* participant EHR BMI measurements over time. A, An upset plot showing the number of participants with BMI measurements within obesity categories (represented by each row) based on CDC-defined BMI value thresholds in Pediatric EHRs. The total number of participants for each category are shown at each horizontal bar, and the vertical bars represent the total number of participants that, throughout their care, had BMI measurements in each obesity category. B, Participants with a BMI measurement in pediatric obesity categories, based on CDC-defined BMI value thresholds in Pediatric EHRs, during childhood and with BMI measurements in their EHRs during adulthood (adult obesity marked by red dotted line). The shading represents the 95% confidence interval among the patients. Abbreviation: BMI: body mass index; CDC: Center for Disease Control; EHR: electronic health record.
ally underrepresented in biomedical research.\textsuperscript{16} Representation from different datatypes across child development stages lays a broad foundation for retrospective pediatric research studies and prospective pediatric data capture.

We show that \textit{All of Us} participant’s pediatric EHR data generate hypotheses regarding pediatric clinical care and public health. For example, there are multiple factors that might contribute to observing abdominal pain as the top trending EHR diagnosis over time. Abdominal pain is one the common symptoms that pediatric patients experience.\textsuperscript{25} The increase in ultrasonography for abdominal pain was shown to correlate with increased usage of nonradiating imaging technologies in pediatric care settings over time.\textsuperscript{22} Abdominal pain is a symptom for multiple common diseases, such as Gastroenteritis and Appendicitis. The retrospective pediatric EHR data in \textit{All of Us} can further evaluate factors contributing to abdominal pain as well as other pediatric patient diagnoses and EHR datatypes we observed in our analysis.

Here we showed that the participant’s pediatric EHR data contains observations relevant to current issues such as the rise in pediatric obesity.\textsuperscript{11} Participants’ pediatric EHR data from \textit{All of Us} shows a similar upward trend, consistent with prior epidemiological studies.\textsuperscript{11,23} We corroborate the diagnoses with longitudinal BMI measurements during pediatric care and their association with obesity in adulthood confirming the association between childhood obesity and adulthood obesity found by other studies.\textsuperscript{24,25} Moreover, our study showed an increasing trend in childhood obesity diagnosis and its positive relationship with adulthood obesity, as shown in a recent pooled analysis.\textsuperscript{26} Our analysis is significant since it demonstrates the additional pediatric investigation that researchers can perform on a diverse dataset by replicating known associations and trends and discovering new ones in the \textit{All of Us} pediatric population especially in underrepresented populations.

Our analysis has a few limitations. First, we relied on the presence of diagnosis codes during childhood which can be incomplete or undocumented due to clinical practice variability. Second, our study is limited toward participants who had pediatric EHR data available, which is not consistent across demographic groups. Third, our longitudinal analysis might have a lower sample since some participants moved to another healthcare organization or state, limiting the number of participants who have diagnosis and measurements from childhood through adulthood.

The \textit{All of Us} Research Program cohort is distinct from other datasets such as the UK biobank\textsuperscript{27} that contain data only for adults 18 years or older and the MVP Million Veterans Program\textsuperscript{28} that do not have clinical data retrospective to pediatric years and lack diversity. While \textit{All of Us} is not yet enrolling minors directly, the program is moving in that direction and available EHR will grow over time. Additionally, external data linkages to other sources such as billing codes or pharmacy databases may provide additional retrospective data. Participants also contribute biosamples at enrollment, and both genotyping and whole genome sequence generation is underway and expected to be available for research use in 2022. These data and longitudinal relationship with this highly engaged, diverse population are of high value to pediatric research.

CONCLUSION

The \textit{All of Us} research program, in addition to serving our aging population, has the potential to accelerate pediatric research. The participant’s EHR data is a significant source of longitudinal, diverse pediatric observations that will continue to grow over time. This data source has enormous potential to accelerate evidence-based precision pediatric medicine and inform public health interventions.

FUNDING

The \textit{All of Us} Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549, 1 OT2 OD026554, 1 OT2 OD026357, 1 OT2 OD026356, 1 OT2 OD026350, 1 OT2 OD 026552, 1 OT2 OD026353, 1 OT2 OD026348, 1 OT2 OD026551, 1 OT2 OD026555, IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 26320160008SU; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205, 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277, 3 OT2 OD025315, 1 OT2 OD025337, 1 OT2 OD025276. Mr Giangreco and Dr Tatonetti are supported by National Institute of General Medical Sciences, Grant/Award Number: R35GM131905. Dr Subbian was supported in part by the National Science Foundation under grant #1838745. Dr Gebo initiated this work in her role as the Chief Medical and Scientific Officer for the \textit{All of Us} Research Program. The other authors received no external funding. In addition, the \textit{All of Us} Research Program would not be possible without the partnership of its participants.

AUTHOR CONTRIBUTIONS

Mr Giangreco conceptualized and designed the study, performed the data extraction and summaries, performed the data analysis and wrote the initial manuscript draft. Dr Ramirez, a member of the \textit{All of Us} Research Program Demonstration Projects Subcommittee, conceptualized and designed the study, substantially contributed to the interpretation of data and critically reviewed and critiqued manuscript drafts, and also contributed to the study design by reviewing, providing recommendations, and approving the study concept. Mr Qian and Ms. Kuoame performed the data extraction and summaries. Dr Suleman performed a critical review of the data analysis methods and code and provided recommended revisions to improve study analysis. Dr Subbian and Dr Tatonetti substantially contributed to the interpretation of data and critically reviewed and critiqued manuscript drafts. Dr Boerwinkle, Dr Cieck, Dr Clark, Dr Cohen, Dr Gebo, Dr Loperena-Cortes, Dr Mayo, Dr Mockrin, and Dr Ohno-Machado, as members of the \textit{All of Us} Research Program Demonstration Projects Subcommittee, contributed to the study design by reviewing, providing recommendations, and approving the study concept. All authors listed also reviewed and revised the article, as needed, for important intellectual content. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

SUPPLEMENTARY MATERIAL

Supplementary material is available at \textit{JAMIA Open} online.

CONFLICTS OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article are available in the Researcher Workbench. All researchers must be authorized and approved via a
3-step process that includes registration, completion of ethics training and attestation to a data use agreement. The Researcher Workbench project where this data and code is available can be shared after approval on the platform. The supplemental material can be found at https://doi.org/10.5061/dryad.j0zpc86g3.

REFERENCES

1. Murray J. The ‘All of Us’ Research Program. *N Engl J Med* 2019; 381 (19): 1884.
2. Ramirez AH, Sulieman L, Schluter DJ, et al. The All of Us Research Program: data quality, utility, and diversity. *medRxiv* 2020; 1–39. doi:10.1101/2020.05.29.20116905.
3. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; 216: 574–8.
4. Sutherland SM, Kaelber DC, Downing NL, et al. Pediatric clinical trials. *Pediatr Clin North Am* 2016; 63 (2): 251–68.
5. Bennett TD, Callahan TJ, Feinstein JA, et al. Data science for child health. *J Pediatr* 2019; 208: 12–22.
6. Bavdekar SB. Pediatric clinical trials. *Perspect Clin Res* 2013; 4 (1): 89–99.
7. Mulugeta LY, Yao L, Mould D, et al. Leveraging big data in pediatric development programs: proceedings from the 2016 American College of Clinical Pharmacology annual meeting symposium. *Clin Pharmacol Ther* 2018; 104 (1): 81–7.
8. Goulooze SC, Zweep LB, Vogt JE, et al. Beyond the randomized clinical trial: innovative data science to close the pediatric evidence gap. *Clin Pharmacol Ther* 2020; 107 (4): 786–95.
9. Forrest CB, Margolis PA, Bailey LC, et al. PEDSnet: a national pediatric learning health system. *J Am Med Inform Assoc* 2014; 21 (4): 602–6.
10. Denburg MR, Razzaighi H, Bailey LC, et al. Using electronic health record data to rapidly identify children with glomerular disease for clinical research. *J Am Soc Nephrol* 2019; 30 (12): 2427–35.
11. Skinner AC, Ravanbakht SN, Skelton JA, et al. Prevalence of obesity and severe obesity in US children, 1999-2016. *Pediatrics* 2018; 141 (3): e20173459.
12. Lange SJ, Kompaniyets L, Freedman D, et al. Longitudinal trends in body mass index before and during the COVID-19 pandemic among persons aged 2–19 years—United States 2018-2020. *MMWR Morb Mortal Wkly Rep* 2021; 70 (37): 1278–83.
13. Denny JC, Rutter JL, Goldstein DB; All of Us Research Program Investigators, et al. The ‘All of Us’ research program. *N Engl J Med* 2019; 381 (7): 668–76.
14. All of Us research hub. https://www.researchallofus.org/. Accessed April 23, 2020.
15. Methods – All of Us research hub. https://www.researchallofus.org/. Accessed April 23, 2020.
16. Mapes BM, Foster CS, Kusnoor SV; the All of Us Research Program, et al. Diversity and inclusion for the All of Us research program: a scoping review. *PLoS One* 2020; 15 (7): e0234962.
17. Williams K, Thomson D, Seto I, et al. Standard 6: age groups for pediatric trials. *Pediatrics* 2012; 129 (Suppl 3): S153–60.
18. Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance System 2013—YRBS data user’s guide; 2014. http://www.cdc.gov/yrbs. Accessed April 26, 2021.
19. CDC. Defining childhood obesity. *Overweight & Obesity*. https://www.cdc.gov/obesity/childhood/defining.html. Accessed January 29, 2021.
20. CDC. Defining adult overweight and obesity. *Overweight & Obesity*. https://www.cdc.gov/obesity/adult/defining.html. Accessed January 29, 2021.
21. Cow P, Aajim H-K, Jam H, et al. Pain in children and adolescents: a common experience. *Pain* 2000; 87: 51–8.
22. Marin JR, Rodean J, Hall M, et al. Trends in use of advanced imaging in pediatric emergency departments, 2009-2018. *JAMA Pediatr* 2020; 174 (9): e202209.
23. World Health Organization. Obesity and overweight. https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight. Accessed January 29, 2021.
24. Freedman DS, Khan LK, Serdula MK, et al. The relation of childhood BMI to adult adiposity: the Bogalusa heart study. *Pediatrics* 2003; 115 (1): 22–7.
25. Bhadoria A, Sahoo K, Sahoo B, et al. Childhood obesity: causes and consequences. *J Family Med Prim Care* 2015; 4 (2): 187–92.
26. Simmons M, Llewellyn A, Owen CG, et al. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obes Rev* 2016; 17 (2): 95–107.
27. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; 12 (3): e1001779.
28. Michael J, Conrado J, Brophy M, et al. Million Veteran Program: a megabiobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016; 70: 214–23.