

# Crowdsourced Health Data: Comparability to a US National Survey, 2013–2015

Veronica Yank, MD, Sanjhavi Agarwal, MPH, Pooja Loftus, MS, Steven Asch, MD, MPH, and David Rehkopf, ScD, MPH

**Objectives.** To determine the generalizability of crowdsourced, electronic health data from self-selected individuals using a national survey as a reference.

**Methods.** Using the world's largest crowdsourcing platform in 2015, we collected data on characteristics known to influence cardiovascular disease risk and identified comparable data from the 2013 Behavioral Risk Factor Surveillance System. We used age-stratified logistic regression models to identify differences among groups.

**Results.** Crowdsourced respondents were younger, more likely to be non-Hispanic and White, and had higher educational attainment. Those aged 40 to 59 years were similar to US adults in the rates of smoking, diabetes, hypertension, and hyperlipidemia. Those aged 18 to 39 years were less similar, whereas those aged 60 to 75 years were underrepresented among crowdsourced respondents.

**Conclusions.** Crowdsourced health data might be most generalizable to adults aged 40 to 59 years, but studies of younger or older populations, racial and ethnic minorities, or those with lower educational attainment should approach crowdsourced data with caution.

**Public Health Implications.** Policymakers, the national Precision Medicine Initiative, and others planning to use crowdsourced data should take explicit steps to define and address anticipated underrepresentation by important population subgroups. (*Am J Public Health*. 2017;107:1283–1289. doi:10.2105/AJPH.2017.303824)

The growth of internet-based sampling and data collection presents an important opportunity for cheaper, and thus, higher volume collection of health-related data. Such crowdsourced data have excited interest from a variety of perspectives. Individuals have turned to the “quantified self” (self-tracking of personal data) to compare their experiences with others.<sup>1,2</sup> Corporations have seen business opportunities, not just in wearable devices, but in the aggregate data they collect. Both Google and Apple have invested heavily in developing mobile device platforms for medical research, the former with the Google Baseline Study and the latter with the Apple ResearchKit.<sup>3</sup> Researchers and policymakers also are understandably eager to use the tools and information derived from mobile devices, social media platforms, and online search engines.<sup>2</sup> They have used Google trends for infection and tobacco research,<sup>4,5</sup> Facebook for social science, psychology, and public

health research,<sup>6,7</sup> Twitter for monitoring of health-related attitudes, and mobile devices to track clinical parameters.<sup>8</sup>

However, how generalizable are associations researchers identify in such data? If certain groups are over- or underrepresented, the research may generate misleading conclusions when extrapolated to larger populations. A debate from the 1990s provides some historical context. At that time, a national consensus emerged that underrepresentation of women and racial and ethnic minority participants in research limited its value. In 1993, the Food and Drug

Administration instructed drug developers that it expected inclusion of women in clinical pharmaceutical trials,<sup>9</sup> and the National Institutes of Health (NIH) quickly followed suit with statements on enrollment of women and racial and ethnic minorities in NIH-sponsored studies.<sup>10</sup> More recent work has emphasized an ongoing need for diversity among study participants in precision health studies. Whole genome sequencing associations developed from predominantly European cohorts may lead to diagnostic or treatment failures among patients who are from traditionally underrepresented groups.<sup>11,12</sup> After investigators found that benign genetic variants were misclassified as pathogenic among patients of African ancestry, they concluded,

These findings show how health disparities may arise from genomic misdiagnosis . . . errors that are related to . . . a historical dearth of populations that include persons of diverse racial and ethnic backgrounds.<sup>11(p662–663)</sup>

Thus, even in this era of precision medicine and the quantified self, attention to whose data we are—and are not—capturing in medical studies remains a research and policy imperative.

How participants enter a particular study can have a great influence on the generalizability of results. Traditionally, population studies have expended considerable effort on targeted recruitment of representative samples and statistical adjustments for over- or underrepresentation of subgroups among those enrolled. This approach of probability-based sampling allows

## ABOUT THE AUTHORS

Veronica Yank and Sanjhavi Agarwal are with the Division of General Internal Medicine, University of California, San Francisco. Pooja Loftus is with the Division of General Medical Disciplines, Stanford University, Stanford, CA. Steven Asch is with the V.A. Palo Alto Health Care System, Palo Alto, CA, and the Division of General Medical Disciplines, Stanford University. David Rehkopf is with the Division of General Medical Disciplines, Stanford University.

Correspondence should be sent to Veronica Yank, MD, 1545 Divisadero Street, Box 0320, University of California, San Francisco, San Francisco, CA 94115 (e-mail: veronica.yank@ucsf.edu). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints” link.

This article was accepted March 25, 2017.  
doi: 10.2105/AJPH.2017.303824

investigators to determine the degree of confidence they can have about the representativeness of the data. Nonprobability-based samples, such as convenience samples, may or may not be representative. Without examining this issue in detail, one cannot be confident about the ways in which nonprobability samples might differ from the general population.

Nonetheless, researchers have begun to capitalize on the public's interest in the quantified self and technology to enroll and study nonprobability-based samples recruited remotely. Crowdsourcing, in which self-selected individuals provide electronic data or feedback, is perhaps the most innovative recent method for study population accrual.<sup>13</sup> Social science and psychology researchers have used it widely over the past 5 years,<sup>14,15</sup> and its use is growing among scientists in other areas of health research.<sup>16,17</sup> The NIH's Precision Medicine Initiative plans to recruit 1 million US persons to its longitudinal patient cohort through nonprobability-based sampling. It will accept self-selected volunteers who will enroll online (and others who will enroll through healthcare organizations). Its working group describes how these volunteers can be "recruited through a number of technologies, such as internet, social media, and mobile technologies" and "that electronic consent across digital platforms will be a core component of cohort recruitment."<sup>18</sup> The cohort is named "All of Us" and will begin recruiting this year (<http://www.nih.gov/precision-medicine-initiative-cohort-program/participation>).

We sought to determine how research participants recruited electronically through crowdsourcing were similar to or different from other persons in the United States with respect to their health profiles. We performed a case study using the world's largest online crowdsourcing platform, Amazon Mechanical Turk (MTurk). We compared demographic and health characteristics of adults recruited through MTurk with those of the US population. We focused on health characteristics that are known risk factors for cardiovascular disease, which is the leading cause of death in the United States and the world.<sup>19</sup> In addition, many cardiovascular disease risk factors are suitable for remote measurement and data collection,

making them a growing focus in crowdsourced studies, including the All of Us cohort.<sup>18</sup>

## METHODS

MTurk has 500 000 registered, anonymous "workers," with an estimated 400 000 based in the United States; 15 000 of these are active on any given day.<sup>15</sup> "Requesters" are businesses or others who would like workers to complete short tasks for compensation. Requesters post descriptions of the tasks on the MTurk Web site. Workers self-select among these (e.g., identifying objects in photographs, translating text, providing survey feedback) based on task title, keywords, expected time involved, and compensation amount, which at the time of our study ranged from \$0.10 to \$0.25 per 10 minutes. Requesters of the task can set a batch limit for the number of respondents they want to complete it and prevent workers from performing the same task more than once.

### Surveys

We used MTurk to recruit a convenience sample of US-based adults (aged  $\geq 18$  years). We designed our task following best practices for MTurk,<sup>15</sup> with the title "Stanford Survey on Health and Well-Being" keywords "survey, demographics, health, medicine, well-being," compensation of \$0.25, and anticipated completion time of 10 minutes. We sought enrollment of 2000 participants—equating to 0.5% of the estimated 400 000 US-based MTurk workers. We posted the task to be completed by sequential batches of 200 respondents. We varied the batch posting times, days, and weeks to target MTurk workers online at different times. Surveys were completed over a 6-week period from June to August 2015. We modestly surpassed our enrollment target because some batches exceeded their 200 respondent limit before closing. We did not attempt to stratify enrollment according to predetermined characteristics (e.g., age, race) or apply sampling weights following recruitment, because our study intent was to mimic what others were already doing with crowdsourcing tools and

determine what kind of sample crowdsourcing (i.e., a nonprobability approach) would produce.

The US Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance System (BRFSS) is the largest continuously performed health survey in the world. It is administered annually in English by landline or cellular telephone to non-institutionalized adult US residents and gathers cross-sectional data on demographic characteristics, health characteristics, and disease risk factors. BRFSS provides survey sampling weights and has been validated to ensure that its data are representative of the US adult population.<sup>20</sup> We used data from the 2013 survey.

We adapted national survey questions for online administration. Selected questions focused on individual characteristics known to influence cardiovascular morbidity and mortality. Sociodemographic characteristics included age, age<sup>2</sup>, gender, race (White, Black, Asian, Pacific Islander, Native American/Alaska Native, Other), ethnicity (Hispanic/Latino, non-Hispanic/Latino), and measures of educational attainment (coded as did not graduate high school, high school degree or equivalent, some college or greater), annual income (selected from list of income ranges), and employment status (not employed, self-employed, employed for wages, employed for salary). Health behavior and chronic disease characteristics encompassed low level of physical activity (defined by reported weekly frequency, duration, or degree of vigor not meeting national guidelines), smoking (both previous and current), diabetes mellitus, hypertension, hyperlipidemia, and overweight or obesity (calculated using reported height and weight with adjusted cutoffs for respondents of Asian ancestry<sup>21</sup>). All data were based on self-reports of the respondents.

### Data Analysis

MTurk routinely collects the internet protocol addresses of workers. We applied geocoding software (Qualtrics Geo IP locator; Qualtrics; Provo, UT) to identify the nearest zip code location—a method estimated to have 70% to 95% accuracy<sup>22</sup>—and compared these by visual inspection to census-determined population density maps

TABLE 1—Cardiovascular Disease Risk Characteristics of Survey Respondents, Crowdsourced Health Data Compared With National Survey: United States, 2013–2015

Characteristics	Aged 18–29 Years		Aged 30–39 Years		Aged 40–49 Years		Aged 50–59 Years		Aged 60–75 Years	
	Crowdsourced	BRFSS	Crowdsourced	BRFSS	Crowdsourced	BRFSS	Crowdsourced	BRFSS	Crowdsourced	BRFSS
No. of respondents by group	992	50 101	587	55 485	223	68 493	151	100 979	62	146 025
Age, y, mean $\pm$ SD	24.8 $\pm$ 2.9	23.2 $\pm$ 3.5	33.7 $\pm$ 2.7	34.3 $\pm$ 2.9	43.8 $\pm$ 2.7	44.4 $\pm$ 2.9	53.5 $\pm$ 2.7	54.3 $\pm$ 2.9	65.2 $\pm$ 3.7	65.9 $\pm$ 4.2
Male, %	54.7	48.6	57.4	49.0	45.3	50.0	42.4	48.8	38.7	47.1
Race and ethnicity, %										
White	73.5	70.4	81.8	72.5	83.9	74.9	94.0	79.5	85.5	82.8
Black	7.7	14.9	5.3	14.1	7.6	13.8	3.3	12.6	6.5	10.8
Asian	13.3	8.1	7.2	7.5	5.4	6.0	2.7	3.8	1.6	3.3
Hispanic/Latino	10.6	22.0	8.7	22.8	4.5	18.0	4.0	13.0	3.2	9.1
Attended college, %	90.3	54.5	89.4	60.9	93.7	59.0	84.7	56.4	90.3	56.6
Income, \$, mean $\pm$ SD	27 055 $\pm$ 26 630	40 106 $\pm$ 26 375	40 916 $\pm$ 28 882	51 016 $\pm$ 29 752	48 664 $\pm$ 56 191	54 368 $\pm$ 30 213	42 479 $\pm$ 37 421	52 996 $\pm$ 30 155	37 509 $\pm$ 27 694	47 732 $\pm$ 27 384
Health behavior, %										
Current smoking	12.4	21.0	18.9	21.9	18.4	19.7	20.5	20.0	21.0	12.8
Low physical activity	58.1	56.1	67.5	59.0	65.0	58.4	74.2	56.6	77.4	55.0
Chronic disease, %										
Diabetes	2.7	1.4	4.6	3.1	5.8	7.7	13.3	13.6	12.9	20.9
Hypertension	8.1	8.7	13.6	15.7	17.9	26.4	34.4	41.1	50.0	65.9
Hypertlipidemia	8.3	5.4	12.4	15.3	22.4	28.4	33.8	42.1	41.9	52.9
Overweight or obesity	46.2	50.9	59.3	67.2	56.5	71.5	71.6	72.5	61.3	72.3

Note: BRFSS = Behavioral Risk Factor Surveillance System. Asian includes all Asian subsets, as well as Pacific Islanders. "Attended college" indicates individuals who attended some college or have a college degree or higher. For non-Asians, body mass index demarcations for overweight and obese were  $\geq 25$ – $29.9$  kg/m<sup>2</sup> and  $\geq 30$  kg/m<sup>2</sup>, respectively. For Asians, in accordance with recommendations of the World Health Organization and others, demarcations for overweight and obese were 23–26.9 kg/m<sup>2</sup> and  $\geq 27$  kg/m<sup>2</sup>, respectively.

based on 2010 US Census data.<sup>23</sup> We used kernel distribution estimation to estimate and visually display the age distributions of crowdsourced and national survey respondents.<sup>24</sup> The kernel distribution estimate approach is a method of visualizing the predicted age distribution of a population using a smooth curve rather than a histogram.

We assessed respondent characteristics by age strata: those aged 18 to 29, 30 to 39, 40 to 49, 50 to 59, and 60 to 75 years. Because there were no crowdsourced respondents older than 75 years, we truncated national survey data at age 75 years. We weighted BRFSS data according to published weights. In univariate analyses, we used the *t*-test and  $\chi^2$  test to compare crowdsourced and BRFSS characteristics of the respondents. All characteristics differed significantly among the groups ( $P < .05$ ) and were included in logistic regression models. For ease of viewing results in Tables 1 and 2, we omitted age<sup>2</sup>, status of not having graduated from high school, employment status, and previous smoking, whose results were congruent with those shown. (Full models are available in the technical appendix in a supplement to the online version of this article at <http://www.ajph.org>.)

We performed 3 types of sensitivity analyses. To assure findings did not vary by gender, we built models assessing men and women separately. Because cardiovascular disease occurs at higher rates among older Americans, we performed sensitivity analyses using data from another representative national survey, the Health and Retirement Study (HRS), which focuses on older adults.<sup>25</sup> We compared crowdsourced versus HRS respondents aged 50 to 75 years. Finally, as another way to identify demographic profiles of respondents differentially represented among crowdsourced versus national survey respondents, we fit age-stratified propensity models within both groups combined, which predicted the likelihood of having a high propensity score for being a crowdsourced respondent. We compared the characteristics of those with high propensity scores with all others. Study analyses were performed using Stata statistical

**TABLE 2—Adjusted Comparisons of Sociodemographic and Health Characteristics Among Crowdsourced Respondents vs US Population Stratified by Age, Crowdsourced Health Data Compared with National Survey: 2013–2015**

Characteristics	Aged 18–29 Years, OR (95% CI)	Aged 30–39 Years, OR (95% CI)	Aged 40–49 Years, OR (95% CI)	Aged 50–59 Years, OR (95% CI)	Aged 60–75 Years, OR (95% CI)
Male	1.2 (1.1, 1.4)	1.5 (1.2, 1.8)	0.9 (0.7, 1.2)	0.8 (0.6, 1.1)	0.6 (0.4, 1.1)
Race/ethnicity					
White	1.4 (1.0, 1.9)	1.3 (0.9, 1.8)	1.9 (0.9, 4.1)	2.9 (1.5, 5.7)	0.7 (0.3, 1.9)
Black	0.6 (0.4, 0.8)	0.3 (0.2, 0.5)	0.9 (0.4, 2.1)	NA	NA
Asian	1.9 (1.3, 2.8)	0.8 (0.5, 1.4)	1.0 (0.4, 2.7)	NA	NA
Hispanic/Latino	0.4 (0.4, 0.6)	0.3 (0.2, 0.5)	0.3 (0.1, 0.5)	NA	NA
Non-White combined	NA	NA	NA	0.4 (0.2, 0.7)	0.4 (0.1, 1.1)
Attended college	5.7 (4.5, 7.1)	5.1 (3.9, 6.8)	9.6 (5.5, 16.9)	6.3 (3.9, 10.0)	10.4 (3.9, 27.6)
Income level	0.96 (0.96, 0.96)	0.97 (0.97, 0.97)	0.97 (0.97, 0.97)	0.97 (0.97, 0.97)	0.97 (0.97, 0.97)
Health behavior					
Low physical activity	1.1 (1.0, 1.3)	1.6 (1.3, 1.9)	1.6 (1.2, 2.1)	2.4 (1.7, 3.4)	3.0 (1.6, 5.5)
Current smoking	0.5 (0.4, 0.6)	0.8 (0.6, 1.0)	1.1 (0.8, 1.6)	1.0 (0.6, 1.5)	2.4 (1.2, 5.1)
Chronic disease					
Diabetes	2.2 (1.4, 3.5)	2.2 (1.4, 3.3)	1.4 (0.8, 2.5)	1.5 (0.9, 2.4)	0.8 (0.4, 1.7)
Hypertension	0.8 (0.6, 1.1)	0.9 (0.7, 1.1)	0.8 (0.5, 1.1)	1.0 (0.7, 1.3)	1.1 (0.6, 2.0)
Hyperlipidemia	1.3 (1.0, 1.7)	0.9 (0.7, 1.1)	0.9 (0.6, 1.3)	0.8 (0.6, 1.1)	0.7 (0.4, 1.3)
Overweight or obesity	0.8 (0.7, 1.0)	0.7 (0.6, 0.9)	0.6 (0.5, 0.8)	1.1 (0.8, 1.6)	0.7 (0.4, 1.1)

Note. CI = confidence interval; NA = not applicable; OR = odds ratio. Results are presented as OR for characteristics being present among crowdsourced respondents vs Behavioral Risk Factor Surveillance System respondents. Asian includes all Asian subsets, as well as Pacific Islanders. For the age groups 50–59 and 60–75 years, there were too few Black, Asian, or Hispanic/Latino crowdsourced respondents to include as separate variables, and we combined Black, Asian, and Hispanic/Latino respondents into a “Non-White combined” group. “Attended college” indicates individuals who attended some college or have a college degree or higher. We transformed income by multiplying by  $10^{-3}$ . Full models included additional covariates of age, age<sup>2</sup>, status of not having graduated from high school, employment type, and previous smoking (see the data available in the technical appendix in a supplement to the online version of this article at <http://www.ajph.org>).

software version 14.2 (StataCorp LP; College Station, TX).

## RESULTS

Of the 2148 crowdsourced respondents who accessed our survey, 2015 (94%) completed it. We compared their data to data from 428 211 BRFSS respondents. Locations of crowdsourced respondents generally corresponded to census-determined population densities and were mapped to 49 US states (Figure A, available as a supplement to the online version of this article at <http://www.ajph.org>).

Crowdsourced respondents ranged in age from 18 to 75 years and reported

cardiovascular risk characteristics (Table 1). They were younger on average (mean  $\pm$  SD = 33  $\pm$  11 y) than US adults aged 18 to 75 years (mean  $\pm$  SD = 44  $\pm$  16 y). As Figure 1 shows, compared with BRFSS, crowdsourced samples tended to be overrepresented in the 20 to 39 years range, and underrepresented in those aged 40 to 75 years. Few were aged 60 years or older: 55 (3%) were aged 60 to 69 years, and 7 (<1%) were aged 70 years or older.

In adjusted models, we found that certain risk characteristics were over- or underrepresented among crowdsourced respondents versus US adults, whereas others were similarly represented (Table 2). Crowdsourced respondents were more likely to be men in the younger groups (aged

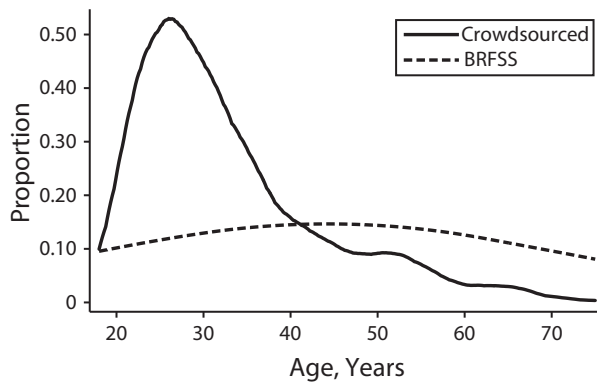
18–29 and 30–39 years). Other than those in the oldest group (60–75 years), crowdsourced respondents were more likely to be White and less likely to be Black or Hispanic. In the youngest group (18–29 years), they were significantly more likely to be Asian. Crowdsourced respondents in all age groups were significantly more likely to be college educated, but they had lower annual incomes.

With respect to health characteristics, crowdsourced respondents were more likely to have low physical activity. Among younger groups (aged 18–29 and 30–39 years), crowdsourced respondents were more likely to have diabetes, although they were less likely to be overweight or obese. Crowdsourced respondents in the youngest group (18–29 years) also had lower tobacco use but a higher prevalence of hyperlipidemia. Among those aged 40 to 49 and 50 to 59 years, crowdsourced respondents were similar to US adults in adjusted odds of smoking, diabetes, hypertension, and hyperlipidemia. Those aged 50 to 59 years were also similar in overweight or obesity. Although they were few in number, crowdsourced respondents aged 60 to 75 years had similar odds of diabetes, hypertension, hyperlipidemia, and overweight or obesity but had higher odds of current smoking.

In sensitivity analyses, separate analyses for men and women generated similar findings to the main findings (Table A, available as a supplement to the online version of this article at <http://www.ajph.org>). For older respondents (aged 50–59 and 60–75 years), comparisons with HRS data were also similar (Tables B and C, available as a supplement to the online version of this article at <http://www.ajph.org>). Finally, in analyses that identified outlier respondents (those with demographic profiles uniquely overrepresented among crowdsourced respondents), outliers consisted of only 0.4% (1000/275 058) of national survey respondents and reinforced the main findings on race and educational attainment (Table D, available as a supplement to the online version of this article at <http://www.ajph.org>).

## DISCUSSION

Our case study of using crowdsourcing for health research identified important



Note. BRFSS = Behavioral Risk Factor Surveillance System. We used the kernel distribution estimation approach<sup>24</sup> available in the Stata software package to estimate age densities (a method of visualizing the estimated age distribution of a population using a smooth curve rather than a histogram; available at: <https://www.stata.com/manuals13/rkdensity.pdf>).

**FIGURE 1—Age Distribution of Survey Respondents, Crowdsourced Health Data Compared With National Survey: United States, 2013–2015**

similarities and differences between crowdsourced respondents and the US population as a whole. We found that the cardiovascular disease risk profile of crowdsourced study participants differed in well-defined ways from the US population. Crowdsourced participants were younger and more likely to be non-Hispanic and White, and had higher levels of educational attainment. Those who were aged 40 to 59 years were most representative with regard to smoking, diabetes, hypertension, and hyperlipidemia, but even they had significant differences with regard to race/ethnicity, education, and physical activity. Crowdsourced data from younger age groups were even less similar, and those aged 60 years and older were difficult to reach by the crowdsourced method.

Some of our sociodemographic findings were similar to previous work. Previous studies identified crowdsourced respondents as simultaneously having higher levels of education but lower household incomes than the general US population<sup>15,26</sup>—perhaps indicating that people who are relatively underemployed for their degree of education are more likely than others to pursue online activities. Other investigators found that crowdsourced respondents were more likely to be White and less likely to be Hispanic or Black.<sup>15,26</sup> Some concluded that crowdsourced respondents were more likely to be Asian, but they did not stratify respondents by age as we did. This allowed

us to determine that Asians were over-represented among those aged 18 to 29 years but were likely underrepresented among those aged 50 to 75 years. Our findings that younger crowdsourced respondents were more likely to have diabetes but less likely to be obese might indicate an underlying characteristic (not measured by our survey) of being sedentary. Being sedentary is an independent risk factor for developing diabetes regardless of physical activity achieved at other times of day.<sup>27</sup>

Although a number of previous studies examined health characteristics of crowdsourced respondents, most focused on psychology<sup>14,15,28,29</sup> and, to a lesser extent, other fields of mental health and addiction.<sup>16</sup> Their findings on comparability to the US population were mixed. A psychology review on MTurk respondents concluded that they were not representative, but were “more diverse than samples typically used” in psychology research (e.g., students).<sup>15</sup> Other investigators demonstrated that online crowdsourcing could be used to anonymously study hard-to-reach groups about sensitive topics (e.g., addiction, sexual practices) and argued that this advantage might outweigh the potential non-representativeness of data.<sup>30,31</sup>

Recently, investigators in public health and other fields of medical researcher have adopted use of certain online technologies. They have used social networking sites, search engines, e-mail networks, and blogs

to perform observational studies of substance abuse, sexually transmitted infections, adolescent health, infection trends, and public attitudes toward vaccination campaigns.<sup>14,15</sup> Furthermore, they have noted the huge potential of crowdsourcing research to extend this existing work,<sup>13,17,32</sup> and some have begun to use it as a research tool.<sup>33–35</sup> However, fewer have done so (thus far) than in the social sciences and mental health fields.

Calls for mass adoption of digital recruitment, data collection, and delivery platforms for clinical and public health endeavors have increased.<sup>8,32</sup> The ongoing Health eHeart study, a partnership between the American Heart Association and the University of California, San Francisco, invites voluntary enrollment and data collection through an online platform (<https://www.health-eheartstudy.org>). It is prefaced on fully online and smartphone-based participation, collecting self-reported data and linking it with data from wearable personal sensors, online social networks, and other importable “big data” to provide “real-life, real-time metrics.” Within the context of these broader trends there are a number of potential implications of the study presented here.

### Public Health Implications

We argue that enthusiasm for technology-based crowdsourcing should be tempered with a suitable degree of caution. Policymakers, funders of research, and researchers themselves should be explicit about the advantages and limitations of relying on crowdsourced data. In cases in which investigators could argue that underlying sociodemographic characteristics or health variables should not matter in terms of their influence on health outcomes, use of crowdsourcing for recruitment and data collection might be useful. Similar arguments might be made about its usefulness in reaching groups that are not meant to be representative but whose responses might still generate key insights into underinvestigated areas.

By contrast, in areas in which representativeness is sought and responses might be expected to vary according to participant sociodemographic or health characteristics, investigators must ensure that the use of crowdsourcing is appropriate.<sup>36</sup> Much as was

done in the 1990s to boost enrollment of women and racial/ethnic minorities in clinical trials, stakeholders (public health experts, researchers, policymakers, and the public) may need to establish guidance for studies that rely on remotely conducted patient recruitment and data collection.

First, stakeholders must determine that the target population for their research or policy question is reachable through the electronic platform. Second, until they show otherwise, they should expect that certain groups will be systematically underrepresented in data collected by these means. From our case study (and the work of others), it was clear that older adults and respondents from certain racial and ethnic groups are likely to be underrepresented among crowdsourced respondents. The same might be said for respondents from rural areas (where internet connections are less available) or respondents with low English-language proficiency, health literacy, or digital access and proficiency. In such cases, alternative recruitment and data collection efforts will be required. Third, stakeholders should build the possible need for statistical adjustment (e.g., weightings) for nonrepresentative samples into each study design. If there are sufficient responses, albeit proportionally underrepresented, from key subgroups, sampling weights can be created and used in analyses.

In the near future, the national Precision Medicine Initiative will begin enrollment into the historic All of Us cohort. In many ways, this cohort is an answer to the call for mass adoption of digital recruitment and data collection approaches, and public enthusiasm for the “quantified self.” It proposes to enroll 1 million persons in the United States, which would make it the largest longitudinal cohort ever undertaken. Notably, it will use crowdsourcing—engaging and enrolling participants online—as one of its recruitment approaches. The Working Group of the Initiative has emphasized the “critical importance of recruiting underrepresented populations.”<sup>18</sup> However, our study found that online crowdsourced recruitment led to systematic gaps in representation of some US adults (and overrepresentation of others). Thus, for the All of Us cohort to reach its full potential—to truly represent all of the United States—it

must anticipate and address potential gaps in participation to be expected from its use of voluntary online enrollment as one of its recruitment approaches. We are encouraged that the cohort program is pursuing alternative recruitment pathways through healthcare organizations. Ideally, these will fill predicted gaps in crowdsourced enrollment by historically underrepresented groups. In addition, if the cohort achieves its target recruitment of 1 million US persons, there may be high enough enrollment, even among these groups, to enable statistical adjustments for underrepresentation. At the same time, we encourage this Initiative and others that plan to use crowdsourcing to carefully consider whether crowdsourced data are those that can best elucidate their research questions, and whether and how they need to augment their recruitment strategies (or statistical adjustment strategies) to best represent the patients they want to impact and the US population as a whole.

If we are to build from the public’s willingness to provide large quantities of remotely collected electronic data, we must familiarize ourselves with—and then expand the pool of—“quantified selves” being detailed. This is a pivotal time. Crowdsourced research has the potential to expand citizen participation but also to exacerbate health disparities. Leaders in public health, medicine, research, and policy must ensure that the data upon which decisions are being made are derived from respondents who have similar characteristics to those they are trying to help, and that large swaths of the US population are not left out of data collection efforts. Public health policymakers can galvanize this effort. They can call for national consensus-building and guidance on how crowdsourcing should be used for health research.

## Limitations

Our study had several limitations. We tested one crowdsourcing platform with relatively low hourly incentive. Although it is the most widely used in the world, predictors of participation might vary by platform or incentive structures, or for platforms that recruit study subjects willing to participate without compensation.

Second, the internet protocol addresses we used for geocoding respondent locations might represent sites of employment or other community hubs (e.g., libraries), which are more likely clustered in urban settings. However, these urban centers are likely to be in proximity to sites of residences of respondents who live in suburban or rural locations. Third, the national survey we used, BRFSS, has known limitations.<sup>20</sup> However, it remains, a gold standard for characterizing the US adult population that is English proficient. Furthermore, we performed sensitivity analyses of findings for older respondents using data from another nationally representative survey.

## Conclusions

Greater attention should be given to determining what populations are and are not reachable using remote, electronic data collection platforms. Studies that rely on crowdsourced respondents need to define the profile of the people generating those data. Assumptions of generalizability should be avoided. As in the 1990s when national policies were undertaken to ensure the inclusion of women and racial/ethnic minorities in clinical trials, proactive efforts need to be made to understand and promote inclusion of underrepresented groups within the Precision Medicine Initiative and similar projects that use crowdsourced recruitment and data collection. **AJPH**

## CONTRIBUTORS

V. Yank originated and designed the study, supervised all aspects, interpreted analyses, and wrote the article. S. Agarwal performed data collection and analyses and assisted with interpretation and writing. P. Loftus supervised statistical analyses and assisted with interpretation. S. Asch assisted with interpretation and writing. D. Rehkopf contributed to study design, analyses, interpretation, and writing.

## ACKNOWLEDGMENTS

V. Yank is supported by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (K23DK097308). D. Rehkopf is supported by a grant from the National Institute of Aging (K01AG047280).

**Note.** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## HUMAN PARTICIPANT PROTECTION

Approval for the research was obtained from the Stanford institutional review board. Crowdsourced participants provided online informed consent before survey initiation.

## REFERENCES

1. Wolf G. The data-driven life. *The New York Times Sunday Magazine*. 2010;MM38.
2. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. *Big Data*. 2013; 1(2):85–99.
3. Comstock J. Apple and Google's different, but complementary, approaches to medical research. 2015. Available at: <http://mobihealthnews.com/41189/apple-and-googles-different-but-complementary-approaches-to-medical-research>. Accessed May 9, 2016.
4. Cavazos-Rehg PA, Krauss M, Spitznagel E, et al. Monitoring of non-cigarette tobacco use using Google Trends. *Tob Control*. 2015;24(3):249–255.
5. Houle JN, Collins JM, Schmeiser MD. Flu and finances: influenza outbreaks and loan defaults in US cities, 2004–2012. *Am J Public Health*. 2015;105(9):e75–e80.
6. Wilson RE, Gosling SD, Graham LT. A review of Facebook research in the social sciences. *Perspect Psychol Sci*. 2012;7(3):203–220.
7. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. *J Med Internet Res*. 2014;16(3):e79.
8. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. *Health Aff (Millwood)*. 2016;35(2):341–347.
9. Merkatz RB, Temple R, Subel S, Feiden K, Kessler DA; Working Group on Women in Clinical Trials. Women in clinical trials of new drugs. A change in Food and Drug Administration policy. *N Engl J Med*. 1993; 329(4):292–296.
10. National Institutes of Health. NIH guidelines on the inclusion of women and minorities as subjects in clinical research. *Fed Regist*. 1994;59(59):11146–11151.
11. Manrai AK, Funke BH, Rehm HL, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med*. 2016;375(7):655–665.
12. Burchard EG. Medical research: missing patients. *Nature*. 2014;513(7518):301–302.
13. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res*. 2012;14(2):e46.
14. Goodman JK, Cryder CE, Cheema A. Data collection in a flat world: the strengths and weaknesses of Mechanical Turk samples. *J Behav Decis Making*. 2013;26(3):213–224.
15. Chandler J, Shapiro D. Conducting clinical research using crowdsourced convenience samples. *Annu Rev Clin Psychol*. 2016;12:53–81.
16. Naslund JA, Aschbrenner KA, Marsch LA, McHugo GJ, Bartels SJ. Crowdsourcing for conducting randomized trials of internet delivered interventions in people with serious mental illness: a systematic review. *Contemp Clin Trials*. 2015;44:77–88.
17. Ranard BL, Ha YP, Meisel ZF, et al. Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *J Gen Intern Med*. 2014;29(1):187–203.
18. Precision Medicine Initiative (PMI) Working Group. The Precision Medicine Initiative Cohort Program – Building a research foundation for 21st century medicine. 2015. Available at: <https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf>. Accessed November 14, 2016.
19. World Health Organization. WHO fact sheet: cardiovascular diseases (CVDs). Available at: <http://www.who.int/mediacentre/factsheets/fs317/en>. Accessed November 18, 2016.
20. Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Med Res Methodol*. 2013;13:49.
21. WHO EC. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet*. 2004;363(9403):157–163.
22. King K. Geoloaction and federalism of the internet: cutting internet gambling's Gordian Knot. *Columbia Sci Technol Law Rev*. 2010;XI:41–75.
23. US 2010 Census. Available at: <http://www.census.gov/2010census/data>. Accessed October 28, 2015.
24. Zambom AZ, Dias R. A review of kernel density estimation with applications to econometrics. 2012. Available at: <https://arxiv.org/abs/1212.2812>. Accessed November 14, 2016.
25. Growing Older in America. *The Health and Retirement Study*. Bethesda: National Institute on Aging; 2015.
26. Huff C, Tingley D. “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Res Politics*. 2015;2(3):1–12.
27. Wilmot EG, Edwardson CL, Achana FA, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia*. 2012;55(11):2895–2905.
28. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspect Psychol Sci*. 2011;6(1):3–5.
29. Choi JH, Lee JS. Online social networks for crowdsourced multimedia-involved behavioral testing: an empirical study. *Front Psychol*. 2016;6:1991.
30. Kristan J, Suffoletto B. Using online crowdsourcing to understand young adult attitudes toward expert-authored messages aimed at reducing hazardous alcohol consumption and to collect peer-authored messages. *Transl Behav Med*. 2015;5(1):45–52.
31. Burkill S, Copas A, Couper MP, et al. Using the web to collect data on sensitive behaviours: a study looking at mode effects on the British National Survey of Sexual Attitudes and Lifestyles. *PLoS One*. 2016;11(2):e0147983.
32. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. *Am J Prev Med*. 2014;46(2):179–187.
33. Smolinski MS, Crawley AW, Baltrusaitis K, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. *Am J Public Health*. 2015;105(10):2124–2130.
34. Santiago-Rivas M, Schnur JB, Jandorf L. Sun protection belief clusters: analysis of Amazon Mechanical Turk data. *J Cancer Educ*. 2016;31(4):673–678.
35. Merchant RM, Griffis HM, Ha YP, et al. Hidden in plain sight: a crowdsourced public art contest to make automated external defibrillators more visible. *Am J Public Health*. 2014;104(12):2306–2312.
36. Vayena E, Mastroianni A, Kahn J. Ethical issues in health research with novel online sources. *Am J Public Health*. 2012;102(12):2225–2230.