

# BMJ Open The Understanding America Study (UAS)

Arie Kapteyn <sup>1,2</sup> Marco Angrisani <sup>1,2</sup> Jill Darling,<sup>1</sup> Tania Gutsche<sup>1</sup>

**To cite:** Kapteyn A, Angrisani M, Darling J, *et al*. The Understanding America Study (UAS). *BMJ Open* 2024;**14**:e088183. doi:10.1136/bmjopen-2024-088183

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-088183>).

Received 30 April 2024

Accepted 09 September 2024

## ABSTRACT

**Purpose** The Understanding America Study (UAS) is a probability-based Internet panel housed at the Center for Economic and Social Research at the University of Southern California (USC). The UAS serves as a social and health sciences infrastructure for collecting data on the daily lives of US families and individuals. The collected information includes survey data, DNA from saliva samples, information from wearables, contextual and administrative linkages, ecological momentary assessments, self-recorded narratives and electronic records of financial transactions. The information collected focuses on a defining challenge of our time—identifying factors explaining racial, ethnic, geographic and socioeconomic disparities over the life course, including racial discrimination, inequalities in access to education and healthcare, differences in physical, economic and social environments, and, more generally, the various opportunities and obstacles one encounters over the life course. The UAS infrastructure aims to optimise engagement with the wider research community both in data dissemination and in soliciting input on content and methods. To encourage input from the research community, we have reserved 100 000 min of survey time per year for outside researchers, who can propose to add survey questions four times a year.

**Participants** The UAS currently comprises about 15 000 US residents (including a 3500-person California oversample) recruited by Address-Based Sampling and provided with Internet-enabled tablets if needed. Surveys are conducted in English and Spanish.

**Findings to date** Since the founding of the UAS in 2014, we have conducted more than 600 surveys, including a sequence of surveys collecting biennial information on health and retirement (the complete Health and Retirement Study instrument), 11 cognitive assessments, personality, knowledge and use of information on Social Security programme rules, work disability and subjective well-being. Several hundreds of papers have been published based on the collected data in the UAS. Studies include documentations of the mental health effects of the COVID-19 pandemic and how this varied across socioeconomic groups; comparisons of physical activity measured with accelerometers and by self-reports showing the dramatic biases in the latter; extensive studies have shown the power of using paradata in gauging cognitive change over time; several messaging experiments have shown the effectiveness of information provision on the quality of decision-making affecting well-being at older ages.

**Future plans** The UAS national sample is planned to grow to 20 000 respondents by 2025, with subsamples

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Probability-based Internet panels such as the Understanding America Study (UAS) avoid selection problems facing opt-in or convenience Internet panels, while several studies confirm their ability to provide population representative information.
- ⇒ Attrition in the UAS is on a par with other population representative panels.
- ⇒ By breaking up surveys in short (30 min or less) modules that are administered frequently, the amount of longitudinal information collected per respondent is substantially larger than in traditional interviewer-administered surveys.
- ⇒ The flexibility of the Internet as a survey mode allows for quick turn-around data collection and dissemination, and hence for fast response to new developments, such as the COVID-19 pandemic.
- ⇒ As with any self-interviewing survey mode, Internet interviewing limits our ability to collect biomarker data, as these must be collected by respondents themselves without the help of an in-person interviewer.

of about 2500 African American, 2000 Asian and 3000 Hispanic participants and an oversample of rural areas. An increasing amount of non-interview data (contextual information, data from a suite of wearables and administrative linkages) is continually being added to the data files.

## INTRODUCTION

The roots of the UAS lie in an early recognition by the Principal Investigator of the Health and Retirement Study (HRS), Robert J Willis, in 2001 that the Internet was going to be an important interview mode for national surveys. This resulted in two consecutive research grants awarded by the National Institute on Ageing (NIA) entitled 'Internet Interviewing and the HRS', which were jointly conducted by the University of Michigan and the RAND Corporation. As a direct result, the American Life Panel (ALP) was founded at RAND. The ALP was a probability-based Internet panel. It was used to conduct several mode experiments in conjunction with the HRS, but gradually became a platform for numerous research projects beyond



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Center for Economic and Social Research, University of Southern California, Los Angeles, California, USA

<sup>2</sup>Department of Economics, University of Southern California, Los Angeles, California, USA

## Correspondence to

Arie Kapteyn; [kapteyn@usc.edu](mailto:kapteyn@usc.edu)

the original intent of providing insights into the potential of Internet interviewing for the HRS. After the PI of the ALP, Arie Kapteyn, moved to University of Southern California (USC), the UAS was founded in 2014 with a structure similar to the ALP, while taking advantage of the insights gained from the experiences with the ALP.

The goal of the UAS is to be a next-generation probability-based population survey research platform that integrates state-of-the-art survey practices for longitudinal and cross-sectional data collections and merges various existing data sources into a flexible, easily accessible infrastructure. In this pursuit, the UAS combines online survey information with non-survey data, including DNA from saliva samples, person-generated health data from wearable devices and paradata available in various UAS data collection modalities. Additionally, it incorporates contextual data from external sources, such as information on green spaces, pollution levels, crime rates, state policies and other geocoded details.

A flexible platform that allows the comprehensive combination of diverse data from multiple sources and domains is essential for the study of health and economic inequalities, given the many mechanisms that create and allow these inequalities to exist.

The substantive focus of the UAS is to understand how health disparities and economic inequality develop over the life course. This means that prospective survey data are supplemented with life history information, which is either directly elicited from respondents or retrieved from external sources (eg, by linking residential histories with geocoded and time-specific external information).

The UAS is both a source of longitudinal information on US residents and an open platform for advanced social science and health research. All data collected in the UAS are available to the broader research community with minimal delay, in principle immediately after a field period ends. Researchers can easily add questions to existing UAS instruments, which can then be combined with earlier collected information. Although researchers contributing outside funding to use the UAS platform may be granted a brief embargo period (up to 6 months, in exceptional cases up to a year), in principle, all data collected in the UAS are available for all registered users independent of the funding source of any particular data set.

With few exceptions, all survey data are collected over the Internet. This allows for flexibility in the timing and duration of any survey. In practice, UAS participants answer surveys once or twice a month. Even though each survey is relatively short, never exceeding 30 min, the high frequency of surveys enables the collection of vast amounts of data, much more than would be possible in interviewer-administered modes.

An important feature of the Internet as the backbone of the UAS infrastructure is the ease with which experiments can be conducted and implemented. With minimal changes in the code underlying a survey, it is straightforward to assign UAS participants to randomly

chosen experimental groups. These groups may receive different kinds of information or monetary incentives, or may differ in terms of how the survey questionnaire is organised or certain questions are phrased. Similarly, one can conduct specific data collections, such as intensive Ecological Momentary Assessments (EMAs) among targeted subgroups. Third, the existing infrastructure permits constant and real-time monitoring of the data collection process. As a result, problems can be identified and resolved as they arise.

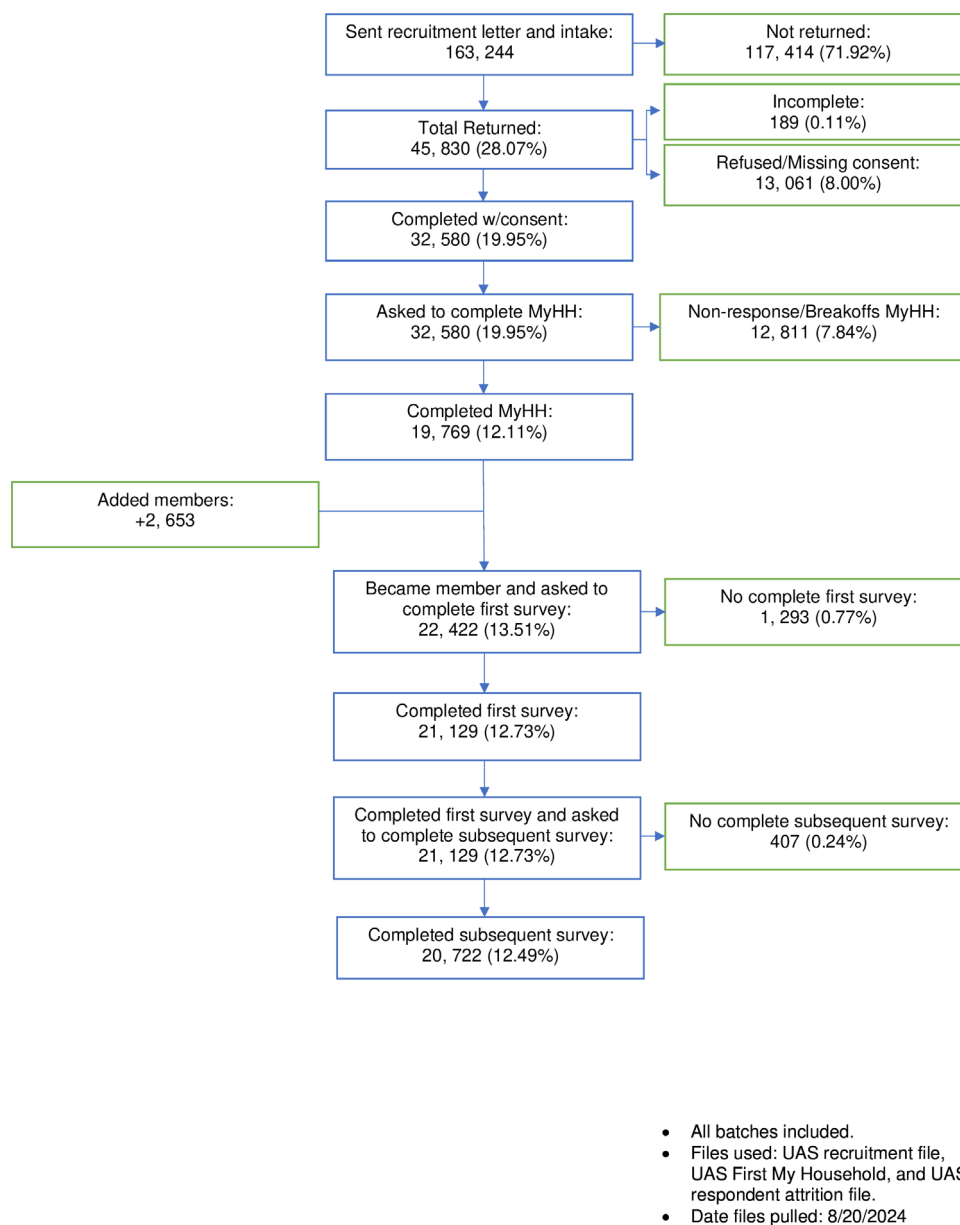
## COHORT DESCRIPTION

The UAS was founded in 2014 and has grown to about 15 000 US residents, including a 3500-person California oversample. We are expanding the panel to at least 20 000 participants by the end of 2025. Surveys are conducted in English and Spanish. Any US resident 18 or older is eligible to participate in the study. Currently, information on individuals younger than 18 is only collected from guardians, although that may change in the future. Apart from that, all information is collected through self-reports, or passively from biomarkers (currently saliva for genotyping respondents), wearables, contextual data and administrative linkages. Respondents are recruited via Address-Based Sampling (ABS) by drawing from post office delivery sequence files provided by a vendor. Using added information by the vendor and Census Tract information about the address of each potential respondent, inclusion probabilities are calculated such that the panel composition satisfies certain desirable characteristics. Specifically, we are aiming for oversamples of African American (2500), Asian (2000) and Hispanic (3000) participants, and residents of rural areas (5000). Individuals selected into the sample are provided with an Internet-enabled tablet if they are interested in participating but do not have the necessary hardware.

Figure 1 shows a flow chart of the recruiting process of all batches since the start of the UAS in 2014. In total, 34 batches have been recruited at the time of writing. The recruiting process is documented in detail on the UAS website. After respondents consent to join the UAS, the first survey they are asked to answer is a basic demographic questionnaire about their background characteristics and their household composition, called 'My Household', abbreviated to MyHH in the diagram. The invitation letter asks one respondent in a household to be the primary respondent, but other household members 18 or over are also invited to join the UAS. These are denoted as 'Added members' in the diagram.

## How often are participants followed?

Core information is collected at a 2 year frequency, akin to the HRS or Panel Study of Income Dynamics (PSID), which both collect core information at a 2 year frequency. As noted, surveys are broken up into short segments, each designed to take no more than 30 min. Currently, the core surveys comprise 22 modules. Respondents join the



**Figure 1** Consort diagram of the UAS recruiting process; all batches combined. MyHH, My Household; UAS, Understanding America Study.

panel at different times and are invited to take the core surveys following a predefined sequence. This means that data from core modules are being collected continuously, as different participants answer at different times. Panel members are offered one or two surveys per month, including the core surveys and various other surveys or experiments. Response rates to individual surveys are on the order of 75%, but can be as high as 90%, when a survey is kept in the field long enough. The median number of surveys answered by respondents per year is 28. Analysis by Jin and Kapteyn<sup>1</sup> finds no overall relation between the number of survey invitations that respondents have received in the past and the probability of responding to subsequent surveys. Moreover, they find a positive relation between the length of a preceding survey (measured by the number of screens shown to a respondent) and the

probability of responding to a new survey. The authors interpret this as indicating that the incentive of \$20 per 30 min of survey time is generally considered sufficient compensation for the effort of responding. As of August 2024, about 640 distinctive surveys have been conducted in the UAS. All the corresponding data sets are publicly available for analysis and can be linked using unique respondent identifiers. Each data set includes weights that align the sample to the US adult population in terms of gender, race/ethnicity, age, education and geographic location (Census regions).

## What has been measured?

### Core surveys

The UAS core surveys include the complete HRS core instrument and modules eliciting personality (the big

five), 11 cognition measures, financial literacy, financial outcomes, labour force status, retirement, knowledge of Social Security rules and subjective well-being. [Table 1](#) provides a summary of current core content.

Over the next year, the research team will considerably expand the core content by adding an extensive Life History Survey, a module on childhood experiences and education, modules on health behaviour, decision-making, mental health, cognitive measurement and physical and social environments. These are also listed in [table 1](#).

### Ecological Momentary Assessments (EMAs)

Several projects have conducted intensive EMAs among subgroups of respondents, eliciting affect, social contact and time use.<sup>2</sup>

### Genotyping the UAS

Supported by a recent grant from NIA, we have started asking UAS participants for consent to provide saliva samples with the goal of deriving a large number of polygenic indices that will be added to the UAS database.

### Contextual data and administrative linkages

UAS has created a contextual data resource which links to individual UAS respondent identifiers. This resource is similar to the HRS Contextual Data Resource Series<sup>3</sup> ([box 1](#))

We link survey data with CMS records for participants who consent (currently 45% of participants enrolled in Medicare or Medicaid). We are in the process of finalising the process to also link Social Security Earnings Records, and the National Death Index. Since linking contextual and programme data to survey data increases the risk of disclosure, linked data sets are available through the NIA supported Data Linkage enclave (<https://www.nia.nih.gov/research/dbsr/nia-data-linkage-program-linkage>).

### Wearables

At this moment, we are collecting data through two wearable devices. Approximately 1000 UAS respondents are wearing a Fitbit Versa,<sup>4</sup> while another 1000 are wearing an Atmotube air quality monitor. Both devices are linked to a smartphone app, which allows us to collect the data in the background. Importantly, the UAS has provided the devices rather than relying on already existing ownership, which would lead to highly selective samples. Consent rates for participation in these projects are 64% in both cases. Analysis of consent by demographics shows few differences, with the exception of age: respondents over 65 are less likely to participate.

### Transactional data

The UAS has pioneered the passive collection of financial information through a financial aggregator in a population-representative panel. Over the years, about 1000 respondents have provided permission to retrieve real-time transactions and balances from at least one of their financial accounts.

**Table 1** Current and planned UAS core content

Description	Notes
Background, household, health history, cognitive abilities	HRS sections A–D
Family, health, care and living arrangements	HRS sections E–H
Current job status, job history, health-related work impairments	HRS sections J–M
Health insurance, healthcare usage and probabilities of events	HRS sections N–P
Income and assets	HRS sections Q–R
Wills, trusts and life insurance policies	HRS sections T–W
Financial literacy; personality; understanding probabilities; numeracy	Various sources
Satisfaction with life domains; well-being yesterday; neighbourhood quality; income comparisons	Various sources
What do people know about social security	Developed in collaboration with SSA
Financial services and decision-making	
Ways people get information on retirement and social security	Developed in collaboration with SSA
Subjective numeracy and consumer financial well-being	
Views and knowledge about the social security disability programme	Developed in collaboration with SSA
Retirement Preparedness Index	Derived variable
Financial health score	Derived variable
<i>Cognition:</i>	
Serial sevens, general knowledge, word recall, probability of cognitive impairment score	From HRS sections
Financial literacy, numeracy	Part of survey mentioned above
Woodcock-Johnson 1—numbers	
Woodcock-Johnson 2—picture vocabulary	
Woodcock-Johnson 3—verbal analogies	
Stop and go switch	
Figure identification	
<i>Planned content as of 2024:</i>	
Life history, including residential history	Various sources
Decision-making	Various sources

Continued



**Table 1** Continued

Description	Notes
Physical/social environment	Various sources
Mental health	Various sources
Disability	Various sources
Economic measures	Various sources
Health behaviour	Various sources
Childhood, education	Various sources
HRS, Health and Retirement Study; SSA, Social Security Administration.	

### Patient and public involvement

The UAS sample is recruited via ABS; thus, all members of the public who speak English or Spanish have a non-zero probability of being chosen to complete our research.

Though the research subjects are not involved in the design or conduct of this internet panel study, we frequently engage our subjects in qualitative research in order to learn how clear our questionnaires are and what topics we may be missing in the various subjects we cover including health, finances, public events, policy and education. The subjects are told in their consent that we make deidentified data available to the research community so that academics and policy makers can learn from their input. We also communicate findings to our subjects via newsletters and posts in their study pages.

### Box 1 Contextual data resources (currently available or planned)

#### Currently available

- ⇒ US Decennial Census and American Community Survey
- ⇒ USDA Food and Environment Access
- ⇒ Street Connectivity
- ⇒ Uniform Crime Reports 1930–2020 (UCR)
- ⇒ Weather Data
- ⇒ Area Health Resource Files
- ⇒ Pollution (Ozone, Nitrogen Dioxide, PM<sub>2.5</sub>)

#### Planned

- ⇒ US Elections Atlas (vote share)
- ⇒ USDA ERS County Typology Codes
- ⇒ National Neighbourhood Data Archive (tract and county amenities and social services)
- ⇒ The State Policies and Politics Database
- ⇒ Gun Ownership Proxy and State-Level Firearm Law Database
- ⇒ Structural racism and related state laws by year (state level)
- ⇒ Racial/Ethnic and economic segregation (computed from ACS/Census data) Racial resentment (state level)
- ⇒ Racial redlining
- ⇒ National Incidence-Based Reporting System 2021–present (NIBRS)

### FINDINGS TO DATE

Currently, the UAS website lists 316 journal publications and working papers, 416 media reports and 64 online posts.

### COVID-19-related papers

A large number of UAS-based publications (130 and counting) have used the COVID-19 tracking data we have collected between March 2020 and July 2022. The rapid adaptation of the questionnaire and the biweekly surveys provided unique data allowing researchers to address a wide spectrum of questions. These papers have investigated mental health effects, the prevalence of long covid, protective behaviour, vaccine hesitancy, discrimination of minorities,<sup>5</sup> race/ethnic differences in illness, job loss and the effect of benefit extensions on the psychological well-being of those who lost a job.<sup>6</sup> Several papers have documented the strain on parents of school age children during a period of remote learning and how COVID-19 has affected educational access of different groups.<sup>7–9</sup> We also estimated the positive psychological well-being effects of getting one's first covid vaccination.<sup>10</sup> Additionally, a detailed analysis of the COVID burst data showed shifts in affect and social contacts in the early months of the pandemic.<sup>11</sup> Other papers examined how the effects of COVID-19 varied by age and how social determinants of health explained infection rates by race/ethnicity. Analyses found that individuals who mainly trust left leaning media were considerably more likely to adopt protective behaviours and less likely to adopt risky behaviours compared with individuals who mainly trust right leaning media.<sup>12–15</sup>

Self-reports of physical activity showed a clear decrease between April 2020 and January 2021, followed by an increase from January 2021 to July 2021.<sup>16</sup>

### Papers using information derived from wearables and non-interview data

The pandemic also offered the opportunity to gauge the power of wearables in detecting illnesses. During the tracking period, respondents reported when they were diagnosed with, or thought they had, COVID-19. The Fitbits worn by UAS respondents picked up changes in sleep patterns and physical activity a few days before respondents reported an infection.<sup>4</sup>

Kapteyn *et al*<sup>17</sup> compared physical activity measured by self-reports and from wrist worn accelerometers across the Netherlands, England and the USA. They found that self-reports suggest little difference in physical activity between countries or across age groups, while the accelerometry suggests large differences; Americans are much less active than their European counterparts and older people are much less active than younger people. Kapteyn *et al*<sup>18</sup> compared physical activity measured by Fitbits, which provide feedback to wearers, with physical activity measured by GeneACTIV accelerometers, which do not provide feedback. They found that feedback provided

by Fitbits increased physical activity by 7%, at least in the short run.

The second-to-fourth digit length ratio of an individual's hand (digit ratio) is a putative biomarker for prenatal exposure to testosterone. Finley *et al*<sup>19</sup> examined the hypothesised negative association between the digit ratio and the preference for risk taking by asking UAS respondents which is longer: their index finger or ring finger. Their empirical findings support the hypothesis and suggest a meaningful biological basis for risk preferences.

## Experiments

As noted, the UAS facilitates experiments. For instance, Brown *et al*<sup>20</sup> conducted a randomised experiment with about 4000 adults in the UAS, in which they manipulated the complexity of how annuities are presented to respondents. They found that increasing the complexity of the annuity choice reduced respondents' ability to value the annuity, measured by the difference between the sell and buy values they assigned to the annuity. When they induced people to think first about how quickly or slowly to spend down assets in retirement, their ability to value an annuity increased.

Similarly Samek *et al*<sup>21</sup> noted that people have difficulty in understanding complex aspects of retirement planning, which leads them to under-use annuities and claim Social Security benefits earlier than is optimal. They developed vignettes about the consequences of different annuitisation and claiming decisions and then evaluated the vignettes using a sample of 2000 UAS respondents. In the experiment, respondents were either assigned to a control group with no vignette, to a written vignette or to a video vignette. They were then asked to give advice to hypothetical persons on annuitisation or Social Security claiming. Being exposed to vignettes led respondents to give better advice.

Perez-Arce *et al*<sup>22</sup> studied comprehension of Social Security terminology. They conducted an experiment in which they presented UAS respondents with different wordings for key Social Security concepts, such as Early Eligibility Age, Full Retirement Age and Delayed Retirement Credits. The content of the information treatments was identical for all respondents, but some were randomly given an alternative set of terms to refer to the key claiming ages (the experimental treatment group), while others were given the current terms (the control group). Despite these minimal changes, there were significant differences in outcomes. Those in the treatment group spent less time reading the information, but their understanding of the Social Security programme improved more than the control group. The effects were particularly strong for those with low levels of financial literacy. The relative gains in knowledge persisted several months after the treatment.

Burke *et al*<sup>23</sup> conducted a survey experiment examining whether short educational interventions can reduce adults' susceptibility to financial fraud. They found that online educational interventions can meaningfully

reduce fraud susceptibility, and that effects persist for at least 3 months if, and only if, a reminder is provided. They found no evidence that the educational interventions reduced willingness to invest generally, but rather increased knowledge, which participants were able to selectively apply. Effects were concentrated among individuals who are more likely to invest, particularly the financially sophisticated.

Two experiments have shown the impact of simple messaging manipulations on parent attitudes towards education policies. The first demonstrated that specific messaging around return-to-school post-COVID significantly impacted parents' attitudes about sending their own child back to in-person learning, if they had previously been unsure.<sup>24</sup> The second demonstrated that presenting parents with a brief statement highlighting the drawbacks of parents opting students out of lessons containing content they disagree with, significantly diminished respondents' tendencies to support schools' honouring those opt-out requests—and this effect was independent of political party identification.<sup>25</sup>

Mas and Pallais<sup>26</sup> conducted a field experiment in which job seekers were presented with different combinations of working hours and wages. The goal was to estimate the marginal value of non-working time. They repeated the experiment by offering UAS respondents similar, but hypothetical, choices. They found that the field experiment and the experiment in the UAS yielded very similar conclusions.

## Public policy

Allatar *et al*<sup>27</sup> exploited the longitudinal nature of the UAS to track the population's knowledge about Social Security programmes.

Bossert *et al*<sup>28</sup> used the UAS to examine the relationship between economic insecurity and political preferences. Constructing measures of economic insecurity from the variation in observed incomes over time, they related insecurity to voting preferences in the 2016 election. They found that economic insecurity increased the probability of voting for Trump.

UAS publications have contributed to highly polarised conversations about the inclusion of contested topics in education curricula. We have measured the accuracy of adults' understanding of Critical Race Theory, adult support for, and resistance to, specific lesbian, gay, bisexual, transgender, queer or questioning-related, race-related and sex-related topics in elementary and secondary school classrooms, opinions on book availability in school classrooms and libraries and sought to understand how much influence parents think they should have in school operations.<sup>29</sup> These surveys have pushed forward our understanding of seemingly intractable differences in how adults feel towards the purpose of the education system and how and what students should be learning in schools.

## Innovations in cognitive assessment

Scholars using the UAS have examined opportunities for web-based survey administration for cost-effective

cognitive functioning measurement. Liu *et al*<sup>30</sup> demonstrated the validity of self-administered online tests of perceptual speed and executive functioning. Gatz *et al*<sup>31</sup> developed a crosswalk to create a web-based classification of UAS participants' probability of cognitive impairment. Other papers have documented that response times to survey questions, which are stored as paradata alongside the answers in online surveys, can be used to infer cognitive abilities and detect mild cognitive impairment in UAS respondents.<sup>32–34</sup>

### Dissemination to the public and policymakers

Findings based on the UAS have appeared in well over 500 media reports, including national TV and radio broadcasts (eg, CBS and NPR) and influential newspapers and magazines (eg, *New York Times*, *The Washington Post*, *The Economist*). Results have been cited in congressional hearings<sup>35</sup> and in Senate discussions.<sup>36</sup>

## STRENGTHS AND WEAKNESSES

### Strengths

The use of the Internet, and the implied participants' self-administered surveys are likely to provide more reliable information on sensitive items as compared with face-to-face interviews.<sup>37 38</sup>

Probability-based Internet panels, such as the UAS mitigate selection problems facing opt-in or convenience Internet panels, may harm representativeness especially in older age groups and more disadvantaged segments of the population. Moreover, we and others have shown that widely used opt-in panels contain small but measurable shares of 'bogus respondents' (about 4%–7%, depending on the source), who tend to select positive answer choices, thus introducing systematic biases.<sup>39 40</sup> Moreover, respondents in convenience panels are incentivised to misrepresent their situation if they suspect that a particular answer may make them eligible for follow-up questions that will yield more financial rewards. It has been found for instance that 25% of Mturk respondents may endorse health conditions that do not exist.<sup>41</sup>

Prior research has shown that probability-based Internet panels provide superior information to convenience panels, even when recruiting rates are low.<sup>42–45</sup> In a comparison of the UAS with the HRS and the Current Population Survey (CPS), we found that UAS estimates of population parameters were very similar to those obtained from these traditional, high-quality population representative surveys.<sup>46</sup>

Attrition in the UAS is on a par with other population representative panels (see below). Other major probability-based Internet panels in the USA, such as Ipsos KnowledgePanel<sup>47</sup> and Amerispeak,<sup>48</sup> neither collect a substantial amount of core information, nor emphasise the collection of repeated observations on the same panel members as a central feature of the data. The KnowledgePanel reports that 'about 33% of the panel turns over annually'. Amerispeak cites an annual retention rate of

85%. The LISS (Longitudinal Internet Studies for the Social Sciences) panel in the Netherlands has a set-up similar to the UAS and reports annual respondent attrition of 12% per year.<sup>49</sup>

Figure 2 shows cumulative attrition of the first batch of UAS respondents, recruited in 2014 (graphs for later batches are very similar, but obviously cover fewer years). A respondent is considered to have attrited after 1 year of inactivity. The sample loss of 44% in 2023 amounts to an annual attrition rate of 6.2% over 9 years, with a concentration in the early years. Attrition in later years is converging to about 2.5% annually. Projecting an attrition rate of 2.5% going forward would mean that after 15 years, approximately 52% of the original sample would remain.

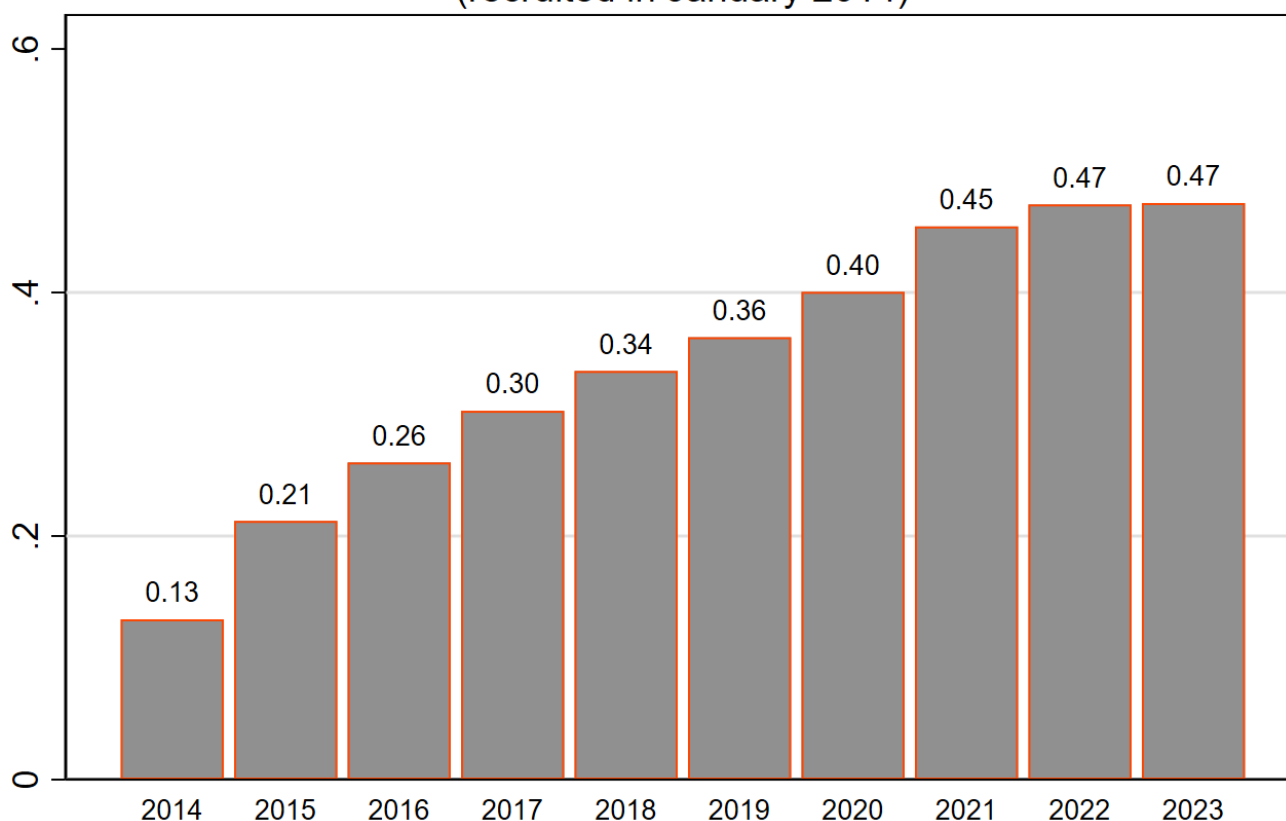
Many of the non-Internet panels in the US cover a specific age range. Most comparable in terms of population coverage are the PSID and the Survey of Income and Programme Participation (SIPP). Zabel<sup>50</sup> noted that first year attrition in the PSID was 11.9%, and 51.2% after 21 years. The implied 48.8% retention rate after 21 years is similar to the estimated retention rate of the UAS after 15 years. The PSID attrition rate has fallen further since Zabel published his analysis. Sastry *et al*<sup>51</sup> report an average attrition of about 3.5% per 2 years in the PSID since 1997. The most recent SIPP redesign was implemented in 2014. Attrition rates between the first and second, and the second and third waves of the SIPP, were 25.8% and 20.0%, respectively.

Thus, other probability-based Internet panels have higher attrition rates, while only the Dutch LISS panel collects a core of longitudinal data on its respondents. The PSID has a lower attrition rate, but also collects much less information at lower frequency. For example, a biennial panel like PSID may have realised three surveys over a 6 year period, while over the same period UAS members may have answered 75 (short) surveys or more. The UAS does not only yield more information, but also with much more granularity, as for instance when collecting data on health and financial well-being during the COVID-19 pandemic. Of course, the caveat here is that the type of information will be partly different, while the UAS and PSID samples may be different, if only because the recruiting response rate in the UAS is much lower than in the PSID (The response rate (AAPOR RR1) to the initial invitation letter is somewhat over 20%, with about a 75% consent rate. Including drop out in the stage from consent to actual participation leads to a recruiting response rate of about 10%, cf. figure 1.)

New UAS data collection can be implemented quickly in response to new events. Prompt and user-friendly data dissemination (both for the research community and the general public) optimises opportunities for the research community to use the data and to influence what is being collected.

The technology allows for quick responses to new developments. Collecting information closer in time to the occurrence of important events, and collecting it

## Cumulative Attrition Batch 1 (recruited in January 2014)



**Figure 2** Attrition over time of the first batch recruited in 2014.

repeatedly reduces the role of measurement error and omitted variables, enabling sharper statistical inference. For example, UAS fielded the first COVID tracking survey on 10 March 2020 and continued tracking until the summer of 2022. Prospective data collection minimises the need for reliance on respondents' recollection of past events. Furthermore, by opening the survey to the wider research community and by insisting on data sharing, any new data collection can build on the wealth of data already available.

### Weaknesses

Relying almost exclusively on self-interviewing and the Internet adds greatly to the flexibility and speed of data collection, while reducing costs associated with maintaining a large corps of interviewers. However, this also limits the types of interactions we can have with participants. The panel recruiting rate currently hovers around 10%. This appears somewhat better than other probability-based Internet panels in the USA, but is clearly lower than traditional population surveys that rely on face-to-face recruiting. Although, as noted above, this does not appear to affect representativeness, a higher recruiting rate would be desirable. Internet interviewing

also limits our ability to collect biomarker data, as these must be collected by respondents themselves without the help of an in-person interviewer. The collection of saliva for the collection of DNA is our first foray in the use of collecting biomarkers without the involvement of an interviewer or health professional.

### DATA ACCESS

UAS survey data and UAS longitudinal and processed files are available for download from the UAS website (<https://uasdata.usc.edu/index.php>) for any researcher who has signed a Data User Agreement (DUA). We also provide restricted data through the UAS enclave.

We distinguish three tiers of data security (<https://uasdata.usc.edu/page/Data+Overview>):

- ▶ Tier 1: coded data that include no direct or indirect identifiers and no open-text fields.
- ▶ Tier 2: sensitive data—for example, political data; UAS data linked to external data at other than state level by UAS staff. No location indicators, or only blinded indicators. Does not include any direct or indirect identifiers.



- Tier 3: restricted access data—available only in the DataLinkage enclave. This includes UAS survey data linked with external programme data from SSA or CMS, the NDI, social media data, location crosswalks used to link at the census track, zip code or county level.

Our philosophy has always been to share UAS data as widely as possible without compromising data security and respondents' privacy. This takes the form of a number of concrete policies:

- In principle, data are made available to the research community immediately after the field period of any survey.
- This applies to all data that are funded by Center for Economic and Social Research (CESR) or by grants acquired by CESR.
- External users with their own funding are also required to follow the same policy, but limited embargo periods (up to 6 months, in exceptional cases up to a year) may be arranged.
- Data from UAS surveys can be linked so that any researcher who proposes a new UAS survey benefits from the wealth of information already available. This creates a user community in which researchers benefit from the efforts of earlier contributors, and at the same time contribute themselves.
- The provision of data right after the end of a field period implies that generally no data processing takes place. Quality control (eg, answers out of range, or inconsistent answers across questions) is part of the survey process itself.
- The main exception to fast dissemination has been the construction of some longitudinal files. A prime example is the 'Comprehensive File' (CF), which comprises all UAS longitudinal core information listed in [table 1](#) and includes derived variables (in particular, income and wealth variables from the HRS modules) and checks for outliers in these variables. Since core information is collected every 2 years and respondents have joined the UAS at different times, new data keep coming in continually. New versions of the CF based on the most recent data are posted every 3 months.
- Naturally, data from the CF and other UAS longitudinal files can be linked to data from other UAS surveys.

## COLLABORATION

The cooperative agreement (1U01AG077280) that funds most of the UAS activities includes more than 30 key personnel from universities in the USA and Europe. Disciplines include economics, psychology, sociology, demography, epidemiology, bioengineering, environmental health, survey methodology, econometrics and political science. Furthermore, the Data Monitoring Committee includes 12 members covering economics, epidemiology, survey methods, health economics, sociology and psychology.

Outside of the data collected through the cooperative agreement, the UAS accommodates surveys by other researchers, who have funds to conduct their own research. The collected data can be linked to the already existing longitudinal data. So far, some 65 researchers have taken advantage of this capability.

The yearly CIPHER conference (Current Innovations in Probability-based Household Internet Panel Research: <https://dornsife.usc.edu/cesr/cipher-2024/>) brings together some 150 researchers with a methodological interest in probability-based Internet panels.

**Correction notice** This article has been corrected since it was first published. Funding statement has been updated.

**X** Tania Gutsche @spack

**Acknowledgements** This paper has benefited greatly from comments by the many coinvestigators of the cooperative agreement, in particular Arthur Stone, Jeremy Burke, Bas Weerman, Bart Oriens, Doerte Junghaenel, Amie Rappaport, Eileen Crimmins, Francisco Perez-Arce, Lila Rabinovich, Stefan Schneider, Margaret Gatz, Jennifer Ailshire, Ying Liu, Ricky Bluthenthal, Brian Finch, David Weir, Rob Scot McConnell, Ashlesha Datar, Shannon Monnat and Evan Sandlin.

**Contributors** AK is the guarantor for the overall content. AK conceptualised The Understanding America Study. AK, TG, MA and JD jointly developed the study. The concept of the article was developed jointly by AK, JD, TG and MA. AK wrote the first draft. JD, TG and MA commented on the first and second drafts. AK finalised the draft. All authors are employed at Center for Economic and Social Research at the University of Southern California.

**Funding** The majority of funding for the UAS comes from a cooperative agreement with the National Institute on Aging of the National Institutes of Health (U01AG077280), with substantive cofunding from the Social Security Administration.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants. Oversight of data collection and dissemination was ceded by University of Southern California IRB (UP-14-00148) to BRANY (Biomedical Research Alliance of New York) IRB (22-030-1044). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. Data are available to the research community, subject to signing data use agreements appropriate for the data tier one is requesting.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Arie Kapteyn <http://orcid.org/0000-0002-1855-5528>

Marco Angrisani <http://orcid.org/0000-0002-6084-6388>

## REFERENCES

- 1 Jin H, Kapteyn A. Relationship Between Past Survey Burden and Response Probability to a New Survey in a Probability-Based Online Panel. *J Off Stat* 2022;38:1051–67.
- 2 Stone AA, Schneider S, Smyth JM. Shedding light on participant selection bias in Ecological Momentary Assessment (EMA) studies: Findings from an internet panel study. *PLoS ONE* 2023;18:e0282591.
- 3 Dick C. The Health and Retirement Study: Contextual Data Augmentation. *Forum Health Econ Policy* 2022;25:29–40.

- 4 Chaturvedi RR, Angrisani M, Troxel WM, *et al.* American Life in Realtime: a benchmark registry of health data for equitable precision health. *Nat Med* 2023;29:283–6.
- 5 Liu Y, Finch BK, Brenneke SG, *et al.* Perceived Discrimination and Mental Distress Amid the COVID-19 Pandemic: Evidence From the Understanding America Study. *Am J Prev Med* 2020;59:481–92.
- 6 Liu Y, Mattke S. Association between state stay-at-home orders and risk reduction behaviors and mental distress amid the COVID-19 pandemic. *Prev Med* 2020;141:106299.
- 7 Haderlein SK, Saavedra AR, Polikoff M, *et al.* Educational access by race/ethnicity, income, and region in the time of COVID: evidence from a nationally representative panel of American families. *AERA Open*, 2021.
- 8 Haderlein SK, Saavedra AR, Polikoff MS, *et al.* Disparities in Educational Access in the Time of COVID: Evidence From a Nationally Representative Panel of American Families. *AERA Open* 2021;7:23328584211041350.
- 9 Silver D, Polikoff M, Saavedra A, *et al.* The Subjective Value of Postsecondary Education in the Time of COVID: Evidence from a Nationally Representative Panel. *Peabody J Educ* 2022;97:344–68.
- 10 Perez-Arce F, Angrisani M, Bennett D, *et al.* COVID-19 vaccines and mental distress. *PLoS ONE* 2021;16:e0256406.
- 11 Mak HW, Wang D, Stone AA. Momentary social interactions and affect in later life varied across the early stages of the COVID-19 pandemic. *PLoS One* 2020;17:e0267790.
- 12 Kim JK, Crimmins EM. How does age affect personal and social reactions to COVID-19: Results from the national Understanding America Study. *PLoS ONE* 2020;15:e0241950.
- 13 Lee H, Andrasfay T, Riley A, *et al.* Do social determinants of health explain racial/ethnic disparities in COVID-19 infection? *Soc Sci Med* 2022;306:115098.
- 14 Wu Q, Ailshire JA, Crimmins EM. Long COVID and symptom trajectory in a representative sample of Americans in the first year of the pandemic. *Sci Rep* 2022;12:11647.
- 15 Zhao E, Wu Q, Crimmins EM, *et al.* Media trust and infection mitigating behaviours during the COVID-19 pandemic in the USA. *BMJ Glob Health* 2020;5.
- 16 Wijngaards I, Pozo Cruz B, Gebel K, *et al.* Exercise frequency during the COVID-19 pandemic: a longitudinal probability survey of the US population. *Prev Med Rep* 2022;25:101680.
- 17 Kapteyn A, Banks J, Hamer M, *et al.* What they say and what they do: comparing physical activity across the USA. *Engl Neth J Epidemiol Community Health* 2018;72:471–6.
- 18 Kapteyn A, Saw H-W, Darling J. Does feedback from activity trackers influence physical activity? evidence from a randomized controlled trial. *JMIR Form Res* 2021;24:34460.
- 19 Finley B, Kalwij A, Kapteyn A. Born to be wild: Second-to-fourth digit length ratio and risk preferences. *Econ Hum Biol* 2022;47:101178.
- 20 Brown JR, Kapteyn A, Luttmner EFP, *et al.* Behavioral Impediments to Valuing Annuities: Complexity and Choice Bracketing. *Rev Econ Stat* 2021;103:533–46.
- 21 Samek A, Kapteyn A, Gray A. Using vignettes to improve understanding of Social Security and annuities. *J Pension Econ Finance* 2022;21:326–43.
- 22 Perez-Arce F, Rabinovich L, Yoong J, *et al.* Three little words? The impact of social security terminology on knowledge and claiming intentions. *J Pension Econ Finance* 2024;23:132–51.
- 23 Burke J, Kieffer C, Mottola G, *et al.* Can educational interventions reduce susceptibility to financial fraud? *J Econ Behav Organ* 2022;198:250–66.
- 24 Polikoff MS, Silver D, Garland M, *et al.* The Impact of a Messaging Intervention on Parents' School Hesitancy During COVID-19. *Educ Res* 2022;51:156–9.
- 25 Saavedra AS, Polikoff MS, Silver D, *et al.* Searching for common ground: widespread support for public schools but substantial partisan divides about teaching contested topics. 2024.
- 26 Mas A, Pallais A. Labor Supply and the Value of Non-Work Time: Experimental Estimates from the Field. *Am Econ Rev Insights* 2019;1:111–26.
- 27 Alattar L, Messel M, Rogofsky D, *et al.* The Use of Longitudinal Data on Social Security Program Knowledge. *Soc Secur Bull* 2019;79:1–9.
- 28 Bossert W, Clark AE, D'Ambrosio C, *et al.* Economic insecurity and political preferences. *Oxf Econ Pap* 2023;75:802–25.
- 29 Polikoff M, Fienberg M, Silver D, *et al.* n.d. Who wants to say 'Gay?' Public opinion about LGBT issues in the curriculum. *J LGBT Youth* 2024;1–22.
- 30 Liu Y, Schneider S, Orriens B, *et al.* Self-administered Web-Based Tests of Executive Functioning and Perceptual Speed: Measurement Development Study With a Large Probability-Based Survey Panel. *J Med Internet Res* 2022;24:e34347.
- 31 Gatz M, Schneider S, Meijer E, *et al.* Identifying Cognitive Impairment Among Older Participants in a Nationally Representative Internet Panel. *J Gerontol B Psychol Sci Soc Sci* 2023;78:201–9.
- 32 Junghaenel DU, Schneider S, Orriens B, *et al.* Inferring Cognitive Abilities from Response Times to Web-Administered Survey Items in a Population-Representative Sample. *J Intell* 2023;11:3.
- 33 Schneider S, Junghaenel DU, Meijer E, *et al.* Using Item Response Times in Online Questionnaires to Detect Mild Cognitive Impairment. *The J Gerontol* 2023;78:1278–83.
- 34 Jin H, Junghaenel DU, Orriens B. Developing Early Markers of Cognitive Decline and Dementia Derived From Survey Response Behaviors: Protocol for Analyses of Preexisting Large-scale Longitudinal Data. *JMIR Res Protoc* 2023;12:e44627.
- 35 Social Security Administration (SSA) commissioner nominee Martin O'Malley's confirmation hearing before the Senate Finance Commission. 2023.
- 36 Warren E, Murphy CS, Booker CA, *et al.* Letter to majority leader schumer and minority leader McConnell. Senate US, 2021.
- 37 Newman JC, Des Jarlais DC, Turner CF, *et al.* The differential effects of face-to-face and computer interview modes. *Am J Public Health* 2002;92:294–7.
- 38 Perlis TE, Des Jarlais DC, Friedman SR, *et al.* Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. *Addiction* 2004;99:885–96.
- 39 Kennedy C, Hatley N, Lau A, *et al.* Assessing the risks to online polls from bogus respondents. Pew Research Center, 2020.
- 40 Schneider S, May M, Stone AA. Careless responding in internet-based quality of life assessments. *Qual Life Res* 2018;27:1077–88.
- 41 Hays RD, Qureshi N, Herman PM, *et al.* Effects of Excluding Those Who Report Having 'Syndromitis' or 'Chekalism' on Data Quality: Longitudinal Health Survey of a Sample From Amazon's Mechanical Turk. *J Med Internet Res* 2023;25:e46421.
- 42 Malhotra N, Krosnick JA. The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples. *Polit anal* 2007;15:286–323.
- 43 Yeager DS, Krosnick JA, Chang L. Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Pub Opin Q* 2011;75:709–47.
- 44 Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality. *Pub Opin Q* 2009;73:641–78.
- 45 MacInnis B, Krosnick JA, Ho AS, *et al.* The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension. *Public Opin Q* 2018;82:707–44.
- 46 Angrisani M, Finley B, Kapteyn A. Can internet match high-quality traditional surveys? Comparing the health and retirement study and its online version. In: *The econometrics of complex survey data: theory and applications*. Emerald Publishing Limited, 2019: 3–33.
- 47 Ipsos. Knowledge panel. n.d. Available: <https://www.ipsos.com/en-us/solutions/public-affairs/knowledgepanel>
- 48 Amerispeak. Available: <https://amerispeak.norc.org/Documents/Research/AmeriSpeak%20Technical%20Overview%202019%2002%2018.pdf>
- 49 Centerdata. LISS panel. n.d. Available: <https://www.lissdata.nl/about-panel/composition-and-response>
- 50 Zabel JE. An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior. *J Hum Resour* 1998;33:479.
- 51 Sastry N, Fomby P, McGonagle KA. Effects on Panel Attrition and Fieldwork Outcomes from Selection for a Supplemental Study: Evidence from the Panel Study of Income Dynamics. *Adv Longit Surv Methodol* 2021;74–99.