Comparing Two Types of Online Survey Samples

Opt-in samples are about half as accurate as probability-based panels

BY *Andrew Mercer and Arnold Lau*

About Pew Research Center

Pew Research Center is a nonpartisan, nonadvocacy fact tank that informs the public about the issues, attitudes and trends shaping the world. It does not take policy positions. The Center conducts public opinion polling, demographic research, computational social science research and other data-driven research. It studies politics and policy; news habits and media; the internet and technology; religion; race and ethnicity; international affairs; social, demographic and economic trends; science; research methodology and data science; and immigration and migration. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

How we did this

Pew Research Center designed this study to assess the current state of online survey sampling methods – both probability-based and opt-in – and determine their accuracy on general population estimates for all U.S. adults and on estimates for key demographic subgroups. To do this, we administered a common questionnaire to samples obtained from three probability-based online panels, one of which was the Center's American Trends Panel (ATP), and three online opt-in sample providers. The surveys were administered between June 14 and July 21, 2021, and included interviews with a total of 29,937 U.S. adults.

The target size for each sample was 5,000 adults to minimize the impact of sampling error on estimates for demographic subgroups that comprise only a small percentage of the total U.S. population. Because the purpose of this study is methodological, the names of the vendors are masked. Instead, we refer to the samples as probability panels 1, 2 and 3 and opt-in samples 1, 2, and 3 throughout this report.

The probability-based panels use traditional probability-based methods for recruiting a random sample of U.S. adults. Specifically, all three panels use address-based sampling (ABS) for panel recruitment. ABS begins with a random sample of households from the U.S. Post Office's Delivery Sequence File, a near complete list of all residential addresses in the United States. Individuals in sampled households are contacted via mail and invited to join the panel and continue taking surveys periodically online. Although all three panels use similar methods for recruitment, differences in the timing and design of recruitments, the use of incentives, sampling procedures for individual panel waves, and panel maintenance practices could result in samples that are not altogether comparable to one another.

The three opt-in samples in this study are based on different but common approaches to online opt-in sampling. Opt-in sample 1 comes from a panel aggregator, or marketplace, in which individual respondents are drawn from many opt-in sample sources that have agreed to make their sample available to the aggregator. Opt-in sample 2 is sourced entirely from a single opt-in panel. Opt-in sample 3 is a blend, with about three-fifths sourced from a single opt-in panel and the remainder sourced from three sample aggregators. All three opt-in samples use a common set of quotas on age by gender, race and ethnicity, and educational attainment.

The same weighting scheme was applied to all six samples following Pew Research Center's standard procedure for weighting the ATP. Complete details of the weighting procedure and the design of the individual samples can be found in the methodology.

Terminology

Probability-based panel: This refers to a national survey panel recruited using random sampling from a database that includes most people in the population. Today, most such panels in the United States recruit by drawing random samples of residential addresses or telephone numbers. Typically, data collection with these panels is done online. However, some of these panels interview a small fraction of respondents (usually about 5% or fewer) using an offline mode such as live telephone. These panels are "probability-based" because the chance that each address or phone number is selected is known. However, the chance that each selected person will join the panel or take surveys after joining is not known.

Online opt-in sample: These samples are recruited using a variety of methods that are sometimes referred to as "convenience sampling." Respondents are not selected randomly from the population but are recruited from a variety of online sources such as ads on social media or search engines, websites offering rewards in exchange for survey participation, or self-enrollment in an opt-in panel. Some opt-in samples are sourced from a panel (or multiple panels), while others rely on intercept techniques where respondents are invited to take a one-off survey.

Benchmark: These are "ground-truth" measures used to assess the <u>accuracy of survey estimates</u>. For example, survey-based estimates for the share of voters who voted for each candidate in the 2020 presidential election are compared to a benchmark based on an official tally by the Federal Election Commission (FEC). Survey estimates are deemed more accurate the closer they are to the benchmark value. In this study, the benchmarks come from high-quality federal surveys such as the American Community Survey or administrative records like the FEC vote tally. Although these benchmarks come from "gold-standard" data sources, they are not entirely free from error. As such, they are not "true" population values but rather the best available approximations.

Error: This is the difference between an individual survey estimate and its corresponding benchmark value. Error can be either positive or negative depending on whether the survey estimate is higher or lower than the benchmark. For example, the FEC benchmark for the share of voters who voted for Donald Trump in the 2020 presidential election is 47%. If a survey estimated that share to be 42%, the error would be -5 percentage points because it came in 5 points below the benchmark. If the estimate were 49%, the error would be +2 points.

Absolute error: This is the absolute value of the error for a survey estimate. It describes the *size* of the error irrespective of its direction (positive or negative). For example, two estimates that have error of +5 and -5 percentage points, respectively, both have an absolute error of 5 points.

Average absolute error (for a benchmark variable): This is a measure that summarizes the average size of errors across all the categories within a single benchmark variable. For example, the smoking status variable has four categories: 1) Smoke every day, 2) Smoke some days, 3) Do not now smoke and 4) Never smoked 100 cigarettes. A survey's estimates for each category will have different levels of error, both positive and negative. For a given survey, the absolute error for the smoking status variable is equal to the sum of the absolute errors for each category divided by the number of categories.

Average absolute error (for a sample): Average absolute error can also be used to summarize the overall level of error across many different benchmarks within a single sample. When used in this context, the average absolute error for a sample is equal to the sum of the average absolute errors for each benchmark variable divided by the total number of benchmark variables.

Example: Measuring error for a benchmark variable

Smoking status	Survey estimate	Benchmark value	Error	Absolute error
Smoke every day	15%	9%	+6	6
Smoke some days	8%	3%	+5	5
Do not now smoke	19%	22%	-3	3
Never smoked 100 cigarettes	55%	63%	-8	8

Average absolute error = (6 + 5 + 3 + 8)/4 = 5.5

Note: Benchmark values for smoking are from the 2020 National Health Interview Survey. The survey estimates shown were created as an example and do not correspond to any of the samples included in this study.

PEW RESEARCH CENTER

Comparing Two Types of Online Survey Samples

Opt-in samples are about half as accurate as probability-based panels

As the field of public opinion research continues its steady movement toward online data collection, probability-based panels and opt-in samples have emerged as the two most common approaches to surveying individuals online. At the same time, the methodologies and industry practices for both kinds of samples are evolving.

To shed light on the current state of online probability-based and opt-in samples, Pew Research Center conducted a study to compare the accuracy of six online surveys of U.S. adults – three from probability-based panels and three from opt-in sources. This is the first such study to include samples from multiple probability-based panels, allowing for their side-by-side comparison.

The surveys in this study were administered between June 14 and July 21, 2021, and included interviews with a total of 29,937 U.S. adults, approximately 5,000 in each sample. Because this is a methodological study, the names of the sample providers are masked.

The study compared each sample's accuracy on 28 benchmark variables drawn from high-quality government data sources. These benchmarks included a variety of measures on topics such as voting, health, and respondents' work, family and living situations. (Refer to the appendix for the full list of benchmarks and their sources.)

The study's key findings include:

On average, error on opt-in samples was about twice that of probability-based panels

For estimates among U.S. adults on 28 benchmark variables, opt-in samples 1, 2 and 3 had average absolute errors of 6.4, 6.1 and 5.0, respectively, for an overall average of 5.8 percentage points. This was about twice that of the probability-based online panels, for which average absolute error was 2.6 points overall (2.3, 3.0 and 2.5 on probability panels 1, 2 and 3, respectively).

Online opt-in samples had especially large errors for 18- to 29-year-olds and Hispanic adults

On 25 variables for which subgroup-level benchmarks were available, the online opt-in samples averaged 11.2 percentage points of error for 18- to 29-year-olds and 10.8 points for Hispanic adults– each about 5 points higher than for U.S. adults overall (6.4 points on the same 25 variables). By comparison, the average absolute error on the probability-based panels was 3.6

Larger errors on online opt-in samples, especially for 18- to 29-year-olds and Hispanic adults

Average absolute error on benchmark variables
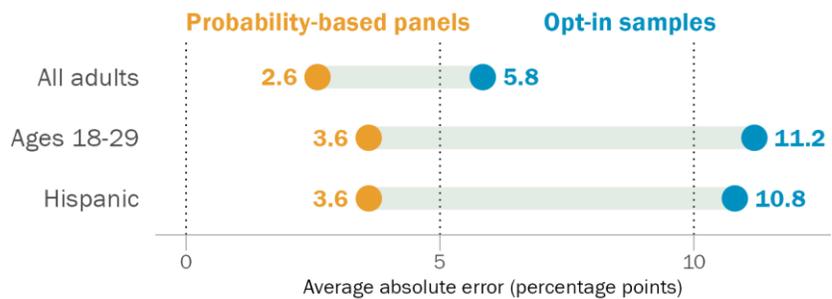
Note: Results shown are the mean value of the average absolute errors across three probability-based panels and across three online opt-in samples. Results for all adults are based on 28 benchmark variables. Results for 18- to 29-year-olds and Hispanic adults are based on 25 variables for which reliable subgroup benchmarks are available.
Source: Pew Research Center analysis of six online samples surveyed June 14-July 21, 2021.

PEW RESEARCH CENTER

points for both young adults and Hispanic adults, less than 2 points higher than the error for all adults. A similar level of error was seen on the probability-based panels for other traditionally hard-to-survey subgroups such as those with no more than a high school education (3.6 points) and non-Hispanic Black adults (3.8 points).

Error was concentrated in a handful of variables on the probability-based panels but widespread on the opt-in samples

On each of the probability-based panels, the number of benchmarks for which average absolute error was greater than 5 percentage points ranged from two to five out of 28. About half (between 14 and 15 benchmarks) had under 2 points of average absolute error. Large errors were more widespread on the opt-in samples, which had between 11 and 17 benchmarks with error greater than 5 points. Fewer benchmarks (between three and seven) on the opt-in samples had average absolute error below 2 points.

Probability-based panels consistently overestimated 2020 voter turnout

The only benchmark that had consistently high error on all three probability-based panels was voter turnout in the 2020 presidential election, which they each overestimated by +8 or +9 percentage points. This suggests that despite the inclusion of volunteerism and voter registration in weighting adjustments, the overrepresentation of politically and civically engaged individuals remains an important challenge for probability-based panels. By contrast, turnout was one of the

most accurate variables on the opt-in samples, two of which came within 1 point of the benchmark while the third exceeded the benchmark by +3 points.

Much of the error on the opt-in samples appears to be due to 'bogus respondents,' who usually say 'Yes' regardless of the question

In the online opt-in samples, an average of 8% of all adults, 15% of 18- to 29-year-olds and 19% of Hispanic adults answered "Yes" on at least 10 out of 16 Yes/No questions that were asked of every respondent. The corresponding shares on the probability-based panels were between 1% and 2% for each group. Similarly large shares reported having received at least three of four types of government benefits (Social Security, food stamps, unemployment compensation or workers' compensation) even though such individuals are virtually nonexistent in the true population. It is highly unlikely that the few individuals who do fit that description are massively overrepresented in online opt-in samples. Instead, this pattern suggests that much of the error on the online opt-in samples is due to the presence of "bogus respondents," who make little or no effort to answer questions truthfully.

What is new about this benchmarking study?

Polling insiders may wonder why we conducted this study. Other research teams (including our own several years ago) have done similar comparisons and arrived at the same result: Probability-based samples tend to be more accurate, even if recent election polls are an exception.

One major reason we conducted this study is because it offers something new. This benchmarking study is the first to estimate the accuracy of *multiple* online probability-based panels in the United States. This allows us to answer a previously unaddressed question: Do different probability-based panels tend to offer similar data quality or not? The three probability-based panels tested in this study performed about the same. The average absolute error ranged from just 2.3 to 3.0 percentage points – good news for researchers seeking a reliable polling method. It's worth noting that this study considered fairly general topics (e.g., employment, marital status), and that the results might differ if a survey focused on a more niche topic, like poverty.

The second major reason we conducted this study was that in the time since our 2016 study, there have been a number of major changes to the ATP's methodology made in response to that study's findings, other challenges that have arisen in the intervening years, and the Center's evolving research needs:

- We changed how we recruit, moving from telephone to an address-based approach.
- We began subsampling the panel rather than routinely interviewing everyone.

- We changed how we weight the panel, adding in adjustments for volunteerism, religion and other factors.
- We began retiring panelists from overrepresented segments of the population.

The study reported here is, in part, our effort to measure whether those improvements made a difference, allowing us to determine how the new, improved ATP stacks up against opt-in samples and against other probability-based panels.

1. Assessing the accuracy of estimates for U.S. adults

To gauge each sample's accuracy on general population estimates for all U.S. adults, we calculated the weighted percentage of adults belonging to 77 categories across 28 different variables and compared them to corresponding benchmarks derived from high-quality government data sources. The benchmarks covered a variety of topics, including voting, health, and respondents' work, family and living situations. (Refer to the appendix for the full list of benchmarks and their sources.)

Because many of the benchmark variables included more than one category, we calculated each

Average error on online opt-in samples was twice that of probability-based panels

Average absolute error on 28 benchmark variables for estimates among U.S. adults

Sample	Average absolute error
Opt-in mean	5.8
Probability mean	2.6
Opt-in 1	6.4
Opt-in 2	6.1
Opt-in 3	5.0
Probability 2	3.0
Probability 3	2.5
Probability 1	2.3

Average absolute error (percentage points)

Source: Pew Research Center analysis of six online samples surveyed June 14-July 21, 2021.
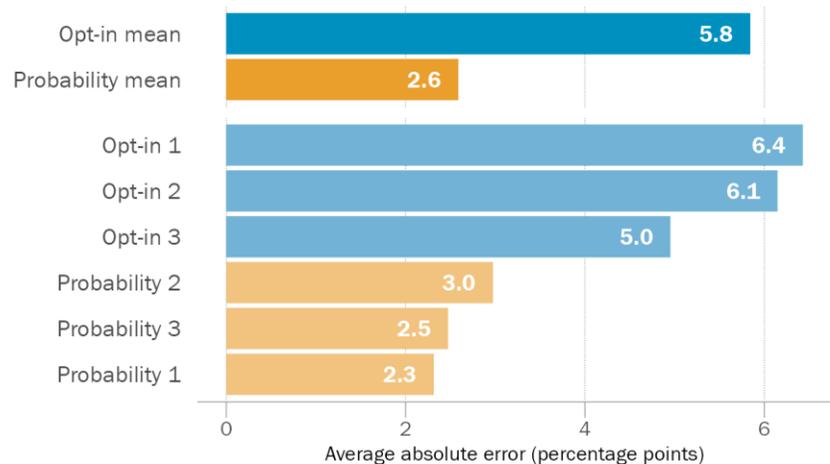
PEW RESEARCH CENTER

variable's average absolute error – that is, the average of the absolute differences between the survey estimate and a corresponding benchmark value for each category – to compare the relative accuracy of variables that have different numbers of categories. To facilitate more general comparisons between samples overall, we also calculated the average absolute error for each sample as the mean of the average absolute errors across all 28 benchmarks.

In any study of this kind, it is important to note that the performance of any given sample depends on numerous factors such as the variables used in weighting or the specific topics included in the benchmarks. It is possible that the relative accuracy of each sample might differ if we had used a different weighting scheme or chosen a different set of benchmarks for comparison. Furthermore, not even "gold-standard" government statistics are entirely free from error. Consequently, the measures of error discussed in this report should be considered approximate.

Across all 28 benchmarks combined, the probability-based panels had a mean average absolute error of 2.6 percentage points for estimates among all U.S. adults. The error for individual probability-based panels ranged from a low of 2.3 points for probability panel 1 to a high of 3.0 points for probability panel 2. The average absolute error for the opt-in samples combined was

about twice as large at 5.8 points. Of these, opt-in sample 3 had the lowest average error at 5.0 points. Opt-in samples 1 and 2 exhibited higher error on average with 6.4 and 6.1 points, respectively.

For ease of explanation, individual benchmark variables whose average absolute error was less than 2 percentage points were classified as having "low" error. Variables with more than 5 points of average absolute error were defined as having "high" error, and the remainder were coded as having "medium" error. These particular groupings were chosen because they each contain about one-third of all benchmarks from all six samples. It is important to note that these designations of low, medium or high error are relative to the specific benchmarks and samples included in this study.

Opt-in samples had many more benchmarks with large errors than probability-based panels

Average absolute error for estimates among U.S. adults on 28 benchmark variables. About half of the benchmarks had more than 5 points of error on the opt-in samples compared with only a handful on the probability-based panels.
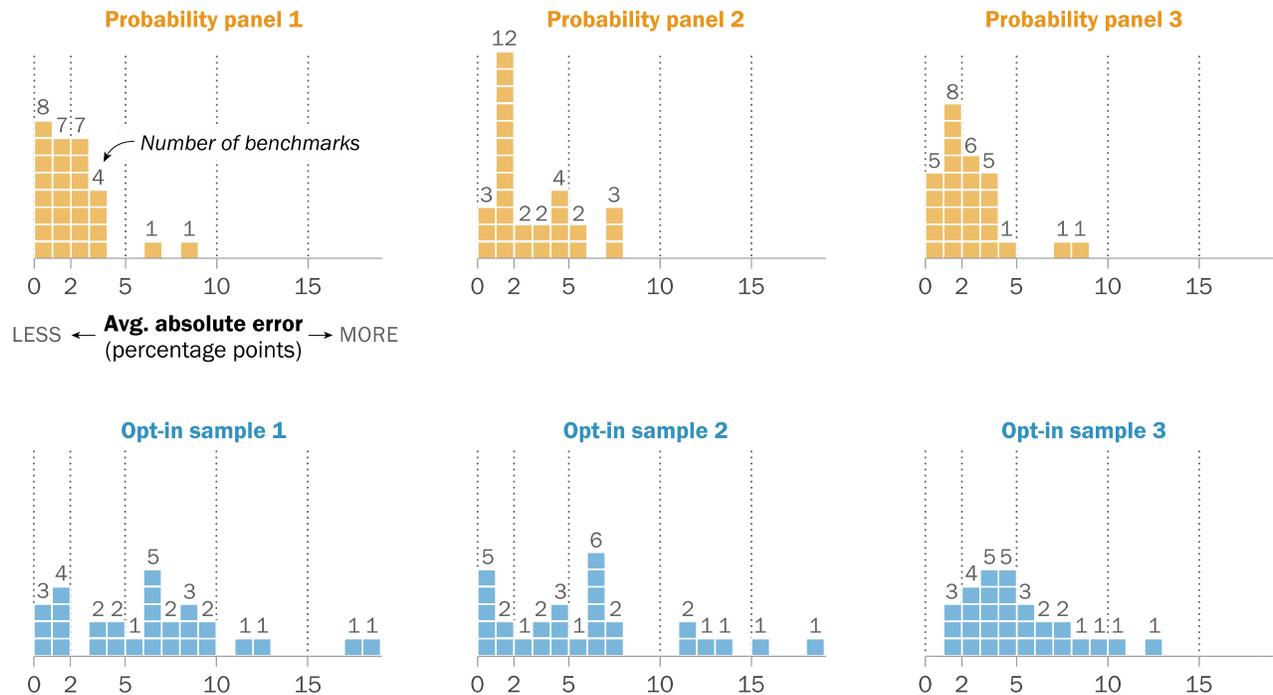
Note: Each box represents one benchmark variable out of 28 that were measured on each sample.
Source: Pew Research Center analysis of six online samples surveyed June 14-July 21, 2021.

PEW RESEARCH CENTER

Collectively, about half of all benchmarks on the probability-based panels fell into the low error category (44 out of 84) while 11% were classified as high error (9 out of 84). The only variable with consistently high error on the probability-based panels was voter turnout in the 2020 presidential election, for which all three samples overestimated the benchmark by +8 or +9 percentage points. This is consistent with our 2016 study, which found civically engaged adults to be overrepresented on the American Trends Panel, and suggests that the problem afflicts ABS-recruited panels more generally.

Level of error on benchmark variables

Number of benchmark variables with low, medium or high average absolute error on estimates among U.S. adults

	Total benchmarks	Low error	Medium error	High error
Probability panel 1	28	15	11	2
Probability panel 2	28	15	8	5
Probability panel 3	28	14	12	2
Total for probability-based panels	**84**	**44**	**31**	**9**
Opt-in sample 1	28	7	4	17
Opt-in sample 2	28	7	6	15
Opt-in sample 3	28	3	14	11
Total for opt-in samples	**84**	**17**	**24**	**43**

Note: Benchmark classifications are based on average absolute error: under 2 points (low), between 2 and 5 points (medium) and greater than 5 points (high).
Source: Pew Research Center analysis of six online samples surveyed June 14-July 21, 2021.

PEW RESEARCH CENTER

This overall pattern was reversed for the opt-in samples, on which 51% of benchmarks fell in the high error category, compared with 20% in the low error category. There were 10 variables with consistently high error on the opt-in samples. Four involved the receipt of certain government benefits in the prior year. On average, the opt-in samples overestimated the shares of adults receiving food stamps (+16 points), Social Security (+15 points), unemployment compensation (+10 points) and workers' compensation (+9 points).

Another three variables with consistently high error related to employment, with the opt-in samples underestimating the share of all adults who worked for pay in the prior week by -12 points on average and the share of adults who worked at a job or business at any time in the prior year by -11 points on average. The opt-in samples also overstated the share of adults who were employed but had been unable to work in the past month because of COVID-19 by an average of +7 points.

Samples of the same type generally had similar levels of error on individual benchmarks

Average absolute error for estimates among U.S. adults
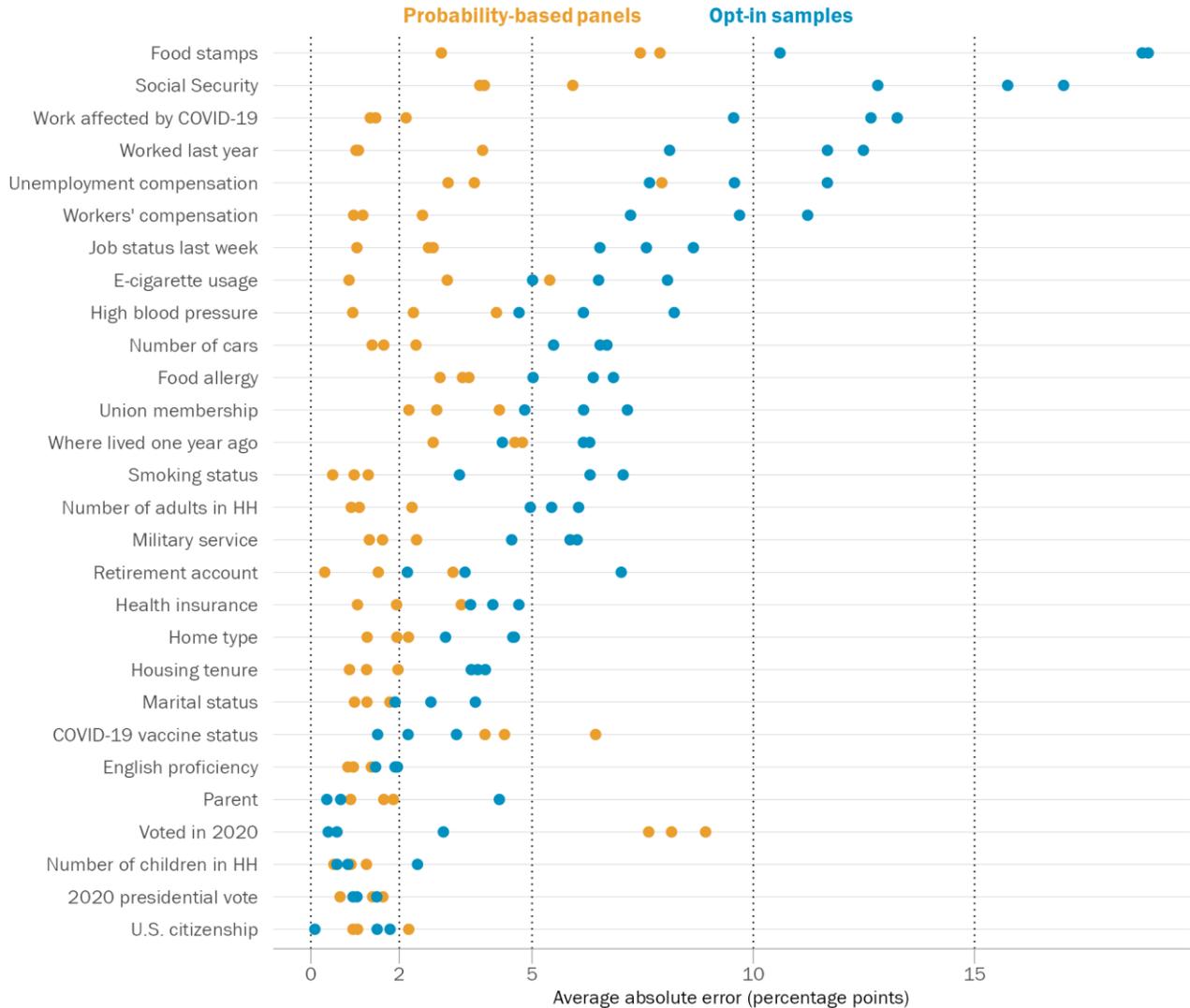
PEW RESEARCH CENTER

Two health-related benchmarks also saw consistently high error on the opt-in samples. Specifically, all of the opt-in samples exceeded the national benchmark for the share of adults with a food allergy (9%) by +6 points on average. They also understated the share of adults who have never tried vaping by -12 points and overstated the share who currently vape some days or every

day by +8 and +5 points on average, respectively. Finally, all three opt-in samples overstated the share of adults who live in carless or single-car households by an average of +6 and +15 points, respectively.

At the other end of the spectrum, there were seven variables with consistently low error on all three probability-based panels. These included parental status, number of children in the household, marital status, housing tenure, smoking status, English language proficiency and candidate vote share in the 2020 presidential election. Two of these, English proficiency and 2020 vote share, also had consistently low error on all three opt-in samples. Citizenship status also had consistently low error on the opt-in samples.

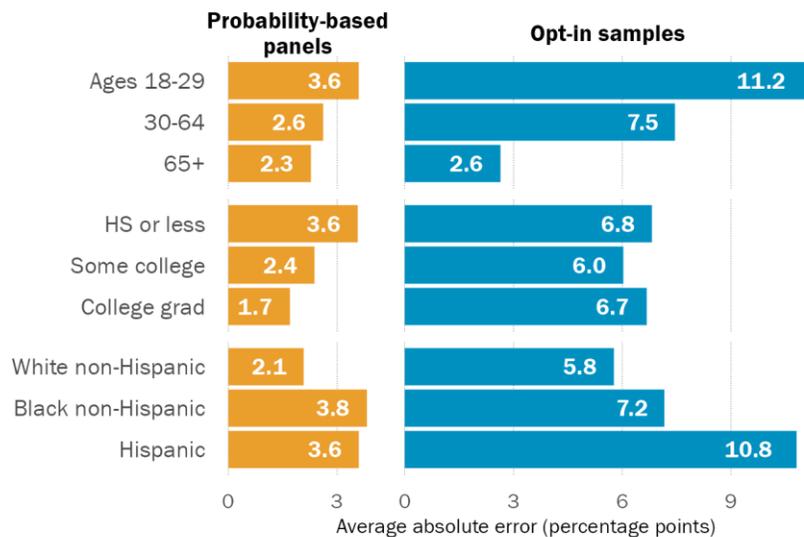# 2. Assessing the accuracy of estimates among demographic subgroups

The accuracy of general population estimates is only one facet of data quality for online samples. Frequently, survey researchers also want to understand the similarities and differences between subgroups within the population. For probability-based panels recruited using ABS, obtaining a sufficiently large sample of respondents belonging to small subgroups can be particularly costly, and one selling point for the use of online opt-in samples is their ability to obtain a large number of interviews with members of hard-to-reach groups at comparatively low cost.

To evaluate the relative accuracy of benchmark subgroup estimates, average absolute error was calculated for each of the samples among subgroups defined by age, education, and race and ethnicity across 25 variables for which subgroup benchmarks were available.[1]

The groups with the largest error across the probability-based panels were adults ages 18 to 29, adults with no more than a high school education, Hispanic adults and non-Hispanic Black adults, which averaged between 3.6 and 3.8 percentage points of error across the three panels. This is about 1 to 2 points less accurate than the other

Large errors for 18- to 29-year-olds and Hispanic adults on opt-in samples

Average absolute error on 25 benchmark variables for estimates among demographic subgroups

	Probability-based panels	Opt-in samples
Ages 18-29	3.6	11.2
30-64	2.6	7.5
65+	2.3	2.6
HS or less	3.6	6.8
Some college	2.4	6.0
College grad	1.7	6.7
White non-Hispanic	2.1	5.8
Black non-Hispanic	3.8	7.2
Hispanic	3.6	10.8

Average absolute error (percentage points)

Note: Results shown are the mean value of the average absolute errors across three probability-based panels and across three online opt-in samples on 25 variables for which reliable subgroup benchmarks are available. Subgroup-level benchmarks are not available for COVID-19 vaccination status, 2020 presidential election turnout and 2020 presidential vote choice and are not included in these figures.
Source: Pew Research Center analysis of six online samples surveyed June 14-July 21, 2021.

PEW RESEARCH CENTER

---

[1] The benchmarks for COVID-19 vaccination status, 2020 presidential vote choice and 2020 presidential election turnout are based on administrative records from the Centers for Disease Control and Prevention and the Federal Election Commission and on figures produced by the United States Elections Project. While these administrative benchmarks are reliable at the national level, they do not allow for the calculation of reliable benchmarks within demographic subgroups and are not used in calculations of average absolute error for subgroups. While the Community Population Survey Voting Supplement can be used to compute subgroup-level estimates for voter turnout, recent scholarship has shown these to be severely biased, substantially overestimating turnout among racial minorities and underestimating turnout among non-Hispanic White voters.

subgroups, which all had average errors between 1.7 and 2.6 points.

By comparison, the opt-in samples had larger error on average for every subgroup, and the differences between the most accurate and least accurate subgroups were much larger. This pattern was most striking for age groups. Here, estimates for adults ages 65 and older on the opt-in samples had an average error of 2.6 points, making them nearly as accurate as the probability-based panels. Their average error was higher for ages 30 to 64, at 7.5 points. Average error was highest for those ages 18 to 29, at 11.2 points – about four times as large as the error for ages 65 and older.

Error for racial and ethnic groups in the opt-in samples showed a similar pattern. Estimates among non-Hispanic White adults were the most accurate, with an overall average absolute error of 5.8 points across the three opt-in samples. Average error among non-Hispanic Black adults was somewhat larger, at 7.2 points, while the average error among Hispanic adults was almost twice as large, at 10.8 points.

Large errors on the opt-in samples were also observed regardless of panelists' level of education. Average errors ranged from 6.0 points for people with some college education to 6.8 points for those with a high school diploma or less.

Opt-in samples had large errors on receipt of government benefits and other variables among 18- to 29-year-olds and Hispanic adults

What factors explain the particularly large errors for 18- to 29-year-olds and Hispanic adults on the opt-in samples? While estimates for both groups are characterized by larger errors on a greater The four benchmarks related to the receipt of government benefits, which had some of the very largest errors for full-sample estimates, had dramatically larger errors for these groups. The opt-in samples overestimated the share of all adults who received food stamps in the past year by an average of +16 percentage points. This overestimation was higher on average among 18- to 29-year-olds (+24) and Hispanic adults (+25). Receipt of Social Security benefits, which had an average error of +15 points for all adults, had errors of +24 for 18- to 29-year-olds and +20 for Hispanic adults. On average, receipt of unemployment compensation was overestimated by +10 points for all adults, +18 points for 18- to 29-year-olds and +21 points for Hispanic adults. Finally, receipt of workers' compensation had an average error of +9 points for all adults, compared with a much higher +23 points for 18- to 29-year-olds and +22 points for Hispanic adults.

For another seven variables, the average absolute error for both 18- to 29-year-olds and Hispanic adults was between 5 and 10 points higher than the error for all adults. These variables included

whether one's work was affected by the COVID-19 pandemic, having a food allergy, union membership, military service, 1-year migration status, parental status and U.S. citizenship.

Similar differences in the magnitude of error were also seen for 18- to 29-year-olds on the benchmarks for high blood pressure, housing tenure and English language proficiency, and for Hispanic adults on e-cigarette usage.

The concentration of disproportionately large errors on so many variables within two specific subgroups raises the question of whether these are primarily errors of representation or measurement. For example, are 18- to 29-year-olds who received food stamps simply overrepresented because some aspect of the data collection process makes them much more likely to participate in online opt-in surveys than other 18- to 29-year-olds? Or are these respondents reporting that they received food stamps when in truth they did not? While this study cannot definitively rule out the possibility these individuals are answering honestly, many respondents to the opt-in surveys answered combinations of questions in ways that are more plausibly explained by individual misreporting than the overrepresentation of certain groups.

On opt-in samples, 18- to 29-year-olds and Hispanic adults show especially large errors on many of the same variables

Average absolute error on opt-in samples for estimates among all U.S. adults, 18- to 29-year-olds and Hispanic adults.
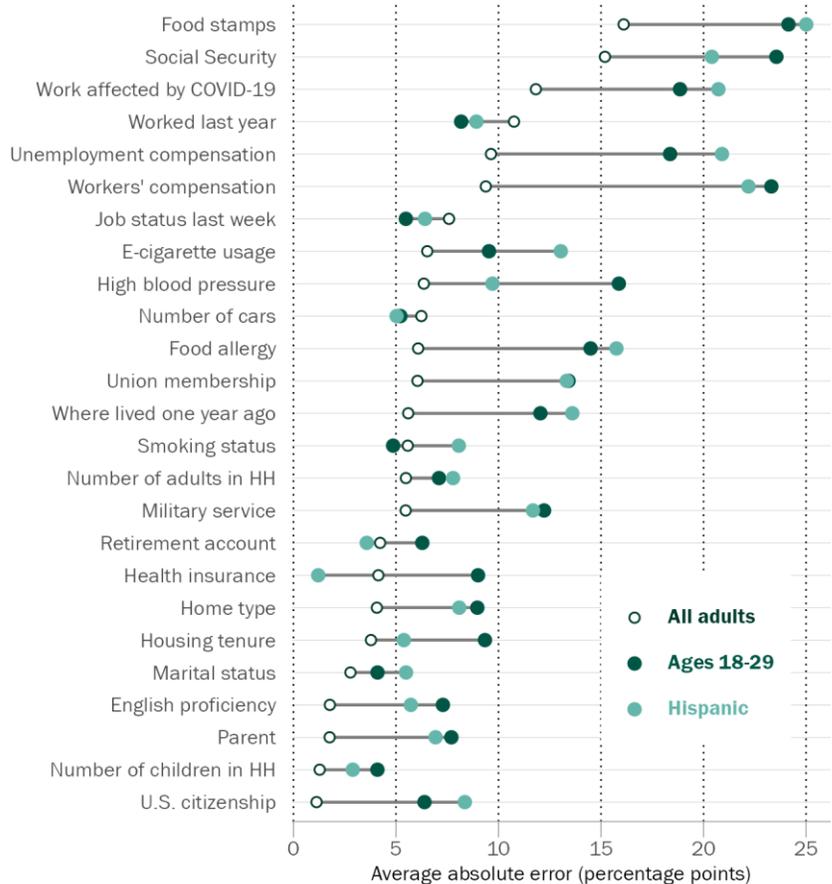
Note: Results shown are the average absolute errors on 25 variables for which reliable subgroup benchmarks are available. Each point represents the mean value across three online opt-in samples. Subgroup-level benchmarks are not available for Covid-19 vaccination status, 2020 presidential election turnout and 2020 presidential vote choice and are not included in these figures.
Source: Pew Research Center analysis of six online samples surveyed June 14-July 21, 2021.

PEW RESEARCH CENTER

The population benchmarks for receipt of government benefits in the previous year (food stamps, Social Security, unemployment compensation and workers' compensation) provide one such example. All of these benchmarks come from the 2021 Current Population Survey Annual Social and Economic Supplement (CPS ASEC), which makes it possible to compute benchmarks for not only the share who received each individual benefit but also for the number of different benefits received. Almost two-thirds of all U.S. adults (62%) did not receive any of these benefits, while 38% received either one or two, according to CPS ASEC data. Adults who received three or four of these benefits comprise only 0.1% of the full U.S. adult population and no more than 0.2% of any demographic subgroup included in this analysis.
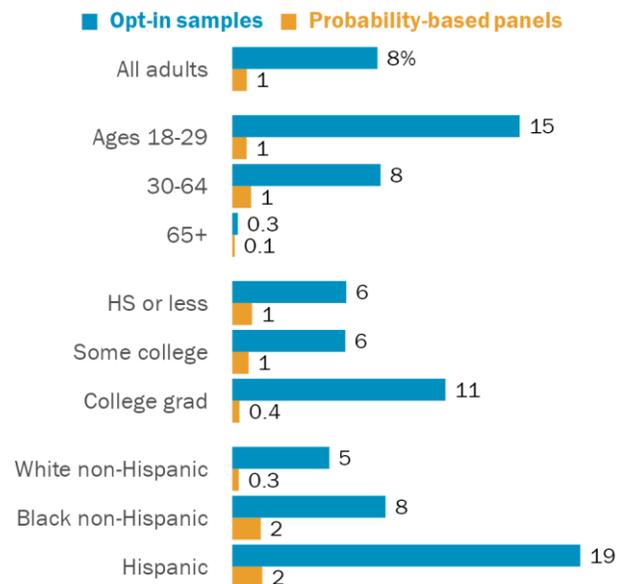
By comparison, the estimated share of adults who received three or four benefits ranged from 6% to 9% on the three opt-in samples. Among 18- to 29-year-olds, estimated shares varied between 15% and 18%; for Hispanic adults, those shares were and between 16% and 19%. On all three probability-based panels, the corresponding estimates were 1% for all adults, between 1% and 2% for 18- to 29-year-olds and between 1% and 3% for Hispanic adults.

It is difficult to see how a group that makes up just a fraction of a percent of the population could come to comprise almost one-in-ten of all respondents, and nearly one-in-five of both 18- to 29-year-old and Hispanic respondents on online opt-in samples. A more straightforward explanation would be a group of respondents who are disproportionately choosing the "Yes" answer rather than answering truthfully. The large errors in estimates among 18- to 29-year-old and Hispanic adults are also consistent with a 2020 Center study that found so-called "bogus respondents" – respondents who make little or no effort to answer survey questions truthfully – disproportionately claimed to be either Hispanic or 18 to 29 years old.

Looking at Yes/No questions more broadly, there remains a consistent pattern. Not counting the question asking respondents if they identify as Hispanic or Latino, there were 12 additional Yes/No questions on the survey that were asked of

Some opt-in respondents tend to answer 'Yes' regardless of the question

Average % of U.S. adults who answered "Yes" to at least 10 of 16 Yes/No questions

■ **Opt-in samples** ■ **Probability-based panels**

	Opt-in	Probability
All adults	8%	1
Ages 18-29	15	1
30-64	8	1
65+	0.3	0.1
HS or less	6	1
Some college	6	1
College grad	11	0.4
White non-Hispanic	5	0.3
Black non-Hispanic	8	2
Hispanic	19	2

Note: Results shown are the mean value across three probability-based panels and across three online opt-in samples. Source: Pew Research Center analysis of six online samples surveyed June 14-July 21, 2021.

PEW RESEARCH CENTER

all respondents, bringing the total number of Yes/No questions to 16. On the probability-based panels, an average of 1% of all adults, 1% of 18- to 29-year-olds and 2% of Hispanic adults answered "Yes" to 10 or more of these questions. On the opt-in samples, the corresponding averages were 8% of all adults, 15% of 18- to 29-year-olds and 19% of Hispanic adults. These results are consistent with the presence of a sizeable group of respondents to the opt-in samples being systematically more likely to answer in the affirmative to Yes/No questions in general.

It is notable that among adults ages 65 and older on the opt-in samples, those saying "Yes" to 10 or more questions comprised only a fraction of a percent on average. This suggests that an absence of bogus respondents within this age group may be a primary reason its accuracy in the opt-in samples was comparable to that of the probability-based panels. One possible reason for this absence may be that on survey measured age by asking respondents to select their year of birth from a drop-down menu with more recent years at the top. Selecting a year of birth corresponding to ages 65 and older would have required more effort than one corresponding to ages 18 to 29, which were much higher up on the list. It is unclear whether a different answer format would have yielded different results.

These findings should not be taken to mean that people who are truly 18 to 29 years old or Hispanic are more likely to misrepresent themselves in online opt-in surveys. It is more likely that individuals who misreport on questions of substantive interest also do so for demographics such as race and age as well. Individuals who are simply attempting to earn survey incentives may be strategically selecting the answer choices they believe are most likely to meet any eligibility criteria for survey participation and least likely to result in being screened out. It is possible that many of the Yes/No questions on the survey resemble the kinds of questions that are commonly used to screen for identify specific subgroups of interest. For example, a bogus respondent seeing a question asking if they have ever vaped may suspect that researchers are conducting a survey of e-cigarette users and that answering "No" would lead to their being screened out. This would be consistent with one recent study that found evidence of widespread falsification intended to get around screening questions in an online opt-in sample. These conclusions are necessarily speculative, as this study was not designed to measure the response strategies of bogus respondents, and this remains an important subject for future research.

However, the fact that such a large portion of the error on the opt-in samples appears attributable to bogus responding that is disproportionately concentrated within specific demographic groups has important implications for practice. The weighting and modeling methods that are most commonly used to adjust for differences between opt-in samples and the larger population are premised on an assumption that the adjustment variables accurately describe the respondents

(e.g. that respondents who say they are Hispanic are, in fact, Hispanic) and that what error exists is small and not strongly correlated with substantive variables of interest.

Here, we have seen that error in adjustment variables like age and Hispanic ethnicity appears to be both widespread in opt-in samples *and* strongly associated with responses to at least Yes/No questions but potentially other kinds of questions where the behavior is not as straightforwardly detectable. While this study did not include trap questions or attention checks, past studies have found such questions to be unsuccessful in identifying bogus respondents. Under such circumstances, there is little reason to expect these kinds of adjustment methods to be successful in the absence of better methods for detecting bogus respondents.

That these kinds of response behaviors appear to be much less common in probability-based panels is heartening and supports a different set of methodological research priorities, particularly correcting the overrepresentation of the most politically and civically engaged respondents.

Acknowledgments

Methodology

The data in this report are drawn from six online surveys of U.S. adults conducted between June 14 and July 21, 2021. Three of the samples were sourced from different probability-based online panels, one of which was Pew Research Center's American Trends Panel (ATP). The remaining three samples came from three different online opt-in sample providers. The study included interviews with a total of 29,937 U.S. adults, approximately 5,000 in each of the samples. Interviews were conducted in both English and Spanish. Because the purpose of this study is methodological, the names of the vendors are masked and the samples are referred to as probability panels 1, 2 and 3 and opt-in samples 1, 2 and 3.

The ATP cases were surveyed using normal procedures. For probability panel 3, the survey was programmed and administered by the vendor who administers that panel. The surveys for the remaining probability-based panel and the three opt-in samples were programmed and administered by a coordinating vendor. The research aims of the study were not discussed with the coordinating vendor and only the questionnaire was provided to the coordinating vendor in advance. On the probability panels, some questions were not asked if a comparable profile variable from a previous survey could be used instead. These questions are identified in the questionnaire.

Source	Field Dates	Sampled	Completes
Probability panel 1	June 14-28, 2021	8,310	5,027
Probability panel 2	June 14-27, 2021	5,722	5,147
Probability panel 3	June 29-July 21, 2021	6,964	4,965
Opt-in sample 1	June 15-25, 2021	n/a	4,912
Opt-in sample 2	June 15-27, 2021	n/a	4,931
Opt-in sample 3	June 15-26, 2021	n/a	4,955
Total			**29,937**

PEW RESEARCH CENTER

Probability panel 1 had a study-specific response rate of 61%. The cumulative response rate to the survey (accounting for nonresponse to recruitment, to the current survey and for panel attrition) was 1.4%.

Probability panel 2 had a study-specific response rate of 90%. The cumulative response rate to the survey (accounting for nonresponse to recruitment, to the current survey and for panel attrition) was 3%.

Probability panel 3 had a study-specific response rate of 71%. The cumulative response rate to the survey (accounting for nonresponse to recruitment, to the current survey and for panel attrition) was 7%.

Sample design

The probability-based panels used in this study all currently recruit using address-based sampling (ABS) in which a random sample of households selected from the U.S. Postal Service's Delivery Sequence File. This Postal Service file has been estimated to cover as much as 98% of the population, although some studies suggest that the coverage could be in the low 90% range.[2]

The ATP cases included in this study are a subset of the respondents to ATP Wave 91. All 11,699 active ATP members were invited to participate in Wave 91, of which 10,606 completed the survey. To achieve an analytic sample of ATP respondents equivalent in size to the other samples included in this study, we drew a stratified random sample of the 11,699 active panelists following the procedure that would have been used to obtain a target sample size of n=5,000 on Wave 91. The panelists who were both selected for this analytic sample and completed Wave 91 are treated as completes for purposes of inclusion in this study. This is the exact set of respondents who would have been observed if only the subsample of the panel had been invited. The remaining probability-based samples were selected following each vendor's normal procedure for achieving a target sample size of n=5,000 U.S. adults.

The three opt-in samples in this study each use a different approach to online opt-in sampling. Opt-in sample 1 comes from a panel aggregator, or marketplace, in which individual respondents are drawn from many opt-in sample sources that have agreed to make their sample available to the aggregator. Opt-in sample 2 is sourced entirely from a single opt-in panel. Opt-in sample 3 is a blend, with about three-fifths sourced from a single opt-in panel and the remainder sourced from three sample aggregators.

All three opt-in samples set quotas for age by gender, race and ethnicity, and education based on estimates from the 2019 American Community Survey.

---

[2] AAPOR Task Force on Address-based Sampling. 2016. "AAPOR Report: Address-based Sampling."

Opt-in sample quotas

Age x Gender	%	Race/Ethnicity	%	Education	%
Men, 18-44	22.8	White/Other/2+ races, non-Hispanic	71.7	High school or less	38.6
Women, 18-44	22.8	Black non-Hispanic	11.8	Some college	30
Men, 45-64	16	Hispanic	16.5	Bachelor's degree or more	31.4
Women, 45-64	17.1				
Men, 65+	9.5				
Women, 65+	11.8				
Total	100		100		100

Note: Quota targets are from the 2019 American Community Survey. No quota was set for the number of respondents who identify as something other than a man or a woman.

PEW RESEARCH CENTER

Data quality checks

No special data quality checks were performed on any of the probability-based panels. For the opt-in samples, the coordinating vendor applied checks for speeding, straightlining and duplicate cases along with other proprietary data quality checks. As a result, a total of 104 cases from opt-in sample 1, 86 from opt-in sample 2, and 69 from opt-in sample 3 were removed for poor data quality.

Weighting dimensions

Variable	Benchmark source
Age x Gender Education x Gender Education x Age Race/Ethnicity x Education Born inside vs. outside the U.S. among Hispanics and Asian Americans Years lived in the U.S.	2019 American Community Survey
Census region x Metro/Non-metro	2020 CPS March Supplement
Volunteerism	2019 CPS Volunteering & Civic Life Supplement
Voter registration	2018 CPS Voting and Registration Supplement
Party affiliation Frequency of internet use Religious affiliation	2020 National Public Opinion Reference Survey (NPORS)

Note: Estimates from the ACS are based on non-institutionalized adults. Voter registration is calculated using procedures from Hur, Achen (2013) and rescaled to include the total U.S. adult population.

PEW RESEARCH CENTER

Weighting

All six samples were weighted following the standard procedure used on ATP Wave 91. For the probability-based panels, this began with a base weight that accounted for differential probabilities of being invited to join the panel, adjustments for panel attrition, and the probability that each panelist was invited to participate in this specific survey. Base weights for each panel were provided by their vendors. Because opt-in samples do not have any known probabilities of selection, all respondents in the opt-in samples were assigned a base weight of 1. The base weight for each sample was then calibrated to align with population benchmarks listed in the accompanying table.

© Pew Research Center 2023

Appendix: Benchmarks and measures of error by sample

2021 American Community Survey

Military service	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Never served in the military	91.9		-2.9	-4.8	-3.7		-12.4	-12.6	-9.5
Only on active duty for training in the Reserves or National Guard	1.2		+0.7	+1.4	+0.7		+5.4	+4.8	+3.2
Now on active duty	0.5		+0.2	+0.9	+0.2		+1.8	+2.6	+1.6
On active duty in the past, but not now	6.4		+1.5	+2.5	+1.9		+4.5	+3.5	+3.9
Average absolute error		*1.8*	*1.3*	*2.4*	*1.6*	*5.5*	*6.0*	*5.9*	*4.5*

Type of dwelling	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
A mobile home	5.0		+0.2	+0.6	+1.4		+4.7	+4.0	+2.4
A one-family house detached from any other house	67.8		-1.5	-4.4	-5.9		-12.0	-10.7	-6.1
A one-family house attached to one or more houses	6.2		+2.1	+3.3	+2.6		+4.2	+4.6	+3.6
A building with two or more apartments	20.9		-2.1	-0.9	+0.2		+0.5	-1.7	-2.2
Boat, RV, van, etc.	0.1		+0.5	+0.6	+0.8		+1.7	+1.8	+1.0
Average absolute error		*1.8*	*1.3*	*1.9*	*2.2*	*4.1*	*4.6*	*4.6*	*3.0*

Housing tenure	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Owned by you or someone in your household with a mortgage or loan (include home equity loans)	43.9		+0.0	+0.9	-3.1		-8.2	-9.0	-5.9
Owned by you or someone in your household free and clear (without a mortgage or loan)	25.3		-1.1	-2.9	-1.4		+0.5	+4.7	+5.6
Rented	29.3		-1.5	+1.1	+2.3		+4.4	-0.3	-1.9
Occupied without payment of rent	1.5		+0.9	+0.2	+1.2		+2.0	+1.8	+1.0
Average absolute error		*1.4*	*0.9*	*1.3*	*2.0*	*3.8*	*3.8*	*3.9*	*3.6*

Note: Benchmark estimates are based on noninstitutionalized adults.

PEW RESEARCH CENTER

2021 American Community Survey (continued)

Number of cars in household	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
No cars	6.1		+1.4	+3.5	+5.2		+7.5	+7.5	+3.8
1 car	23.8		+2.7	+1.9	+2.7		+15.3	+14.8	+15.0
2 cars	39.0		-1.6	-2.0	-2.5		-7.8	-8.2	-5.2
3 cars	19.1		-2.5	-1.7	-3.6		-9.1	-9.5	-7.9
4 cars	8.0		-1.1	-1.2	-0.9		-5.0	-4.4	-4.2
5 cars	2.6		-0.1	-0.7	-1.1		-1.5	-1.3	-1.4
6+ cars	1.4		+0.4	-0.5	-0.6		-0.7	+0.0	-0.8
Average absolute error		*1.8*	*1.4*	*1.6*	*2.4*	*6.2*	*6.7*	*6.5*	*5.5*

English proficiency	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Speaks very well	12.7		+1.8	+3.2	+0.2		+4.4	+2.7	+3.8
Speaks well	4.4		-0.3	+0.0	+0.0		+0.0	+0.1	-1.5
Does not speak well	3.2		+0.0	-1.1	-1.5		-2.2	-1.4	-2.5
Does not speak at all	1.5		-0.9	-1.4	-1.3		-1.2	-1.0	-1.2
Speaks only English at home	78.2		-1.2	-1.1	+1.7		-1.9	-2.1	+0.6
Average absolute error		*1.1*	*0.8*	*1.4*	*1.0*	*1.8*	*2.0*	*1.5*	*1.9*

U.S. citizenship	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
U.S. citizen	92.5		-2.2	-1.0	+1.1		+1.5	+0.1	+1.8
Average absolute error		*1.4*	*2.2*	*1.0*	*1.1*	*1.1*	*1.5*	*0.1*	*1.8*

Number of adults in household	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
1 adult	18.1		+1.4	-3.2	+3.7		+10.0	+8.5	+11.6
2 adults	50.5		+0.4	+0.2	-0.8		-2.4	-0.5	-3.4
3-4 adults	26.8		-0.6	-0.5	-4.3		-7.2	-9.5	-7.7
5+ adults	4.6		-1.2	-0.5	-0.3		-2.2	-1.4	-1.6
Average absolute error		*1.4*	*0.9*	*1.1*	*2.3*	*5.5*	*5.4*	*5.0*	*6.1*

Note: Benchmark estimates are based on noninstitutionalized adults.

PEW RESEARCH CENTER

2021 American Community Survey (continued)

Number of children in household	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
No children	66.9		+3.1	-3.3	+0.2		+0.6	+0.6	+5.5
1 child	14.6		-0.1	-0.6	-0.7		-0.3	-1.8	-3.0
2 children	11.6		-1.7	-0.4	-1.2		-0.9	+0.0	-1.3
3-4 children	6.1		-1.1	+0.3	+0.3		-0.9	-1.6	-2.2
5+ children	0.7		-0.2	+0.0	-0.2		-0.2	-0.2	+0.0
Average absolute error		*0.9*	*1.3*	*0.9*	*0.5*	*1.3*	*0.6*	*0.8*	*2.4*

Note: Benchmark estimates are based on non-institutionalized adults.

PEW RESEARCH CENTER

Centers for Disease Control and Prevention COVID Data Tracker

COVID-19 vaccination status	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
June 29, 2021									
Fully vaccinated	57.2		+8.9	+5.7			-0.1	+2.7	+4.7
Partly vaccinated	9.0		-6.0	-4.6			-2.5	-1.8	-2.8
Not vaccinated	33.8		-4.4	-2.8			+2.0	-2.1	-2.4
July 21, 2021									
Fully vaccinated	59.6				+5.4				
Partly vaccinated	8.8				-6.4				
Not vaccinated	31.6				+0.0				
Average absolute error		*4.9*	*6.4*	*4.4*	*3.9*	*2.3*	*1.5*	*2.2*	*3.3*

Note: Due to its later field date, probability panel 3 was compared to a benchmark dated July 21, 2021. All other samples were compared to a benchmark from June 29, 2021.

PEW RESEARCH CENTER

2021 Current Population Survey Annual Social and Economic Supplement

Worked last year	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Worked at a job or business in 2020	64.2		-3.9	+1.0	-1.1		-11.7	-12.5	-8.1
Average absolute error		*2.0*	*3.9*	*1.0*	*1.1*	*10.8*	*11.7*	*12.5*	*8.1*

Unemployment compensation	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Received state or federal unemployment compensation at any time during 2020	9.3		+3.1	+7.9	+3.7		+9.6	+11.7	+7.7
Average absolute error		*4.9*	*3.1*	*7.9*	*3.7*	*9.6*	*9.6*	*11.7*	*7.7*

Workers' compensation	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Received workers' compensation or other payments as a result of a job-related injury or illness in 2020	0.4		+1.0	+2.5	+1.2		+9.7	+11.2	+7.2
Average absolute error		*1.6*	*1.0*	*2.5*	*1.2*	*9.4*	*9.7*	*11.2*	*7.2*

Food stamps	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Anyone in household received benefits from the Supplemental Nutritional Assistance Program (SNAP) or used a SNAP benefit card in 2020	11.1		+2.9	+7.9	+7.4		+18.9	+18.8	+10.6
Average absolute error		*6.1*	*2.9*	*7.9*	*7.4*	*16.1*	*18.9*	*18.8*	*10.6*

Social Security	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Received Social Security payments in 2020	21.8		+3.8	+5.9	+3.9		+17.0	+15.7	+12.8
Average absolute error		*4.6*	*3.8*	*5.9*	*3.9*	*15.2*	*17.0*	*15.7*	*12.8*

Note: With the exception of union membership, benchmark estimates are weighted using the entropy balanced weights to correct for nonresponse due to the coronavirus pandemic. The benchmark for union membership uses the weight for the outgoing rotation group.

PEW RESEARCH CENTER

2021 CPS ASEC (continued)

Where lived one year ago	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Same house or apartment	92.4		-4.5	-7.1	-7.6		-9.7	-10.5	-7.0
Lived somewhere else in the U.S. or Puerto Rico	7.4		+3.1	+6.0	+5.7		+5.4	+5.3	+3.8
Lived outside the U.S.	0.2		+0.7	+0.8	+1.0		+3.4	+3.0	+2.2
Average absolute error		*4.1*	*2.8*	*4.6*	*4.8*	*5.6*	*6.2*	*6.3*	*4.3*

Marital status	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Married	51.8		+2.0	+1.0	-0.9		-10.7	-5.9	-7.6
Living with a partner	8.1		-0.2	+2.0	+3.5		+2.6	+1.6	-0.9
Divorced	8.4		-1.1	+0.1	+1.9		+2.8	+2.1	+2.0
Separated	1.6		-0.3	+0.5	-0.2		+1.4	+1.1	+0.8
Widowed	5.9		-1.7	-1.1	-0.7		-0.9	-0.6	+0.8
Never been married	24.2		+0.6	-2.9	-3.5		+3.9	+0.2	+4.3
Average absolute error		*1.3*	*1.0*	*1.3*	*1.8*	*2.8*	*3.7*	*1.9*	*2.7*

Union membership	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Member of a labor union or employee association similar to a union	5.6		+2.8	+4.3	+2.2		+6.2	+7.2	+4.8
Average absolute error		*3.1*	*2.8*	*4.3*	*2.2*	*6.1*	*6.2*	*7.2*	*4.8*

Note: With the exception of union membership, benchmark estimates are weighted using the entropy balanced weights to correct for nonresponse due to the coronavirus pandemic. The benchmark for union membership uses the weight for the outgoing rotation group.

PEW RESEARCH CENTER

June/July 2021 Basic Monthly Current Population Survey

Job status last week	Benchmark (%)	Prob. Mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
June 2021									
Worked for pay last week	56.1		-4.0	-1.2			-13.0	-11.9	-9.6
Absent from work last week	2.9		+0.9	+1.7			+2.9	+3.5	+3.0
Not employed last week	40.3		+3.1	-0.2			+10.0	+7.3	+7.0
July 2021									
Worked for pay last week	55.5				-4.1				
Absent from work last week	4.0				+1.5				
Not employed last week	39.9				+2.7				
Average absolute error		*2.2*	*2.7*	*1.0*	*2.8*	*7.6*	*8.6*	*7.6*	*6.5*

Work affected by COVID-19*	Benchmark (%)	Prob. Mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
June 2021									
Employed last week but unable to work during the last four weeks because employer closed due to the coronavirus	1.3		+0.6	+1.5			+7.4	+8.8	+6.0
Employed last week and work was not affected by the coronavirus in the last four weeks	57.7		-3.7	-1.2			-17.9	-17.7	-13.1
July 2021									
Employed last week but unable to work during the last four weeks because employer closed due to the coronavirus	1.1				+0.1				
Employed last week and work was not affected by the coronavirus in the last four weeks	58.3				-2.8				
Average absolute error		*1.7*	*2.2*	*1.3*	*1.5*	*11.8*	*12.7*	*13.3*	*9.6*

* Estimates are based on all adults; however, the category "Not employed last week" is excluded from the average absolute error. This because that category is already included in the benchmark for job status last week.

Note: Due to its later field date, probability panel 3 is benchmarked against the July 2021 Basic Monthly CPS data. All other surveys are benchmarked against the June 2021 Basic Monthly CPS data.

PEW RESEARCH CENTER

2020 Presidential Election

Voter turnout	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Voted (among eligible voters)	66.2		+8.9	+7.6	+8,2		-0.6	-0.4	+3.0
Average absolute error		*8.2*	*8.9*	*7.6*	*8.2*	*1.3*	*0.6*	*0.4*	*3.0*

Presidential vote choice	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Voted Trump	46.9		-5.0	-0.8	-3.7		-3.8	-2.6	-2.6
Voted Biden	51.3		+0.3	-1.6	+0.1		+1.4	+0.1	-0.1
Voted Jorgensen	1.2		+1.6	+0.3	+1.1		+0.9	+1.2	+1.3
Voted Hawkins	0.3		+0.5	+0.2	+0.8		+1.3	+0.7	+0.9
Voted other	0.4		+0.7	+0.4	+1.3		-0.1	+0.2	+0.2
Average absolute error		*1.2*	*1.6*	*0.7*	*1.4*	*1.2*	*1.5*	*1.0*	*1.0*

Note: The benchmark for turnout in the 2020 presidential election comes from the U.S. Elections Project as of May 16, 2021. The benchmarks for presidential vote choice are from the Federal Election Commission.

PEW RESEARCH CENTER

2010 National Health and Nutrition Examination Survey

Food allergy	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Has a food allergy	9.4		+2.9	+3.6	+3.4		+6.8	+6.4	+5.0
Average absolute error		*3.3*	*2.9*	*3.6*	*3.4*	*6.1*	*6.8*	*6.4*	*5.0*

PEW RESEARCH CENTER

2020 National Health Interview Survey

Health insurance	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Covered by health insurance	90.8		-1.1	-3.4	-1.9		-4.1	-4.7	-3.6
Average absolute error		*2.1*	*1.1*	*3.4*	*1.9*	*4.1*	*4.1*	*4.7*	*3.6*

Smoking status	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Now smoke every day	9.2		-1.4	+0.4	+0.3		+10.8	+9.2	+5.2
Smoke some days	3.0		+0.0	+1.7	+0.8		+3.9	+3.7	+2.2
Do not now smoke	22.2		+1.1	+1.3	+0.4		-3.9	-5.4	-1.3
Never smoked 100 cigarettes	63.5		+1.4	-1.9	-0.4		-9.7	-7.0	-4.7
Average absolute error		*0.9*	*1.0*	*1.3*	*0.5*	*5.6*	*7.1*	*6.3*	*3.4*

E-cigarette usage	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Now vape every day	1.7		+0.8	+1.8	+0.8		+4.8	+4.6	+4.4
Vape some days	1.9		+0.5	+3.8	+2.1		+10.5	+8.4	+6.1
Do not now vape	13.4		+0.9	+5.7	+4.0		+1.1	-1.2	-2.6
Never vaped	81.0		-1.2	-10.3	-5.5		-15.8	-11.8	-6.9
Average absolute error		*3.1*	*0.9*	*5.4*	*3.1*	*6.5*	*8.1*	*6.5*	*5.0*

High blood pressure	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Ever diagnosed with high blood pressure	31.1		-0.9	+4.2	+2.3		+8.2	+6.2	+4.7
Average absolute error		*2.5*	*0.9*	*4.2*	*2.3*	*6.4*	*8.2*	*6.2*	*4.7*

Parental status	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Parent of child in household	26.0		-1.6	+1.9	-0.9		-0.4	-0.7	-4.3
Average absolute error		*1.5*	*1.6*	*1.9*	*0.9*	*1.8*	*0.4*	*0.7*	*4.3*

PEW RESEARCH CENTER

2021 Survey of Income and Program Participation

Retirement account	Benchmark (%)	Prob. mean	Prob. 1	Prob. 2	Prob. 3	Opt-in mean	Opt-in 1	Opt-in 2	Opt-in 3
Had a retirement account such as a 401(k), 403(b), IRA or other account designed specifically for retirement savings at any time during 2020	49.9		+3.2	+1.5	-0.3		-7.0	-3.5	-2.2
Average absolute error		*1.7*	*3.2*	*1.5*	*0.3*	*4.2*	*7.0*	*3.5*	*2.2*

Note: The question wording and benchmark value for these estimates was originally taken from the Current Population Survey (CPS) Annual Social and Economic Supplement. Initial analyses showed all samples exceeding the CPS benchmark by a large margin. Investigation of this finding revealed several studies showing that the CPS severely underestimates retirement income and account possession, and that the large errors were most likely due to an unreliable benchmark. Consequently, this report instead uses a comparable benchmark from the Survey of Income and Program Participation determined to be more reliable.

PEW RESEARCH CENTER

QUESTIONNAIRE
2021 BENCHMARKING STUDY

ASK OPT-IN SAMPLES (OPTIN=1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS AND REPLACED WITH PROFILE VARIABLE.* ***
YOBMOD   In what year were you born?

[PROGRAMMING NOTE: 1 DROP DOWN WITH YEAR [2020 - 1920, IN DESCENDING ORDER]

[FORCE ANSWER]

Create Data Only Variable AGE_DOV [Numeric]
AGE_DOV=2021-YOBMOD

[IF AGE_DOV < 18 TERMINATE]

ASK OPT-IN SAMPLES (OPTIN=1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS AND REPLACED WITH PROFILE VARIABLE.* ***
LANG   In which language would you like to complete your survey?
            ¿Prefiere llenar su encuesta en español o en inglés?

    1   English
    2   Español

[PROGRAMMING NOTE:
If LANG=1, show English version of the survey for Opt-in sample
If LANG=2, show Spanish version of the survey for Opt-in sample]

[FORCE ANSWER]

ASK OPT-IN SAMPLES (OPTIN=1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS AND REPLACED WITH PROFILE VARIABLE.* ***
EDUC_ACS   What is the highest degree or level of school that you have COMPLETED?

        1   No schooling completed
        2   Nursery school
        3   Kindergarten
        4   Grade 1 through 11
        5   12th Grade – NO DIPLOMA
        6   Regular high school diploma
        7   GED or alternative credential
        8   Some college credit, but less than 1 year of college credit
        9   1 or more years of college credit, no degree
        10   Associate's degree (for example: AA, AS)
        11   Bachelor's degree (for example: BA, BS)
        12   Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA)
        13   Professional degree beyond a bachelor's degree (for example: MD, DDS, DVM, LLB, JD)
        14   Doctorate degree (for example: PhD, EdD)

[FORCE ANSWER]

ASK OPT-IN SAMPLES (OPTIN=1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS AND REPLACED WITH PROFILE VARIABLE.* ***
HISP   Some questions can be sensitive in nature. We would like to remind you that your participation is strictly voluntary and that your responses are used for research purposes only. You may skip any question that you prefer not to answer.

Are you of Hispanic, Latino, or Spanish origin, such as Mexican, Puerto Rican or Cuban?

> 1   Yes
> 2   No

ASK OPT-IN SAMPLES (OPTIN=1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS AND REPLACED WITH PROFILE VARIABLE.* ***
RACEMOD   Some questions can be sensitive in nature. We would like to remind you that your participation is strictly voluntary and that your responses are used for research purposes only. You may skip any question that you prefer not to answer.

What is your race or origin?

[Check all that apply]

> 1   White
> 2   Black or African American
> 3   Asian or Asian American
> 4   American Indian or Alaska Native
> 5   Native Hawaiian or other Pacific Islander
> 6   Some other race or origin **[TEXT BOX]**

Soft prompt if select 6 and left blank: You did not provide a response in the text box. If you would like to skip, click Next.

ALLOW REFUSED FOR OPT-IN AND TERMINATE IF NON-HISPANIC AND RACEMOD IS REFUSED (HISP=2,REFUSED AND RACEMOD=REFUSED)

ASK OPT-IN SAMPLES (OPTIN=1):
GENDER_SCR   Some questions can be sensitive in nature. We would like to remind you that your participation is strictly voluntary and that your responses are used for research purposes only. You may skip any question that you prefer not to answer.

Do you describe yourself as a man, a woman or in some other way? {W72}

> 1   A man
> 2   A woman
> 3   In some other way

TERMINATE IF GENDER_SCR=REFUSED FOR OPT-IN

ASK ALL:
TRACK   Would you say things in this country today are...

     [RANDOMIZE AND RECORD ORDER OF RESPONSES]

  1 Generally headed in the right direction
  2 Off on the wrong track

ASK ALL:
SOCTRUST Which statement comes closer to your own view, even if neither is exactly right?

  1 Most people can be trusted
  2 You can't be too careful in dealing with people

ASK ALL:
VOL12_CPS In the past 12 months, did you spend any time volunteering for any organization or association? (This includes activities people may not think of, such as infrequent activities or for children's schools)

  1 Yes
  2 No

ASK ALL:
FOLGOV   Would you say you follow what's going on in government and public affairs…

  1 Most of the time
  2 Some of the time
  3 Only now and then
  4 Hardly at all

ASK ALL:
FAVPOL   What is your overall opinion of… **[RANDOMIZE ITEMS]**
       [REVERSE RESPONSE OPTIONS FOR RANDOM HALF SAMPLE]

 a. Joe Biden

   1 Very favorable
   2 Mostly favorable
   3 Mostly unfavorable
   4 Very unfavorable

 b. Kamala Harris

   1 Very favorable
   2 Mostly favorable
   3 Mostly unfavorable
   4 Very unfavorable

 c. Donald Trump

   1 Very favorable
   2 Mostly favorable
   3 Mostly unfavorable

4 Very unfavorable

d. Mike Pence

1 Very favorable
2 Mostly favorable
3 Mostly unfavorable
4 Very unfavorable

e. Anthony Fauci

1 Very favorable
2 Mostly favorable
3 Mostly unfavorable
4 Very unfavorable

ASK ALL:
KNOWSITU   What is one thing that you would like politicians in Washington, D.C. to know about your own situation when they are writing laws and setting policy? Please share as much detail as you can.

[PROGRAMMING NOTE: OPEN END TEXT BOX]

[PROGRAMMING NOTE: ALLOW RESPONDENT TO ENTER AS MUCH TEXT AS THEY WOULD LIKE WITH NO CHARACTER LIMIT]

ASK ALL:
INSURANCE   Are you currently covered by any form of health insurance or health plan?

1   Covered by health insurance
2   Not covered by health insurance

ASK ALL:
COVID_VAXDMOD   Have you received a vaccine to prevent COVID-19?

1   Yes, have had all the shots needed to be fully vaccinated
2   Yes, have had one shot but still need one more
3   No, have not received a vaccine

ASK ALL:
EVSMK_NHIS   Have you smoked at least 100 cigarettes in your <u>entire life</u>?

1   Yes
2   No

ASK IF EVER SMOKED 100 CIGARETTES (EVSMK_NHIS=1):
NOWSMK_NHIS   Do you now smoke cigarettes...

1   Every day

2   Some days
3   Not at all

ASK ALL:
EVVAPE_NHIS   Have you ever used an e-cigarette or other electronic vaping product, even just one time, in your entire life?

1   Yes
2   No

ASK IF USED E-CIGARETTE (EVVAPE_NHIS=1):
NOWVAPE_NHIS   Do you now use e-cigarettes or other electronic vaping products…

1   Every day
2   Some days
3   Not at all

ASK ALL:
BLOODPR   Have you ever been told by a doctor or other health professional that you had hypertension, also called high blood pressure?

1   Yes
2   No

ASK ALL:
FDALL_NHANES   Do you have any food allergies?

1   Yes
2   No

ASK ALL:
LWWORK_CPS   Last week, did you do any work either for pay or profit?

1   Yes
2   No

ASK IF DID NOT WORK LAST WEEK OR REFUSED (LWWORK_CPS=2,99):
LWABSNT_CPS   Last week, did you have a job either full or part time? Include any job from which you were temporarily absent.

1   Yes
2   No

ASK IF WORKED OR HAD JOB LAST WEEK (LWWORK_CPS = 1 OR LWABSNT_CPS = 1):
COVID2_CPS   At any time in the last 4 weeks, were you unable to work because your employer closed or lost business due to the Coronavirus?

1   Yes
2   No

ASK ALL:
JOBLASTYR_CPS   These next few questions are about things that happened *last year in 2020*.

Did you work at a job or business at any time during 2020?

1   Yes
2   No

ASK ALL:
RETACCT_CPS   At any time during 2020 did you have any retirement accounts such as a 401(k), 403(b), IRA, or other account designed specifically for retirement savings?

1   Yes
2   No

ASK ALL:
EMPCOMPMOD At any time during 2020, did you receive any State or Federal unemployment compensation?

1   Yes
2   No

ASK ALL:
WRKRCMPMOD   During 2020 did you receive any Worker's Compensation payments or other payments as a result of a job-related injury or illness?

1   Yes
2   No

ASK ALL:
SOCSEC_CPS   During 2020 did you receive any Social Security payments from the U.S. Government?

1   Yes
2   No

ASK ALL:
FDSTMP2_CPS   At any time during 2020, did *you or anyone in your household* receive benefits from SNAP (the Supplemental Nutritional Assistance Program) or the Food Stamp program, or use a SNAP or food stamp benefit card?

1   Yes
2   No

ASK ALL:
ERNLAB_CPS   Are you a member of a labor union or of an employee association similar to a union?

1   Yes
2   No

ASK ALL:
MIL_ACS   Have you ever served on active duty in the U.S. Armed Forces, Reserves, or National Guard?

1   Never served in the military
2   Only on active duty for training in the Reserves or National Guard
3   Now on active duty
4   On active duty in the past, but not now

ASK ALL:
HOMETYPE_ACS   Which best describes the building where you currently live? (Include all apartments, flats, etc., even if vacant.)

1   A mobile home
2   A one-family house detached from any other house
3   A one-family house attached to one or more houses
4   A building with 2 or more apartments
5   Boat, RV, van, etc.

ASK ALL:
OWNRENTMOD   Which of the following describes the house, apartment or mobile home where you live?

1   Owned by you or someone in your household with a mortgage or loan (include home equity loans)
2   Owned by you or someone in your household free and clear (without a mortgage or loan)
3   Rented
4   Occupied without payment of rent

ASK ALL:
MIG1YR_CPS   Were you living in this house or apartment 1 year ago?

1   Yes, this house
2   No, different house in the United States or Puerto Rico
3   No, outside the United States or Puerto Rico

ASK IF LIVED HERE 1 YEAR AGO (MIG1YR_CPS=1):
MIG5YR_CPS   Were you living in this house or apartment 5 years ago?

1   Yes, this house
2   No, different house in the United States or Puerto Rico
3   No, outside the United States or Puerto Rico

ASK ALL:
CARS_ACS   How many automobiles, vans, and trucks of one-ton capacity or less are kept at home for use by members of your household?

1   None

2	1
3	2
4	3
5	4
6	5
7	6 or more

ASK ALL:
NONENG_ACS   Do you speak a language other than English at home?

    1   Yes
    2   No

ASK IF SPEAKS A LANGUAGE OTHER THAN ENGLISH AT HOME (NONENG_ACS=1):
ENGPROF_ACS How well do you speak English?

    1   Very well
    2   Well
    3   Not well
    4   Not at all

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
CITIZEN   Are you a citizen of the United States?

    1   Yes
    2   No

ASK CITIZENS (CITIZEN = 1):
REG   Which of these statements best describes you?

    1   You are ABSOLUTELY CERTAIN that you are registered to vote at your current address
    2   You are PROBABLY registered, but there is a chance your registration has lapsed
    3   You are NOT registered to vote at your current address

---

ASK CITIZENS (CITIZEN = 1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
VOTED   Which of the following statements best describes you?

    1   I did not vote in the 2020 presidential election
    2   I planned to vote but wasn't able to
    3   I definitely voted in the 2020 presidential election

ASK IF VOTED (VOTED=3):

*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
VOTEGEN_POST   In the 2020 presidential election, who did you vote for?

[RANDOMIZE OPTIONS 1 AND 2 FIRST FOLLOWED BY RANDOMIZED OPTIONS 3 AND 4, WITH OPTION 5 ALWAYS LAST]

    1   Donald Trump, the Republican
    2   Joe Biden, the Democrat
    3   Jo Jorgensen, the Libertarian Party candidate
    4   Howie Hawkins, the Green Party candidate
    5   Another candidate [**only show if respondent tries to skip**]

[PROGRAMMING NOTE: FOR VOTEGENPOST Display choice 5 Voted for another candidate only if the question is skipped without selecting choice 1,2,3 or 4]

Soft Prompt: If you voted for someone else, please select answer choice "Another candidate". If you would like to skip, click Next.

ASK ALL:
WON20   As you may know, Donald Trump contested the results of the 2020 election through legal challenges in various states and the Supreme Court. Those challenges were not successful.

            Who do you think won the 2020 election – that is, who received the most votes cast by eligible voters in enough states to win the election?

[REVERSE ORDER FOR RANDOM HALF OF RESPONDENTS]

    1   Joe Biden definitely won
    2   Joe Biden probably won
    3   Donald Trump probably won
    4   Donald Trump definitely won

ASK ALL:
INTFREQ   About how often do you use the internet?

    1   Almost constantly
    2   Several times a day
    3   About once a day
    4   Several times a week
    5   Less often

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
RELIG   On another topic…

        What is your present religion, if any?

        1   Protestant (for example, Baptist, Methodist, Non-denominational, Lutheran, Presbyterian, Pentecostal, Episcopalian, Reformed, Church of Christ, etc.)
        2   Roman Catholic
        3   Mormon (Church of Jesus Christ of Latter-day Saints or LDS)
        4   Orthodox (such as Greek, Russian, or some other Orthodox church)

5   Jewish
6   Muslim
7   Buddhist
8   Hindu
9   Atheist
10   Agnostic
11   Something else, Specify:_____
12   Nothing in particular

Prompt if response 11 selected and no text entered: You did not specify a response for "Something else". If you are sure you want to skip, click Next.

ASK IF SOMETHING ELSE OR NO RESPONSE TO RELIG (RELIG=11 OR REFUSED):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
CHR   Do you think of yourself as a Christian?

1   Yes
2   No

ASK IF CHRISTIAN (RELIG =1-4 OR CHR=1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
BORN   Would you describe yourself as a born-again or evangelical Christian?

1   Yes, born-again or evangelical Christian
2   No, not born-again or evangelical Christian

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
MARITAL   Which of these best describes you?

1   Married
2   Living with a partner
3   Divorced
4   Separated
5   Widowed
6   Never been married

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
PARTY   In politics today, do you consider yourself a…

1   Republican
2   Democrat
3   Independent
4   Something else

ASK IF INDEP/SOMETHING ELSE (PARTY=3 or 4) OR REFUSED (PARTY=refused):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***

PARTYLN   As of today do you lean more to…

      1   The Republican Party
      2   The Democratic Party

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
IDEO   In general, would you describe your political views as…

[REVERSE RESPONSE OPTIONS FOR RANDOM HALF SAMPLE]

      1   Very conservative
      2   Conservative
      3   Moderate
      4   Liberal
      5   Very liberal

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
NATIVITY   Where were you born?

      1   U.S. – 50 states, District of Columbia
      2   U.S. – Puerto Rico
      3   U.S. – other territory
      4   Another country

ASK IF BORN OUTSIDE OF US (NATIVITY=2,3,4):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
YEARSINUS   How many years have you lived in the United States (excluding Puerto Rico or other U.S. territories)?

[Enter 0 for less than one year]

[PROGRAMMING NOTE: INSERT NUMERIC BOX PROGRAMMED TO ACCEPT NUMBERS FROM 0-97]

Soft prompt if left blank: We would like to have your answer to this question.

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
HH1   How many people, including yourself, live in your household?

[Numeric text box, range 1-8]

[Prompt if HH1>8: If 8 or more people are living in your household, please enter 8 and hit 'Next'.]

ASK IF MORE THAN ONE PERSON IN HOUSEHOLD (HH1>1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
HH3   How many, including yourself, are adults, age 18 and older?

[PROGRAMMING NOTE: NUMERIC TEXT BOX; CANNOT BE GREATER THAN HH1]

ASK ALL:
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
KIDS1   Are you the parent or guardian of any children under age 18?

       1   Yes
       2   No

ASK IF PARENT OR GUARDIAN OF CHILD UNDER AGE 18 (KIDS1=1):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
PARENT   Are any of those children under 18 now living in your household?

       1   Yes
       2   No

ASK IF PROBABILITY PANEL (OPTIN=0):
*** *NOTE: NOT ASKED ON PROBABILITY PANELS IF PROFILE VARIABLE WAS AVAILABLE* ***
GENDER   Do you describe yourself as a man, a woman or in some other way?

       1   A man
       2   A woman
       3   In some other way

ASK OPT-IN SAMPLES (OPTIN-1):
ZIPCODE   What is your zip code?

     [ENTER NUMBER FROM 00000 to 99999]

[SOFT PROMPT IF ENTERS LESS THAN 5 DIGITS: You did not provide a valid response. If you would like to skip click Next.

ASK ALL:
FEEDBACKMOD We strive to ask questions that are politically neutral. Would you say that the
                questions in this survey…

[RANDOMLY DISPLAY CHOICES 1-4 OR 4-1]

       1   Clearly favored one political side
       2   Seemed to favor one political side
       3   Seemed to be politically neutral
       4   Clearly were politically neutral

CLOSING SCREEN
Thank you for participating in this survey.