FOR RELEASE May 2, 2016

Evaluating Online Nonprobability Surveys

Vendor choice matters; widespread errors found for estimates based on blacks and Hispanics

BY *Courtney Kennedy, Andrew Mercer, Scott Keeter, Nick Hatley, Kyley McGeeney,* **AND** *Alejandra Gimenez*

About Pew Research Center

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping America and the world. It does not take policy positions. The center conducts public opinion polling, demographic research, content analysis and other data-driven social science research. It studies U.S. politics and policy; journalism and media; internet, science and technology; religion and public life; Hispanic trends; global attitudes and trends; and U.S. social and demographic trends. All of the center's reports are available at www.pewresearch.org. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

Evaluating Online Nonprobability Surveys

Vendor choice matters; widespread errors found for estimates based on blacks and Hispanics

As the costs and nonresponse rates of traditional, probability-based surveys seem to grow each year, the advantages of online surveys are obvious – they are fast and cheap, and the technology is pervasive. There is, however, one fundamental problem: There is no comprehensive sampling frame for the internet, no way to draw a national sample for which virtually everyone has a chance of being selected.

The absence of such a frame has led to lingering concerns about whether the fraction of the population covered by nonprobability approaches can be made to look representative of the entire population. For roughly 15 years, independent studies suggested that the answer to that question was generally "no" if the goal was to make accurate population estimates.[1] Over time, though, researchers and sample vendors have developed technologies and statistical techniques aimed at improving the representativeness of online nonprobability surveys. Several recent case studies suggest a future (some would argue a present) in which researchers need not have an expensive, probability-based sample to make accurate population estimates.[2]

To better understand the current landscape of commercially available online nonprobability samples, Pew Research Center conducted a study in which an identical 56-item questionnaire was administered to nine samples supplied by eight different vendors.

Nearly all of the questions (52) were also asked on waves of the Center's probability-based American Trends Panel (ATP), which is conducted predominantly online but

Key elements of the study
Design
9 online nonprobability samples
Comparison with an RDD-recruited panel
56 measures including 20 benchmarks
Analysis
Estimated bias on full sample results
Estimated bias on subgroup results
Estimated accuracy of regression models
Demographic profile by sample
Political profile by sample
Variability of estimates across samples

[1] See Reg Baker, Stephen J. Blumberg, J. Michael Brick, Mick P. Couper, Melanie Courtright, J. Michael Dennis, Don Dillman, Martin R. Frankel, Philip Garland, Robert M. Groves, Courtney Kennedy, Jon Krosnick, Paul J. Lavrakas, Sunghee Lee, Michael Link, Linda Piekarski, Kumer Rao, Randall K. Thomas, and Dan Zahs. 2010. "AAPOR Report on Online Panels." *Public Opinion Quarterly* 74(4):711–81; Neil Malhotra and Jon A. Krosnick. 2007. "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." *Political Analysis* 15:286–323; and David S. Yeager, Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." *Public Opinion Quarterly* 75:709–47.
[2] See Wei Wang, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections: Comparing Prediction Markets, Polls, and their Biases." *International Journal of Forecasting*, 31(3): 980–991; Stephen Anolabehere and Brian Schaeffner. 2015. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis*, 22(3): 285-303; and Stephen Ansolabehere and Douglas Rivers. 2013. "Cooperative Survey Research." *Annual Review of Political Science*, Vol. 16, 307-329.

features mail response for adults who do not have internet access. The samples were evaluated using a range of metrics, including estimated bias on 20 full sample survey estimates for which high quality government benchmarks are available, estimated bias for major demographic subgroup estimates, and predictive accuracy of four different regression models. Among the most important findings of this study are the following:

- ***Online nonprobability surveys are not monolithic.*** The study finds, as a starting point, that the methods used to create online nonprobability samples are highly variable. The vendors differ substantially in how they recruit participants, select samples and field surveys. They also differ in whether and how they weight their data. These design differences appear to manifest in the samples' rankings on various data quality metrics. In general, samples with more elaborate sampling and weighting procedures and longer field periods produced more accurate results. That said, our data come from just nine samples, so the effects of these factors are not well isolated, making these particular conclusions preliminary at best.

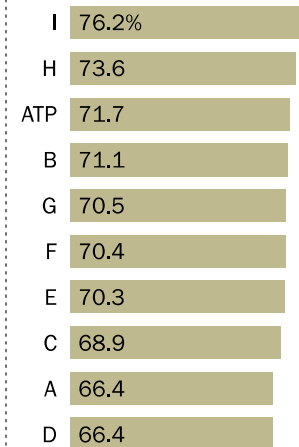## Notable differences in data quality across online samples

Average estimated bias in benchmarking analysis …

Values for each sample represent the average of the absolute differences between the population benchmarks and weighted sample estimates
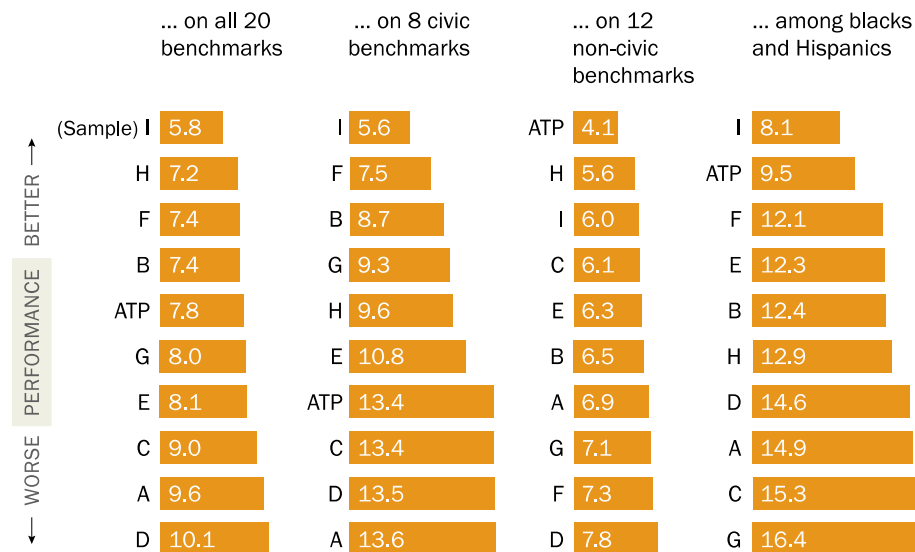
Average % correctly classified in regressions

How well regression models from online samples predict outcomes on benchmark samples (average across four outcomes)

	… on all 20 benchmarks		… on 8 civic benchmarks		… on 12 non-civic benchmarks		… among blacks and Hispanics			
(Sample) I	5.8	I	5.6	ATP	4.1	I	8.1	I	76.2%
H	7.2	F	7.5	H	5.6	ATP	9.5	H	73.6
F	7.4	B	8.7	I	6.0	F	12.1	ATP	71.7
B	7.4	G	9.3	C	6.1	E	12.3	B	71.1
ATP	7.8	H	9.6	E	6.3	B	12.4	G	70.5
G	8.0	E	10.8	B	6.5	H	12.9	F	70.4
E	8.1	ATP	13.4	A	6.9	D	14.6	E	70.3
C	9.0	C	13.4	G	7.1	A	14.9	C	68.9
A	9.6	D	13.5	F	7.3	C	15.3	A	66.4
D	10.1	A	13.6	D	7.8	G	16.4	D	66.4

BETTER ↑ PERFORMANCE WORSE ↓

Note: Black and Hispanic averages exclude the driver's license benchmark which is not available for racial or ethnic subgroups. See Appendix D for details on individual benchmark items.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

- ***Some biases are consistent across online samples, others are not.*** All the samples evaluated include more politically and civically engaged individuals than benchmark sources indicate should be present. The biases on measures of volunteering and community problem-solving were very large, while those on political engagement were more modest. Despite concerns about measurement error on these items, it is accepted that these errors are real because several studies have documented a link between cooperation with surveys and willingness to engage in volunteer activities.[3]

   There is also evidence, though less consistent, that online nonprobability samples tilt more toward certain lifestyles. Most of the samples have disproportionately high shares of adults who do not have children, live alone, collect unemployment benefits and are low-income. In some respects, this squares with a stereotype one might imagine for people who find time to participate in online survey panels, perhaps akin to a part-time job. On other dimensions, however, the online nonprobabilty estimates are either quite accurate (e.g., have a driver's license or length of time at current residence) or the biases are not in a consistent direction across the samples (e.g., daily smoking).

- ***Widespread errors found for estimates based on blacks and Hispanics.*** Online nonprobability survey vendors want to provide samples that are representative of the diversity of the U.S. population, but one important question is whether the panelists who are members of racial and ethnic minority groups are representative of these groups more broadly. This study suggests they are not. Across the nine nonprobability samples, the average estimated bias on benchmarked items was more than 10 percentage points for both Hispanics (15.1) and blacks (11.3). In addition, the online samples rarely yielded accurate estimates of the marginal effects of being Hispanic or black on substantive outcomes, when controlling for other demographics. These results suggest that researchers using online nonprobability samples are at risk of drawing erroneous conclusions about the effects associated with race and ethnicity.

- ***A representative demographic profile does not predict accuracy***. For the most part, a sample's unweighted demographic profile was not a strong predictor of the accuracy of weighted survey estimates. For example, the two samples with the lowest overall accuracy ranked very highly in terms of how well their unweighted demographics aligned with population benchmarks.[4] The implication is that what matters is that the respondents in each

---

[3] See Katherine G. Abraham, Sara Helms and Stanley Presser. 2009. "How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States." *American Journal of Sociology* 114: 1129-1165; and Roger Tourangeau, Robert M. Groves and Cleo D. Redline. 2010. "Sensitive Topics and Reluctant Respondents: Demonstrating a Link between Nonresponse Bias and Measurement Error." *Public Opinion Quarterly* 74: 413-432.

[4] Online nonprobability survey vendors typically apply some form of quota sampling during data collection to achieve pre-specified distributions on age, gender and Census region. However, vendors differ on the details of how this is implemented, which for some involves balancing the sample on variables that go beyond basic demographics.

demographic category are reflective of their counterparts in the target population. It does not do much good to get the marginal distribution of Hispanics correct if the surveyed Hispanics are systematically different from Hispanics in the larger population.

▪ ***One of the online samples consistently performed the best.*** Sample I consistently outperformed the others including the probability-based ATP, ranking first on nearly all of the dimensions considered.[5] This top-performing sample was notable in that it employed a relatively elaborate set of adjustments at both the sample selection and weighting stages. The adjustments involved conditioning on several variables that researchers often study as survey outcomes, such as political ideology, political interest and internet usage. Our impression is that much of sample I's success stems from the fact that it was designed (before and/or during fielding) to align with the population benchmarks on this broader array of dimensions. Unfortunately, we cannot rigorously test that assertion with the data at hand because we have just one survey from that vendor and the relevant design features were not experimentally manipulated within that survey. While the fact that sample I was conditioned on variables that are often treated as survey outcomes raises important questions, it still appears that the sample I vendor has developed an effective methodology. The results from this study suggest that they produce a more representative, more accurate national survey than the competition within the online nonprobability space.

▪ ***Relative to nonprobability samples, results from the ATP are mixed***. Pew Research Center's probability-based panel, the ATP, does not stand out in this study as consistently more accurate than the nonprobability samples, as its overall strong showing across most of the benchmark items is undermined by shortcomings on estimates related to civic engagement. It had the lowest average estimated bias on measures unrelated to civic engagement (4.1 percentage points), but was essentially tied with three other samples as having the largest bias on those types of questions (13.4 points). A likely explanation for this pattern is that the ATP is tilted toward more civically engaged adults as a consequence of being recruited from a 20-minute telephone survey about politics. While the civic engagement bias is concerning, additional analysis indicates that it is not generating large errors on estimates for other domains. When we re-weight the ATP to align with the Current Population Survey (CPS) to eliminate that bias, there is very little impact on other survey estimates, including estimates of voting, party identification, ideology and news consumption.[6]

---

[5] Because the overarching goals of the study were to evaluate the performance of the different samples on a range of metrics and to learn what design characteristics are associated with higher or lower data quality, rather than to single out individual vendors as particularly good or bad, we have anonymized the names of the sample vendors and labeled each with a letter.

[6] This finding is consistent with a highly similar exercise Pew Research Center conducted in a 2012 telephone RDD nonresponse study.

In this study the ATP is not intended to represent all probability samples in any meaningful way, but rather provides one point of comparison. It is an open question as to how a one-off telephone random-digit-dial (RDD) survey or some other probability-based survey would stack up in this analysis.

▪ ***All of the online samples tell a broadly similar story about Americans' political attitudes and recreational interests***. All of the samples indicate that more U.S. adults consider themselves Democrats than Republicans, though as a group they all tilt more Democratic than dual frame telephone RDD surveys. In addition, all of the samples show that Democrats and Republicans are polarized with respect to their attitudes about the proper scope of government. To be sure, there are some notable differences in certain point estimates – e.g., the share of Republicans who say government is doing too many things better left to businesses and individuals is either 64% or 82%, depending on whether one believes sample F or sample I. The broad contours of Americans' political atittudes, however, are arguably similar across the samples. By the same token, results from a battery of 11 personal interest items – ranging from gardening to hip-hop music – show that the top-ranking items tend to be the same from one online sample to the next.

This report focuses on the online nonprobability survey market as it currently exists. But much of the current academic and applied research on this subject is focused on how such samples can be improved through modeling. Aside from relatively simple "raking" adjustments, this study did not examine the potential benefits of more elaborate methods for correcting biases.

To address this, additional research reports on online nonprobability sampling are being planned. One will examine a variety of methods of adjustment to determine how well the accuracy and comparability of estimates across nonprobability samples can be improved. The research underway will test different and more complex approaches to weighting (some of which have been employed by researchers in other organizations) and assess the efficacy of these in reducing bias.

A second study will examine the reliability of repeated measurement over time using online nonprobability samples. The ability to track change over time has been one of the key strengths of probability surveys.[7] The nature of the methods employed by many of the nonprobability samples examined here may or may not produce the levels of reliability that consumers currently rely on from probability samples to detect changes in important attitudes and behaviors.

---

[7] Two waves of a large 2014 Pew Research Center telephone survey administered within a few weeks of each other with 90 identical questions produced a correlation of 0.996 between the measures.

What a 'probability' sample does (and does not) mean for data quality

In this report we make a distinction between samples recruited from a design in which nearly everyone in the population has a known, nonzero chance of being selected ("probability-based") versus samples recruited from advertisements, pop-up solicitations and other approaches in which the chances that a given member of the population is selected are unknown ("nonprobability"). For decades, survey researchers have tended to favor probability samples over nonprobability samples because probability samples, in theory, have very desirable properties such as approximate unbiasedness and quantifiable margins of error that provide a handy measure of precision. For researchers who study trends in attitudes and behaviors over time, the sheer stability of probability-based sampling processes represents an additional crucial property.

While the differences between probability and nonprobability samples may be clear conceptually, the practical reality is more complicated. The root of the complication is nonresponse. If, for example, 90% of the people selected for a probability sample survey decline to respond, the probabilities of selection are still known but the individual probabilities of response are not. In most general population surveys, it is extremely difficult to estimate probabilities of response with a high degree of accuracy. When researchers do not know the probabilities of response, they must rely on weighting to try to correct for any relevant ways in which the sample might be unrepresentative of the population.

Increasingly, researchers are pointing out that when a probability-based survey has a high nonresponse rate, the tools for remediation and the assumptions underpinning the survey estimates are similar if not identical to those used with nonprobability samples. Nonprobability surveys and probability surveys with high nonresponse rates both rely heavily on modeling – whether a raking adjustment, matching procedure, or propensity model – to arrive at what researchers hope are accurate, reliable estimates.

1. Assessing the accuracy of online nonprobability surveys

To better understand the current landscape of commercially available online nonprobability samples, Pew Research Center conducted a study in which an identical questionnaire was administered to nine samples supplied by eight different vendors along with the Center's probability-based online panel. A benchmarking analysis – in which a subset of each survey's results was compared to those from gold-standard government sources – reveals substantial variation across online sample providers in the accuracy of weighted estimates.

The top performing sample was nearly 1.5 percentage points more accurate on average than the second best performing sample (average estimated bias of 5.8 percentage points for sample I versus 7.2 for sample H). The most poorly performing samples yielded estimates that were about 10 percentage points off from the benchmark values on average.

In general, samples with more elaborate sampling and weighting procedures and longer field periods produced more accurate results. The less accurate samples tended to be selected (or "balanced") only with respect to gender, age and region. The best performing samples, by contrast, were balanced not just on those characteristics but also on variables such as education and income. This latter set of samples also tended to be in the field longer, which is likely indicative of the fact that applying more rigorous selection procedures is more time consuming than using less stringent procedures. The limitations of this study's design, however, make these conclusions preliminary at best. Our data come from just nine samples, none of which experimentally manipulated these design features. Consequently, the effects of those features are not well isolated.

In total, 20 benchmark measurements were used in this study (see Appendix D). They touch on a number of different topics including smoking, health care coverage, income, participation in civic or recreational organizations, voting, household composition, internet usage and more. The benchmarks were derived from high quality federal sources based either on national surveys or administrative data. While these are good gauges of accuracy, it is important to keep in mind that

Accuracy of online survey estimates varies substantially across vendors

Average estimated bias of 20 benchmarked weighted survey estimates, in percentage points

(Sample) I	5.8
H	7.2
F	7.4
B	7.4
ATP	7.8
G	8.0
E	8.1
C	9.0
A	9.6
D	10.1
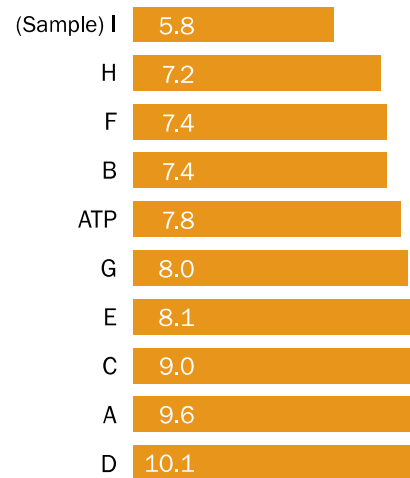
Note: See Appendix D for details on the 20 benchmark items.

Source: Pew Research Center analysis of nine online probability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

measures of political attitudes are frequent targets of surveys but only weakly related with many of these benchmark variables, and thus not necessarily subject to the same biases. For example, the fact that sample G yielded an average estimated bias of 8.0 across the 20 benchmarks does not mean that we would necessarily observe that level of bias in estimates from that sample about, for example, Americans' views on immigration.

The sample showing the lowest average estimated bias overall (I) is also the only nonprobability sample for which the vendor-provided weights performed better than the standardized weighting protocol that we developed to align with the raking used in the Center's probability-based online panel, the American Trends Panel (ATP). The rule employed in this study was to apply whichever weight (vendor-provided or our standardized weight) performed better in terms of minimizing the average estimated bias. This rule sacrifices a clean sample comparison in favor of a "best available package" comparison that allows for the possibility that the vendors might be able to weight their own sample more effectively than we could.

Across the nine nonprobability samples in the study, vendors provided weights for five (B, C, E, F and I). Vendors supplying the other four samples (A, D, G and H) declined to provide weights, signaling that the sample balancing (e.g., quotas) is sufficient for producing a nationally representative survey. The benchmarking results suggest that imposing a few broad quotas is not, in fact, sufficient for at least some of these samples.

While multiple vendors have the ability to sample or weight on a range of variables that go beyond standard demographics, sample I was unusual in this respect. Two of the 20 benchmarks, voter registration and internet usage, were among the variables on which sample I was adjusted. For various reasons, the weighted estimates from that sample hit neither of the benchmark values exactly. The mere fact that the sample was conditioned on these variables, however, calls into question the comparability of sample I's performance in this benchmarking analysis relative to the other samples evaluated. Specifically, it raises the question of whether the other samples would have performed better if they too had been selected and weighted the way that sample I was. In the interest of not putting our thumb on the scale, particularly since the ATP is one of the comparison points, we allowed both of those variables to remain in the benchmarking analysis and for sample I to benefit from its better performing vendor weight.

To understand what effect those decisions have on the benchmarking results, we re-ran the analysis using just the 18 variables which, to the best of our knowledge, were not used in the sampling or weighting of any of the samples. We also re-ran the analysis imposing the standardized weight on sample I, rather than the vendor weight. In each instance, sample I still showed the smallest average estimated bias. This indicates that the superiority of sample I is not

simply a function of the vendor's weighting protocol; it stems also from recruitment and/or sample selection processes.

Tension between conditioning and measurement

While sample I performed the best on the benchmarking analysis, the manner in which that outcome was achieved highlights a critical issue for survey researchers in this era of ever-growing reliance on models to fix sample deficiencies. The design of sample I conditioned on several variables that many social scientists study as survey outcomes – political party, ideology, political interest, voter registration and internet usage. When such variables are "balanced on," "matched on," or otherwise "adjusted for" in the survey, they cease to be random variables estimated by the survey; instead, the survey designer has predetermined what the survey estimates (or at least the possible range of the estimates) for those variables will be. In this case, two of the variables used in the selection and weighting for sample I, voter registration and internet usage, were among the benchmarks outcomes used in the analysis.

Based on our experience commissioning these surveys, the possibility that a sample vendor would predetermine one or more variables that a researcher was intending to study is a real concern. Historically, this has been a relatively minor issue as survey vendors would typically adjust only for demographic variables (e.g., gender, age, race, region) understood by knowledgeable survey consumers to not be the key outcomes estimated in the survey. In recent years, however, there is a trend toward adjusting samples on a greater number and diversity of variables – a trend that is particularly pronounced for some online sample vendors.

Today numerous online survey vendors condition their samples on nondemographic variables in an effort to make them more representative.[8] When implemented carefully and with full consideration of the survey objectives, this practice may help to improve data quality.[9] If, however, the vendor adjusts the sample on attitudes or behaviors without regard for the analytic plan, there appears to be a risk of unintentional influence on study outcomes. Careful coordination between the vendor and the client researchers seems essential to avoid this problem.

---

[8] See Charles A. DiSogra, Curtiss Cobb, Elisa Chan, and J. Michael Dennis. 2011. Calibrating Non-Probability Internet Samples with Probability Samples Using Early Adopter Characteristics. In JSM Proceedings, Survey Methods Section. Alexandria, VA: American Statistical Association, 4501-4515; Mansour, Fahimi, Frances M. Barlas, Randall K. Thomas, Nicole Buttermore. 2015. "Scientific Surveys Based on Incomplete Sampling Frames and High Rates of Nonresponse." *Survey Practice* 8(6); Matthias Schonlau. 2004. "Will Web Surveys Ever Become Part of Mainstream Research?" *Journal of Medical Internet Research* 6(3); and Matthias Schonlau, Arthur van Soest, Arie Kapteyn. 2007. "Are 'Webgraphic' or attitudinal questions useful for adjusting estimates from Web surveys using propensity scoring?" *Survey Research Methods*, 1(3),155-163.

[9] Stephen Ansolabehere and Douglas Rivers. 2013. "Cooperative Survey Research." *Annual Review of Political Science*, Vol. 16, 307-329.

Some biases are quite consistent across online samples, others are not

While the range in the average estimated biases (from a low of 5.8 percentage points to a high of 10.1) demonstrates clear differences across the online nonprobability samples, the direction of the biases reveal some commonalities.

All of the samples include more politically and civically engaged individuals than the benchmark sources indicate should be present. The biases on measures of volunteering and community problem-solving were very large, while those on political engagement were more modest. For example, the nine online nonprobabilty samples overestimated the share of adults who worked with neighborhoods to fix a problem or improve a condition in their community or elsewhere during the past year by an average of 20 percentage points. These same samples overestimated the share of adults who always vote in local elections by an average of 9 points. Despite concerns about measurement error on these items, it is accepted that these errors are real because several studies have documented a link between cooperation with surveys and willingness to engage in volunteer activities.

Online samples tend to display a distinct socioeconomic profile

Weighted % of adults in online samples that belong to each category compared to federal benchmarks
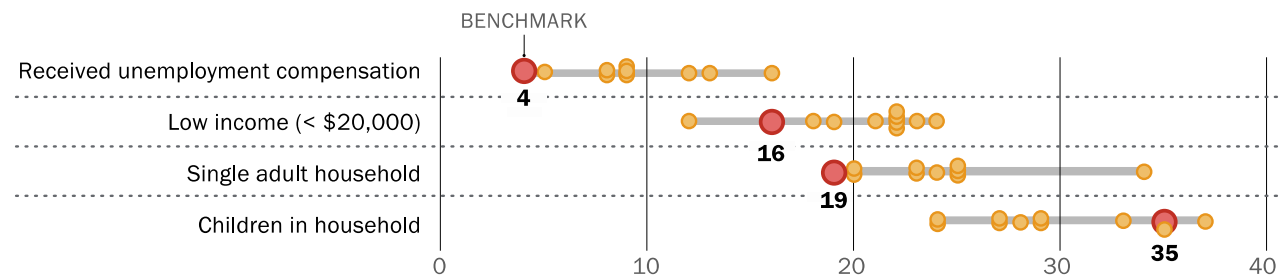
Source: 2015 Current Population Survey Annual Social and Economic Supplement; 2014 American Community Survey; Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

There is also evidence, though less consistent, that online nonprobability samples tend to tilt more toward certain lifestyles. In particular, most of the samples have disproportionately high shares of adults who live alone, collect unemployment benfits, do not have children and are low-income. For example, according to the Department of Labor's Current Population Survey (CPS), 4% of U.S. adults live in a household in which someone received state or federal unemployment compensation during the past year. The average of the weighted estimates from the nine

nonprobability samples, by contrast, was 10% and ranged from a low of 8% (samples H and I) to a high of 16% (sample D). On other topics, however, the online nonprobabilty estimates are either quite accurate or the biases are not in a consistent direction. For example, all of the samples yielded weighted estimates that were reasonably close (within 4 percentage points) to the benchmark incidence of having a driver's license (86%).
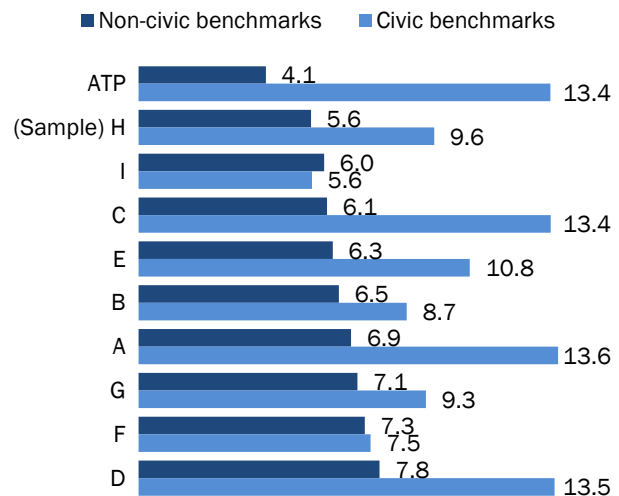
Performance of the American Trends Panel

The American Trends Panel, Pew Research Center's national panel of adults recruited at the end of a large, dual frame RDD survey, is the only probability-based sample in the study. Like the other samples evaluated, the majority of respondents participated online, but the ATP differs in that it also features mail response for adults who do not have internet access. All members of the ATP are asked to complete each of the surveys, which are administered roughly monthly. All of the nonprobability samples, by comparison, select potential respondents for a given survey by subsampling from their panel and, for some, from river sources.[10]

In this study the ATP is not intended to represent all probability samples in any meaningful way, but rather provides one point of comparison. The cumulative response rate for a typical survey on the ATP is 3.5%, reflecting the fact that substantial attrition has taken place even after the recruitment telephone surveys with response rates around 9% are completed. How a one-off dual frame RDD sample or some other probability-based approach would stack up in this analysis is an open question. Future Pew Research Center work will bring data to bear on this issue.

In this analysis, the lone probability-based panel – the ATP – does not stand out as consistently more accurate than the nonprobability samples, as its overall strong showing across most of the benchmark items is undermined by shortcomings on civic-related topics. Overall, the ATP ranked fifth in average estimated bias among the 10 samples evaluated. It had the lowest average bias on measures unrelated to political and civic

ATP shows larger errors for civic and political estimates than others

Average estimated bias of weighted survey estimates

■ Non-civic benchmarks   ■ Civic benchmarks

Sample	Non-civic	Civic
ATP	4.1	13.4
(Sample) H	5.6	9.6
I	6.0	5.6
C	6.1	13.4
E	6.3	10.8
B	6.5	8.7
A	6.9	13.6
G	7.1	9.3
F	7.3	7.5
D	7.8	13.5

Notes: Civic and political items include frequency of talking with neighbors, working with members of your community to solve a problem, membership in a community association, civic association or recreational association, volunteered in the last year, always vote in local elections and registered to vote. See Appendix D for details on individual benchmark items.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

---

[10] "River sample" is a term used when internet users are invited to take a survey through an advertisement or webpage without being required to join a panel. In some cases, answering survey questions allows them to access content that they would otherwise have to pay for, in an arrangement known as a "survey wall."

engagement (4.1 percentage points), but was essentially tied with three other samples as having the largest bias on those types of questions (13.4 points).

A likely explanation for this pattern is that the ATP is biased toward more civically engaged adults as a consequence of being recruited from a 20-minute telephone survey about politics. As Pew Research Center has previously reported, people who engage in volunteer activity are more likely to agree to take part in surveys than those who do not. It is logical that cooperation with a lengthy telephone survey on politics narrowed the potential pool of ATP members to those more inclined toward civic and political engagement. The panel recruitment, in turn, may have been further narrowed to those who viewed their telephone survey experience favorably.

There is some evidence in this study for these compounding factors. Based on our estimate from the CPS, about 69% of all U.S. adults are registered to vote. The registration estimate from the telephone survey used to recruit the ATP was 73%, and the ATP estimate used in this study (from Wave 10) was 76%. Registered voters were more likely than the unregistered to join the panel, and over time the unregistered adults in the panel have been slightly more likely to drop out.[11]

---

[11] In light of the ATP results in this study, we are exploring two possible changes to the panel. In the near term, we may add a civic engagement question to the weighting protocol to mitigate the overrepresentation of more civically minded adults. For a more permanent solution, we are exploring changing the panel recruitment from the the end of a political telephone survey to recruitment through the mail using an address-based sample.

Estimates for Hispanics, blacks, young adults tend to be especially biased

Topline estimates are important, but surveys also try to characterize opinions and behaviors of key population subgroups. This raises the question of whether the average bias levels observed for full sample estimates vary across key subgroups. To gauge this, we computed the benchmarks for major subgroups defined by gender, age, education, race and ethnicity and repeated the analysis for each subgroup. This analysis uses all of the benchmarks except for having a driver's license (no microdataset was available to compute subgroup benchmark values for that characteristic).

Online nonprobability sample estimates based on Hispanics and blacks show particularly large biases. Across the nine nonprobability samples, the average deviation from the benchmarks was 15.1 percentage points for Hispanic estimates and 11.3 percentage points for estimates for blacks. Sample I and the ATP are the only samples examined that have average benchmark deviations in the single digits for both of these subgroups.

Estimated biases were also particularly large for young adults. The pattern of larger average biases for younger adult estimates than older adult estimates (11.8 points for ages 18-29 versus 9.6 points for ages 65 and older) is somewhat surprising given that young adults have much higher levels of internet usage, suggesting that they might be better represented in online panels.

Estimated bias also varied by gender. All of the samples in this study had larger biases when making inferences about men than about women. Across the nine nonprobability samples, the average deviation was 9.9 percentage points for men versus 7.6 points for women.

Estimates for Hispanics and blacks show the largest biases of all major subgroups

Average absolute bias of weighted survey estimates from federal benchmarks by race and Hispanic ethnicity
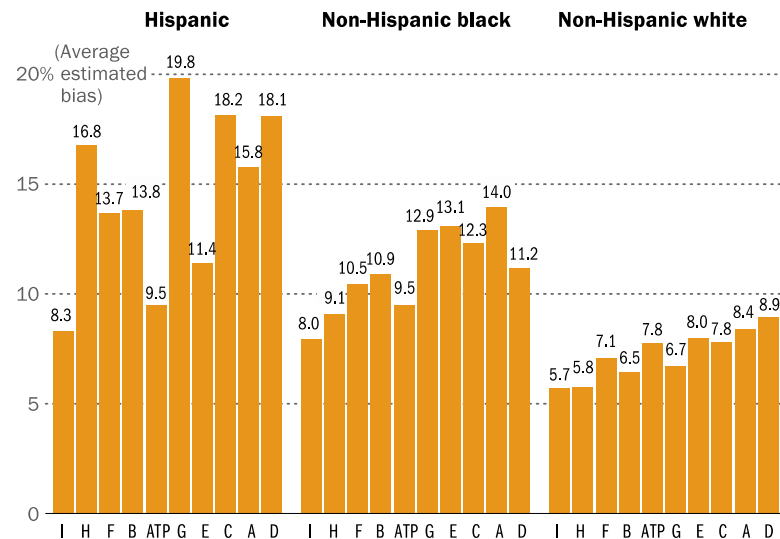
Note: Estimate exclude the driver's license benchmark which is not available for subgroups. See Appendix D for details on individual benchmark items.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

Differences across education categories were not too dramatic, though the average estimated biases tend to be somewhat larger for estimates based on adults with a high school education or less than for estimates based on adults with more formal education.
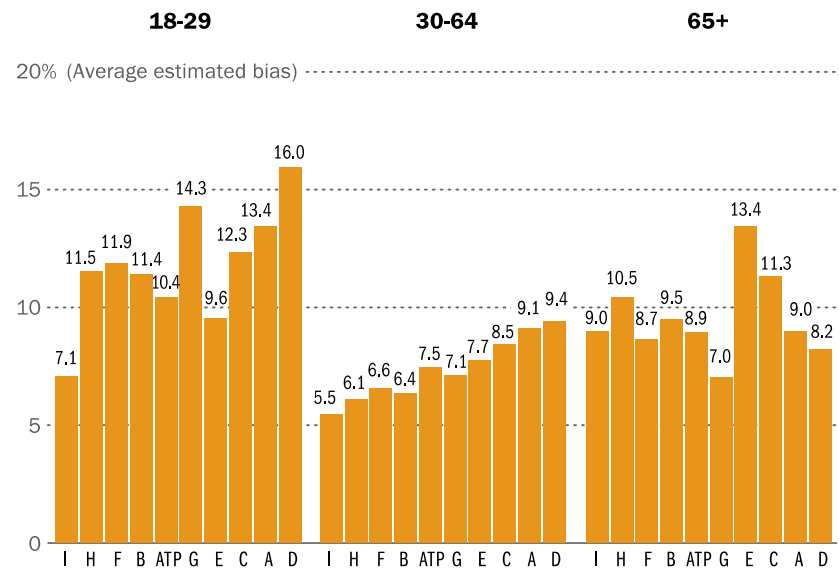
Caveats about benchmarks

Assessing bias in surveys requires an objective standard to which survey findings can be compared. Election polling has such a standard, at least for measures of voting intention: the outcome of the election. Administrative records, such as the number of licensed drivers as used in this report, can provide others. But most such benchmarks are taken from other surveys. Aside from the number of licensed drivers, the benchmarks used here are

Biases tend to be larger for younger adults than for older adults

% Average absolute bias of weighted survey estimates from federal benchmarks by age

Note: Estimate exclude the driver's license benchmark which is not available for subgroups. See Appendix D for details on individual benchmark items.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

drawn from large government surveys that are conducted at considerable expense and with great attention to survey quality. But they are nevertheless *surveys* and are subject to some of the same problems that face surveys like the American Trends Panel and the nonprobability surveys being examined here.

Government surveys tend to have very high response rates compared with probability samples conducted by commercial vendors or nonprofit organizations like Pew Research Center. Accordingly, the risk of nonresponse bias is generally thought to be lower for these government surveys, though it still exists. More relevant is the fact that all surveys, no matter the response rate, are subject to measurement error. Questions asked on government surveys are carefully developed

and tested, but they are not immune to some of the factors that create problems of reliability and validity in all surveys. The context in which a question is asked – the questions that come before it – often affects responses to it. Given that our study selects benchmarks from more than a dozen different government surveys, it is impossible to re-create the exact context in which each of the questions was asked. Similarly, all survey items may be subject to some degree of response bias, most notably "social desirability bias." Especially when an interviewer is present, respondents may sometimes modify their responses to present themselves in a more favorable light (e.g., by overstating their frequency of voting). All of these factors can affect the comparability of seemingly identical measures asked on different surveys.

One other issue: Benchmarks are generally unavailable for questions about attitudes and behaviors that the government does not study. As a result, this analysis uses benchmarks for only a subset of the questions asked on the survey. Moreover, Pew Research Center's work – and the work of other polling organizations conducting political and social research – tends to focus on subjects and questions other than the ones for which benchmarks are available. The generally good record of public polling in presidential elections, including Pew Research Center's surveys, suggests that well-designed surveys using either probability or nonprobability samples can provide accurate measures of political preferences. But election polling's record is hardly unblemished, and candidate choice is but one phenomenon among many we study. Assessing the quality of data is an inexact process at best. It is therefore important to bear in mind that benchmarking provides measures of *estimated* bias and is highly dependent on the particular set of measures included.
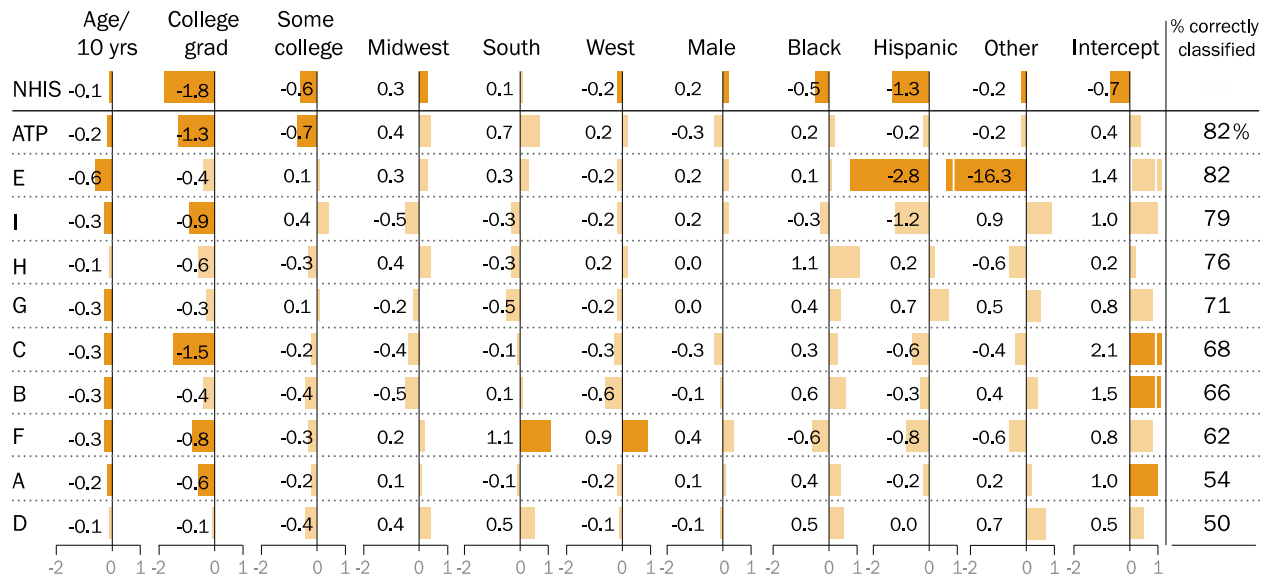
2. Accuracy in estimating multivariate relationships

In addition to point estimates (e.g., % approving of President Barack Obama's job performance), public opinion polls are often used to determine what factors *explain* a given attitude or behavior. For example, is education level or gender more predictive of Obama approval? This type of analysis involves testing the effects of multiple variables simultaneously. One possibility that researchers have discussed is that while some nonprobability samples may not provide very accurate point estimates, they might be able to provide accurate information about how different factors relate to one another (i.e., multivariate relationships).

Substantial variability across samples in predicting smoking

Logistic regressions predicting the probability of smoking daily

How to read this graph: Each row shows the estimated effects of age, education, region, sex and race/ethnicity on the outcome variable for a particular sample, starting with the NHIS benchmark. Negative numbers indicate a lower probability of smoking, and positive numbers a higher one. The length of the bars indicates the size of the effect. Statistically significant effects have darker shading. For example, looking down the first column, all online samples matched the NHIS in finding a negative effect for age, although H and D did not match on significance. The percent correctly classified indicates how often the online sample's model can correctly predict smoking for respondents in the NHIS. The ATP sample is correct 82% of the time, while sample D is correct only half of the time (i.e., as effective as guessing at random).

	Age/10 yrs	College grad	Some college	Midwest	South	West	Male	Black	Hispanic	Other	Intercept	% correctly classified
NHIS	-0.1	-1.8	-0.6	0.3	0.1	-0.2	0.2	-0.5	-1.3	-0.2	-0.7	
ATP	-0.2	-1.3	-0.7	0.4	0.7	0.2	-0.3	0.2	-0.2	-0.2	0.4	82%
E	-0.6	-0.4	0.1	0.3	0.3	-0.2	0.2	0.1	-2.8	-16.3	1.4	82
I	-0.3	-0.9	0.4	-0.5	-0.3	-0.2	0.2	-0.3	-1.2	0.9	1.0	79
H	-0.1	-0.6	-0.3	0.4	-0.3	0.2	0.0	1.1	0.2	-0.6	0.2	76
G	-0.3	-0.3	0.1	-0.2	-0.5	-0.2	0.0	0.4	0.7	0.5	0.8	71
C	-0.3	-1.5	-0.2	-0.4	-0.1	-0.3	-0.3	0.3	-0.6	-0.4	2.1	68
B	-0.3	-0.4	-0.4	-0.5	0.1	-0.6	-0.1	0.6	-0.3	0.4	1.5	66
F	-0.3	-0.8	-0.3	0.2	1.1	0.9	0.4	-0.6	-0.8	-0.6	0.8	62
A	-0.2	-0.6	-0.2	0.1	-0.1	-0.2	0.1	0.4	-0.2	0.2	1.0	54
D	-0.1	-0.1	-0.4	0.4	0.5	-0.1	-0.1	0.5	0.0	0.7	0.5	50

Notes: Statistically significant coefficients are indicated with darker coloring. Percent correctly classified is the percent of respondents in the NHIS sample whose smoking status is correctly predicted by the models fit using online survey data.

Source: 2014 National Health Interview Survey; Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

To test this, we identified four outcomes (smoke daily, volunteered in past 12 months, always vote in local elections, and has health coverage) measured in all of the study samples, as well as in a high quality federal survey. By design, this set of outcomes includes variables that we know from the benchmarking were very difficult to estimate accurately (e.g., volunteering), as well as variables that were generally estimated accurately (e.g., health coverage). We also identified a set of explanatory variables common to these surveys: age, education, gender, race/ethnicity and region. For each sample, we estimated four regression models (one for each of the outcomes) using the same set of explanatory demographics. While these models are simplistic, they are consistent across the study samples.

All samples perform well in predicting health insurance coverage

Logistic regressions predicting the probability of being covered by health insurance

	Age/ 10 yrs	College grad	Some college	Midwest	South	West	Male	Black	Hispanic	Other	Intercept	% correctly classified
ACS	0.3	1.4	0.7	-0.2	-0.6	-0.2	-0.3	-0.4	-1.1	-0.4	0.8	
H	0.3	1.1	0.3	0.7	-0.6	0.0	-0.5	0.6	-0.1	2.5	0.7	89%
D	0.3	1.0	0.8	-1.1	-1.4	-0.4	-0.1	0.0	-0.2	0.4	1.0	88
E	0.2	1.8	0.9	0.6	-0.4	-0.4	-0.1	-0.1	-0.6	0.2	0.7	88
C	0.2	1.6	0.5	-0.2	-0.8	0.0	0.2	0.2	-0.8	0.9		88
I	0.4	0.8	0.2	0.1	-0.4	0.1	-0.3	-0.4	-0.5	-0.9	0.4	88
A	0.4	0.7	0.3	-0.9	-1.4	-0.7	-0.4	0.3	0.1	-0.9	1.2	88
ATP	0.4	1.5	0.6	-0.1	-0.8	0.0	0.0	-0.2	-1.0	0.3	0.3	88
G	0.4	0.8	-0.2	-0.5	-0.8	-0.4	-0.2	-0.2	0.2	0.4	0.2	88
F	0.4	1.9	1.3	-0.5	-1.2	-0.3	-0.2	-1.1	-0.2	-0.3	0.1	87
B	0.4	1.3	0.5	-0.1	-1.0	0.0	0.1	-1.0	-0.4	-0.7	0.0	87

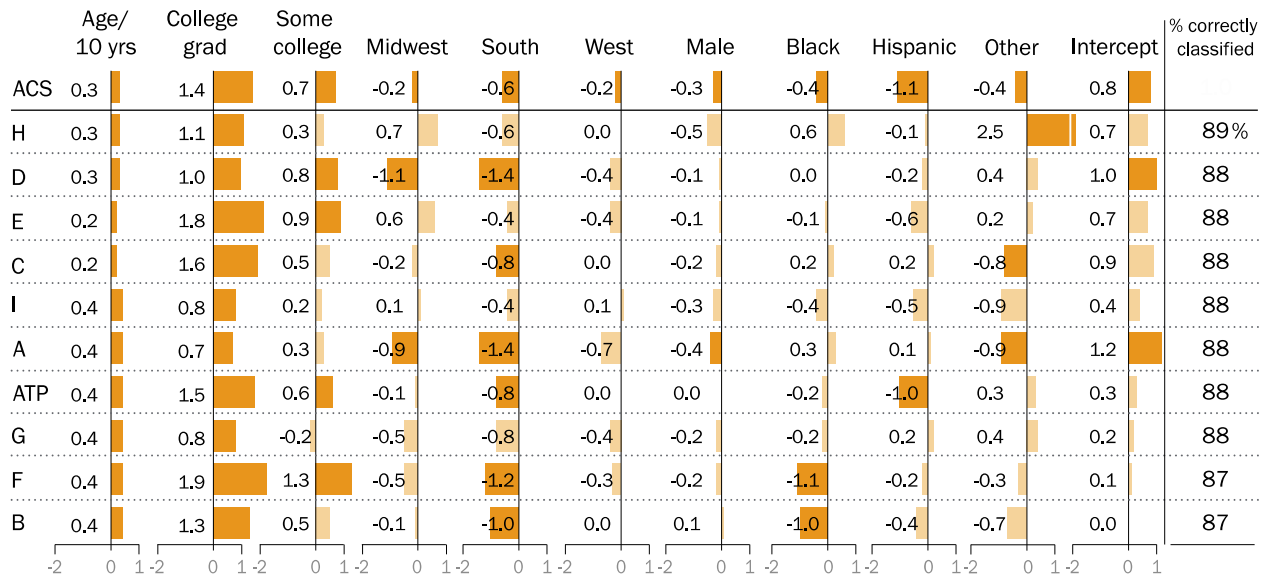Notes: Statistically significant coefficients are indicated with darker coloring. Percent correctly classified is the percent of respondents in the ACS sample whose health care coverage is correctly predicted by the models fit using online survey data.

Source: 2014 American Community Survey; Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
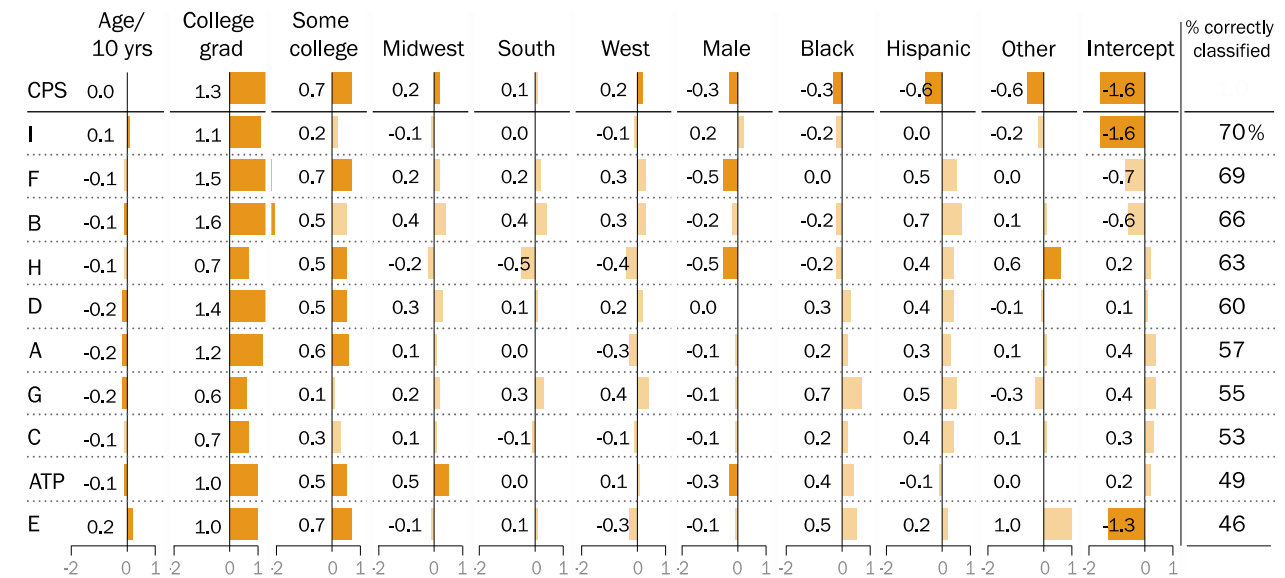"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

With each of these models estimated, the key question was then, *How well do these models, created using online samples, explain the actual behavior of a representative sample of U.S. adults?* We used the microdataset for the federal survey as that representative sample. For each of the 40 models (10 samples, four outcomes), we took the coefficients from the nonprobability

samples and the American Trends Panel (ATP) and applied them to responses in the federal dataset, generating a predicted value for the outcome. We then calculated the rate at which the models that were fit using nonprobability samples were able to correctly predict the true value for each respondent in the benchmark sample (% correctly classified).

Substantial variability across samples in predicting volunteering

Logistic regressions predicting the probability of having volunteered in the last year

	Age/ 10 yrs	College grad	Some college	Midwest	South	West	Male	Black	Hispanic	Other	Intercept	% correctly classified
CPS	0.0	1.3	0.7	0.2	0.1	0.2	-0.3	-0.3	-0.6	-0.6	-1.6	
I	0.1	1.1	0.2	-0.1	0.0	-0.1	0.2	-0.2	0.0	-0.2	-1.6	70%
F	-0.1	1.5	0.7	0.2	0.2	0.3	-0.5	0.0	0.5	0.0	-0.7	69
B	-0.1	1.6	0.5	0.4	0.4	0.3	-0.2	-0.2	0.7	0.1	-0.6	66
H	-0.1	0.7	0.5	-0.2	-0.5	-0.4	-0.5	-0.2	0.4	0.6	0.2	63
D	-0.2	1.4	0.5	0.3	0.1	0.2	0.0	0.3	0.4	-0.1	0.1	60
A	-0.2	1.2	0.6	0.1	0.0	-0.3	-0.1	0.2	0.3	0.1	0.4	57
G	-0.2	0.6	0.1	0.2	0.3	0.4	-0.1	0.7	0.5	-0.3	0.4	55
C	-0.1	0.7	0.3	0.1	-0.1	-0.1	-0.1	0.2	0.4	0.1	0.3	53
ATP	-0.1	1.0	0.5	0.5	0.0	0.1	-0.3	0.4	-0.1	0.0	0.2	49
E	0.2	1.0	0.7	-0.1	0.1	-0.3	-0.1	0.5	0.2	1.0	-1.3	46

Notes: Statistically significant coefficients are indicated with darker coloring. Percent correctly classified is the percent of respondents in the CPS sample whose volunteering is correctly predicted by the by the models fit using online survey data.

Source: 2013 Current Population Survey Volunteer Supplement; Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
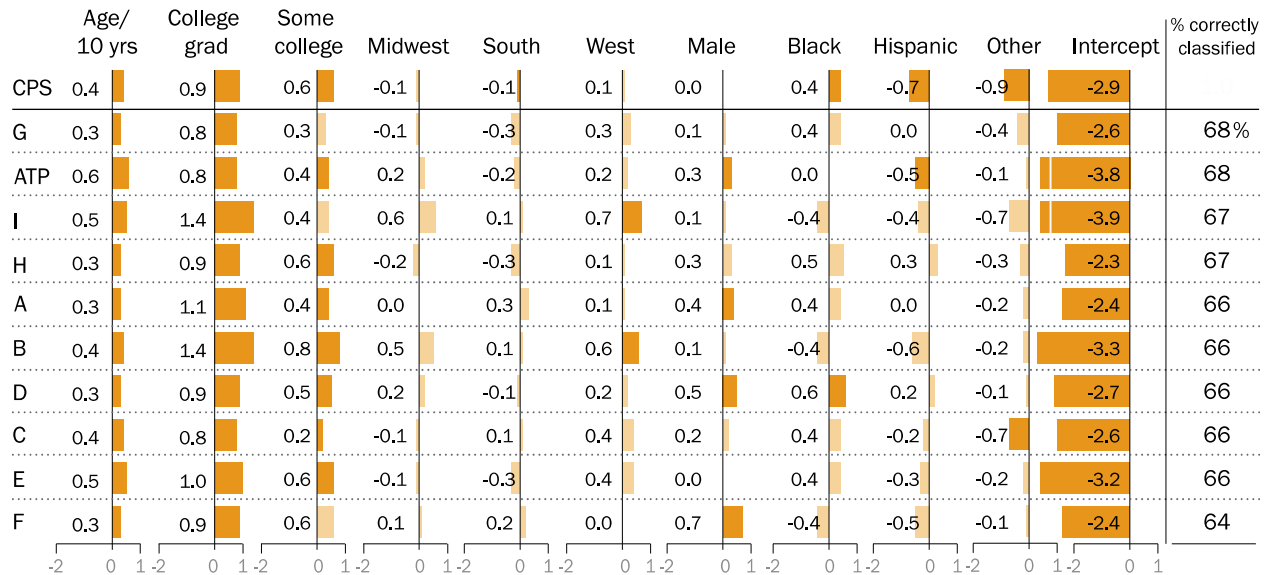"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

Samples that performed relatively well in the benchmarking also performed relatively well in this analysis. Samples I, H and the ATP had the lowest average biases in the benchmarking analysis, and they yielded the regression estimates that were the most likely to correctly classify a randomly sampled, benchmark survey respondent on these four outcomes. As presented at the top of this report, the average share of the federal survey respondents classified correctly across the four outcomes is 76% for sample I, 74% for sample H and 72% for the ATP. The bottom of this ranking is also highly consistent with the benchmarking. The three samples showing the largest average errors in the benchmarking (C, A and D) yielded the regression estimates that are the least likely to correctly classify a randomly sampled adult on these four outcomes. The average percentage classified correctly is 69% for sample C and 66% each for samples A and D.

Looked at individually, the four outcomes vary considerably in the extent to which they differentiate among the samples. For daily smoking, the percentage of federal survey respondents correctly classified was 82% for the best models and 50% for the worst. Similarly, for the volunteering measure, the share of federal survey respondents correctly classified was 70% for the best model and 46% for the worst. For both of these outcomes, the samples that were most accurate on the point estimates also exhibited the most predictive accuracy in the regressions. The reverse was also true: Samples with the least accurate point estimates were also least accurate in the regressions. This demonstrates that the assumption that multivariate relationships will be accurate even when point estimates are not does not hold up.

Samples are similar and limited in their ability to predict voting in local elections

Logistic regressions predicting the probability of always voting in local elections

	Age/ 10 yrs	College grad	Some college	Midwest	South	West	Male	Black	Hispanic	Other	Intercept	% correctly classified
CPS	0.4	0.9	0.6	-0.1	-0.1	0.1	0.0	0.4	-0.7	-0.9	-2.9	
G	0.3	0.8	0.3	-0.1	-0.3	0.3	0.1	0.4	0.0	-0.4	-2.6	68%
ATP	0.6	0.8	0.4	0.2	-0.2	0.2	0.3	0.0	-0.5	-0.1	-3.8	68
I	0.5	1.4	0.4	0.6	0.1	0.7	0.1	-0.4	-0.4	-0.7	-3.9	67
H	0.3	0.9	0.6	-0.2	-0.3	0.1	0.3	0.5	0.3	-0.3	-2.3	67
A	0.3	1.1	0.4	0.0	0.3	0.1	0.4	0.4	0.0	-0.2	-2.4	66
B	0.4	1.4	0.8	0.5	0.1	0.6	0.1	-0.4	-0.6	-0.2	-3.3	66
D	0.3	0.9	0.5	0.2	-0.1	0.2	0.5	0.6	0.2	-0.1	-2.7	66
C	0.4	0.8	0.2	-0.1	0.1	0.4	0.2	0.4	-0.2	-0.7	-2.6	66
E	0.5	1.0	0.6	-0.1	-0.3	0.4	0.0	0.4	-0.3	-0.2	-3.2	66
F	0.3	0.9	0.6	0.1	0.2	0.0	0.7	-0.4	-0.5	-0.1	-2.4	64

Notes: Statistically significant coefficients are indicated with darker coloring. Percent correctly classified is the percent of respondents in the CPS sample whose frequency of voting in local elections is correctly predicted by the by the models fit using online survey data.

Source: 2013 Current Population Survey Civic Engagement Supplement; Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

For health coverage, all of the samples were relatively close to the benchmark, and all display similarly high levels of predictive accuracy, with the best and worst models differing by only 2 percentage points (89% vs. 87%). The voting models also yielded a narrow range in the percent correctly classified (64% to 68%), but the overall level of accuracy was lower.

Marginal effects associated with race and ethnicity are rarely correct

Substantively, the conclusions one would draw from the coefficients of these regression models using the nonprobability samples would likely differ from the conclusions drawn using the benchmark survey. While the nonprobability samples often succeed in capturing the effects associated with education and age, they rarely capture the effects associated with race and ethnicity that one finds in the benchmark surveys.
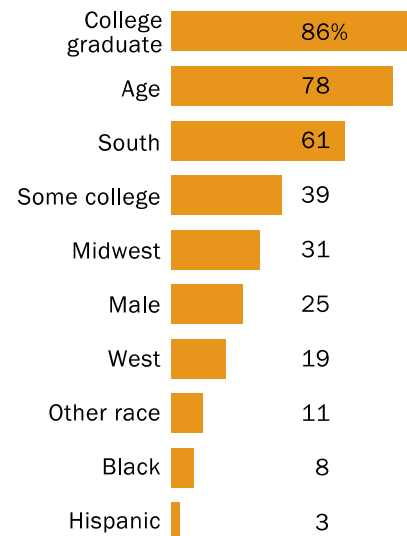
For example, according to the NHIS model estimates, daily smoking is negatively associated with age, education, Hispanic ethnicity and being black. For the most part, the nonprobability sample models show significant, negative effects for age and education. Only one of the nine nonprobability samples, however, has significant negative effects for both Hispanic ethnicity and black race.

This general pattern is seen in models for all four outcomes. We can quantify the pattern by leveraging the fact that, for each category, we have 36 nonprobability sample estimates (four outcomes modeled separately with nine samples). We noted whether each of those 36 estimated effects was consistent with the same effect estimated from the benchmark dataset with respect to direction and significance. If the estimated effect in the benchmark survey was nonsignificant, then the nonprobability effect estimate was coded as "correct" if and only if it was also nonsignificant, regardless of direction. If, by contrast, the estimated effect in the benchmark survey was statistically significant, then the nonprobability effect was coded as "correct" if and only if it was also statistically significant and in the same direction.

By this measure, the nonprobability samples "correctly" estimated the effect of being a college graduate 86% of the time (31 of 36 estimates) and "correctly" estimated the effect of age 78% of the time (28 of 36 estimates). These successes are no doubt related to the fact that age and education truly have strong associations with the four outcomes used in this analysis.

Online nonprobability samples struggle with marginal effects of race and ethnicity

% of coefficients estimated with nonprobability samples that match the significance and sign of the coefficient estimated with the benchmark survey

College graduate	86%
Age	78
South	61
Some college	39
Midwest	31
Male	25
West	19
Other race	11
Black	8
Hispanic	3

Note: Each value is based on 36 estimated coefficients from regression models on four outcome variables run separately on nine online nonprobability samples.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

The nonprobability samples did not fare nearly as well in estimating the marginal effects from race and ethnicity. Based on the benchmark survey data, each of the four models show statistically significant effects for both Hispanic ethnicity and black race. Across the 36 nonprobability sample estimates, the Hispanic effect was "correct" only once (sample E on daily smoking). The effect associated with being black was "correct" only 8% of the time (once each for samples B, D and F). These results indicate that researchers using online nonprobability samples are at risk of drawing erroneous conclusions about the effects associated with race and ethnicity.

In other analyses presented in this report, focusing on the collective performance of the nonprobability samples tells, at best, only part of the discernable story about data quality. This particular analysis is different, however, because all of the nonprobability samples tend to estimate the marginal effects from age and education reasonably well and estimate the marginal effects from race and ethnicity poorly. Notably, sample I, which stands out as a top performer on several other metrics, looks unremarkable here.

By this standard, the results for the American Trends Panel are fairly similar to those from the nonprobability samples, but the ATP was more successful in capturing the marginal effect associated with Hispanic ethnicity. The ATP's estimated Hispanicity effect was accurate for both health coverage and voting in local elections. This result is likely related to the fact that the ATP features Spanish-language as well as English administration, whereas the nonprobability samples were English only. The ATP also has a sample size advantage relative to the nonprobability samples. Each of the nonprobability samples features about n=1,000 interviews, whereas the average ATP sample size was roughly n=2,800, making it easier to detect a statistically significant difference. To test whether the ATP results are explained by the larger sample size, we replicated these regressions on 15 random subsamples of 1,000 ATP respondents and found that the conclusions were not substantively affected.

3. Demographic, political and interest profiles

Many nonprobability sample vendors have the ability to provide samples of respondents that, by design, are forced to align with characteristics of the U.S. population. Often those characteristics are demographics such as gender and age, though some vendors also use nondemographic variables. When a vendor forces the sample to match the population on a particular characteristic (e.g., % female), the survey estimate of that characteristic is no longer informative about the quality of the sample because it was predetermined. This is analogous to the situation in probability-based surveys where weighting a variable to match a particular distribution through raking means that the variable could no longer be considered an informative outcome variable.

Most samples substantially underrepresent less-educated adults, on an unweighted basis

While the capacity to predetermine the demographic profile of the sample is common in nonprobability web surveys, vendors vary dramatically on what variables they use to do so. Consequently, the unweighted demographic profiles of the 10 samples show large differences.

One striking pattern is that all of the unweighted samples, with the notable exception of sample I, substantially underrepresent adults with less formal education.[12] According to the Census Bureau's American Community Survey, 40% of U.S. adults have a high school education or less. Among the 10 samples examined here, the average unweighted incidence of adults in this education group was about half that (21%). Another common, though less consistent, pattern in these samples is overrepresentation of non-Hispanic whites and adults ages 65 and older. The directions of these unweighted demographic biases are quite common in U.S. surveys across a range of designs, not just online nonprobability samples.

---

[12] In personal communications with vendors, we gathered that a number of them (not just the sample I vendor) have the ability to select on education at the sampling stage, but that is done upon request, not by default.

Most samples overrepresent whites and college graduates, on unweighted basis

% of demographic characteristics in the ACS and percentage point differences by survey. The estimates for nonprobability samples reflect their use of quotas to match population benchmarks.

	ACS	ATP	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample I
	%										
HS or less	40	-21	-14	-22	-21	-18	-31	-19	-17	-24	-1
Some college	32	-3	10	2	5	7	1	6	5	4	1
College grad+	28	25	3	19	16	11	30	12	11	19	-1
White, non-Hispanic	66	12	4	16	3	-1	2	12	17	16	5
Black, non-Hispanic	11	-3	1	-5	-4	1	-2	-6	-5	-5	0
Hispanic	14	-6	-3	-8	-6	1	1	-6	-11	-9	-2
Other	8	-2	-2	-3	6	0	0	0	-1	-2	-4
Age 18-29	21	-8	2	-7	7	4	-2	-12	-11	-7	0
Age 30-49	34	-7	2	-3	6	3	-5	-8	-8	-6	1
Age 50-64	27	6	0	-1	-4	-4	5	5	3	4	0
Age 65+	17	10	-4	12	-8	-3	1	15	16	10	-1
Male	48	1	-1	7	-5	0	-2	-12	-14	1	5
Female	52	-1	0	-7	5	0	2	11	14	-1	-5
Sample size		3,147	1,022	1,049	1,178	1,005	1,022	1,008	1,010	1,007	1,000

Source: 2014 American Community Survey; Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

A good-looking sample might not translate into better survey estimates

The unweighted demographic profiles reveal a curious pattern. For the most part, the demographic representativeness of a sample – on gender, age, race, ethnicity and education – is not a strong predictor of how well that sample performed in the benchmarking or in the regression analysis. Samples D and A rank worst in the benchmarking and regression analyses but rank second and third, respectively, in average deviation from population benchmarks on the five demographics.

That said, sample I ranked first in all three: benchmarking, regression and unweighted sample representativeness. The implication is that what matters is that the respondents in each demographic category are reflective of their counterparts in the target population. It does not do much good to get the marginal distribution of Hispanics correct if the surveyed Hispanics are systematically different from Hispanics in the larger population.

Unweighted demographics are largely unrelated to accuracy

Average estimated bias on benchmarks vs. average estimated bias for unweighted demographics
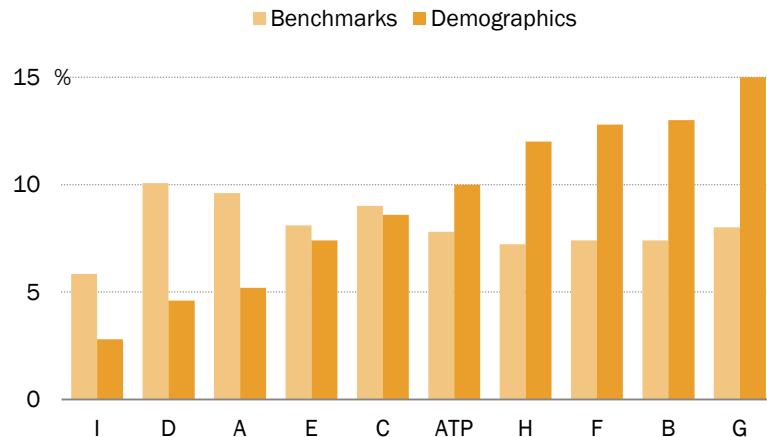
Notes: Demographic values represent the average absolute deviation (in percentage points) between the population benchmark and the unweighted survey estimate for one category of each of the four demographic variables. See Appendix D for details on individual benchmark items.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

Measures of political attitudes and engagement

Much of Pew Research Center's work focuses on politically relevant attitudes and behavior, including civic engagement. Several questions in the current study focused on these topics. Three items asked about political engagement: voter registration, voter turnout in local elections and contacting an elected official. Five items asked about civic engagement, including participation in community, civic or recreational groups or associations; volunteering; and working with others to solve a community problem. Measures of political attitudes included party affiliation, ideological identification and opinion about the scope of the federal government. All five civic engagement items have comparable government benchmarks, as do the two political engagement items. No benchmarks are available for the measures of political attitudes.

All samples overestimate volunteering, but some do so more than others

Weighted % of adults who volunteered in the last year
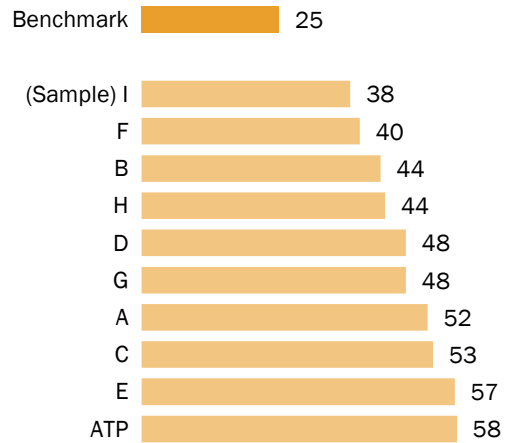
Benchmark	25
(Sample) I	38
F	40
B	44
H	44
D	48
G	48
A	52
C	53
E	57
ATP	58

Source: 2013 Current Population Survey Volunteer Supplement; Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

A well-known bias in political surveys based on probability samples is that they overrepresent the politically engaged. This bias stems from at least three sources. First, the topic of political surveys is considered more salient to politically engaged individuals, leading them to be more likely to participate in the interview.[13] Social desirability bias may also introduce measurement error by leading respondents to say that they are more politically engaged than they are. And, more generally, surveys tend to underrepresent the young and the less-educated, groups that are less interested and engaged in politics than the average. Weighting may not fully correct this bias. All three of these factors may be present in nonprobability samples as well. All of the samples in this study appear to include more politically and civically engaged individuals than the benchmarks indicate should be present.

---

[13] Robert M. Groves, Stanley Presser and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68: 2-31; and Roger Tourangeau, Robert M. Groves and Cleo D. Redline. 2010. "Sensitive Topics and Reluctant Respondents: Demonstrating a Link between Nonresponse Bias and Measurement Error." *Public Opinion Quarterly* 74: 413-432.

Among all benchmark items, most of the civic and political engagement measures have above-average bias, with the civic engagement items showing bigger biases than the political items.[14] Participation in volunteer activity for or through a group in the past 12 months is the item with the largest bias of all 20 benchmark measures compared in this study, averaging 23.1 percentage points and ranging from 13 to 33 points. Even greater bias in relative – though not absolute – terms is seen in a question about working with others to solve a community or neighborhood problem; the mean overstatement was 20.4 points (relative to a benchmark of 7.7%) and ranged from 13 to 26 points.

Participation in each of three types of associations also reflect sizable biases in relative terms, with the share who say they have taken part in activities for a service or civic association averaging nearly double the benchmark (a mean reading of 13% vs. a benchmark of 6%). One of the nonprobability samples (sample I) actually matched the benchmark on participation in a recreational or sports organization and came within 1 point on participation in school group, neighborhood or community association.

Measures of political engagement were subject to similar, though smaller, biases. Voter registration is overstated in eight of the nine nonprobability samples and the American Trends Panel, with the surveys yielding estimates higher than the benchmark of 69%; the average absolute bias is 5.7 percentage points. Sample I produced an estimate of 62%, but weighted the data to match that figure. On a measure of regularity of voting in local elections, all of the samples produced an estimate higher than the benchmark (32% say they "always" vote), with an average bias of 8.1 percentage points.

Online samples tended to overestimate voter registration

Weighted % of adults who are registered to vote

Benchmark	69
(Sample) I	62
G	71
C	71
A	73
F	74
E	76
ATP	76
D	76
H	77
B	77
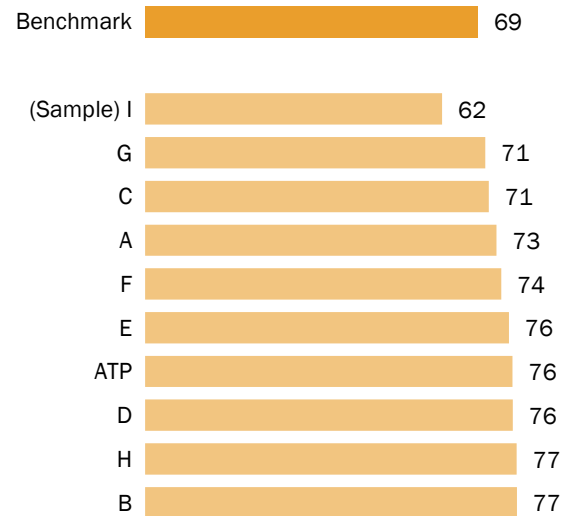
Source: 2014 Current Population Survey Voting and Registration Supplement (adjusted); Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

---

[14] Katherine G. Abraham, Sara Helms and Stanley Presser. 2009. "How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States." *American Journal of Sociology* 114: 1129-1165.

Online samples yield roughly similar results in describing the countors of U.S. political attitudes

No benchmark exists for the three political attitude questions: party affiliation, self-identified ideology and opinion about the appropriate scope of government. Most of the samples produced relatively similar estimates of the Democratic and Republican shares of the public, with Democrats outnumbering Republicans in all of the samples. All of the online samples yielded higher estimated shares of adults identifying as Democrats than was found in an analysis of Pew Research Center RDD telephone surveys conducted during 2015 and 2016. Most of the online samples also yielded fewer independents (respondents who declined to affiliate with one of the two major parties) than were found in the analysis of recent telephone surveys. This could be a mode effect, though there was no statistically significant mode effect on party affiliation in the Center's randomized mode experiment conducted in 2014 with many of these same respondents.

Estimates of the liberal-conservative divide were similar across the online samples, with conservatives outnumbering liberals in each sample. All of the nonprobability samples produced responses that were more politically liberal than conservative on a question that asked respondents about their preference regarding the scope of government.

Online samples tilt more Democratic than telephone RDD samples

Weighted % of adults who consider themselves to be a …
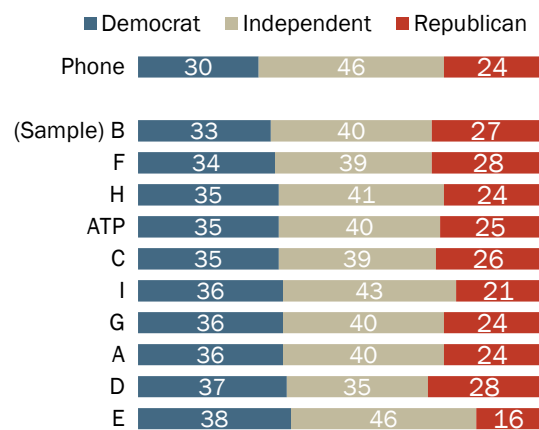
Note: The phone estimate is the average over all Pew Research Center political surveys for 2015-2016.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
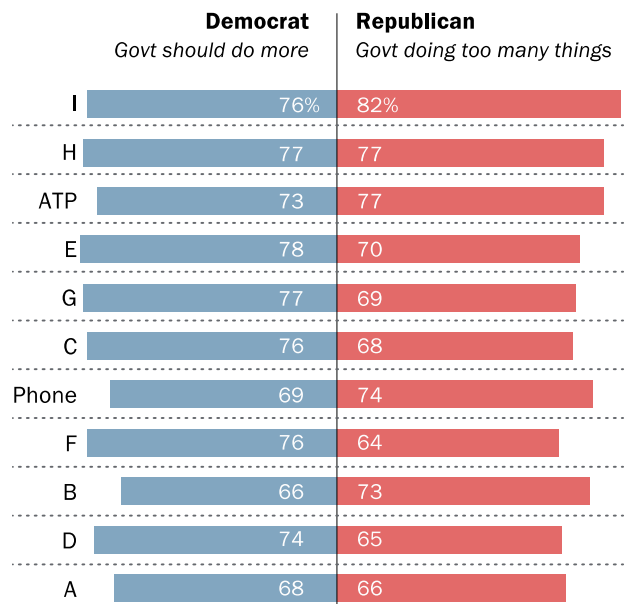"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

The American Trends Panel found 48% favoring a government that does more to solve problems (rather than believing that government does too many things better left to businesses and individuals), while all of the nonprobability samples found this share to be higher than 50%, with an average of 54%.

A key question in contemporary political polling concerns how polarized certain opinions are by party affiliation. Although there is no benchmark available for comparison, we can examine how Democrats and Republicans differ on attitudes measured in the survey. We expect Republicans to identify as conservative and to believe that government is doing too many things, while Democrats are more likely to self-identify as liberal and to prefer a government that does more things.

All of the samples display the expected discrimination between Democrats and Republicans on these two measures, but the extent of the division varies considerably. Sample I shows the greatest discrimination, with 82% of Republicans saying the government is doing too many things, while 76% of Democrats say the government should do more. Sample A had the smallest ideological gap on this question, with 68% of Democrats taking the liberal position and 66% of Republicans taking the conservative one. A large [Pew Research Center telephone survey](#) conducted in late 2015 found 74% of Republicans saying that government is doing too many things, while 69% of Democrats believe that government should do more. This telephone poll result falls roughly in the middle of the results for the online samples.

Partisan views on government

% of Democrats who say govt should do more to solve problems vs. % of Republicans who say govt is doing too many things better left to businesses and individuals

	Democrat *Govt should do more*	Republican *Govt doing too many things*
I	76%	82%
H	77	77
ATP	73	77
E	78	70
G	77	69
C	76	68
Phone	69	74
F	76	64
B	66	73
D	74	65
A	68	66

Note: Phone estimate is from the Survey on Government conducted Aug. 27-Oct. 4, 2015.

Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

Interests and hobbies are generally consistent across samples

Many surveys seek to measure attitudes and lifestyle characteristics. Although no real benchmarks could be said to exist for these sorts of questions, one would expect them to be largely consistent across sample sources if the surveys are accurately representing the same population. In this study, we presented respondents with a list of 11 different activities and interests (e.g. reading the Bible, gardening, working out, celebrity news and gossip), and asked them to select each of the items of interest to them. Respondents could also select "None of the above."

The results were broadly consistent across the different sources. Working out was the top-ranked item for all but sample F, in which it came in second. Gardening was the second-highest-ranked item for all but samples E and F, where it ranked third and first respectively. Rankings of the least popular items were also fairly stable. For all samples, the three lowest-ranked items are comprised of only four items: NBA, NASCAR, hip-hop and "none of the above." None of the above was consistently the least frequently selected item except for samples E, G and I, for which the least commonly selected item was NASCAR. Interest in travel was the most variable category, ranging in rank from third in samples B and H to eighth in sample A. Art and theater was similarly variable, ranging from second in sample E to sixth in samples F and H.

Another way to evaluate these items is in the consistency of pairwise comparisons between individual items (e.g., which is more popular, reading the Bible or hunting and fishing?). Comparing every item to every other item yields a total of 66 comparisons. Of those 66 possible comparisons, all 10 samples agree on the more popular item for 48 of them (73%). For example, reading the Bible is more popular than hunting and fishing in all 10 of the samples, and international travel is always more popular than hip-hop. The 18 comparisons where samples do not agree perfectly are items that all tend to be within a few percentage points of each other in most samples. For instance, hip-hop is more popular than NASCAR in seven of 10 samples, and their respective levels of interest are generally within a few points of one another in each sample.

Despite general consistency, some patterns are apparent in certain samples. The American Trends Panel and sample E both show a tendency for respondents to favor choices such as art and theater and international travel. Sample E rates country music and NASCAR lower than the other samples (at 18% and 8% respectively). NASCAR and country music rate highly for sample A, while sample D stands out on interest in the NBA and hip-hop. This suggests that despite relatively good agreement overall for relative comparisons and rankings, idiosyncrasies in sample composition become apparent when the items are viewed in absolute terms.

# 4. Overall variability in estimates across samples

Generally, we expect that surveys representing the same target population should produce estimates that are consistent with one another; that is, measures taken at roughly the same time should not vary dramatically from survey to survey.

For some of the items in this study, this is the case. For example, the highest estimate for the proportion never married is only 4.4 percentage points higher than the lowest. The proportion living at their residence for less than a year and the proportion with a driver's license are similarly consistent across samples. At the other end of the spectrum, estimates of the proportion who knew that the Republican Party controlled both the House of Representatives and Senate span over a range of 26 percentage points, from 52% (sample I) to 78% (American Trends Panel) with the others dispersed in between.

Other items exhibit more idiosyncratic patterns. The estimates of the share of adults who worry that computers and technology are being used to

Estimates from different online samples run the gamut from highly variable to highly consistent

Deviation between each weighted survey estimate and the grand mean of estimates from all 10 online samples
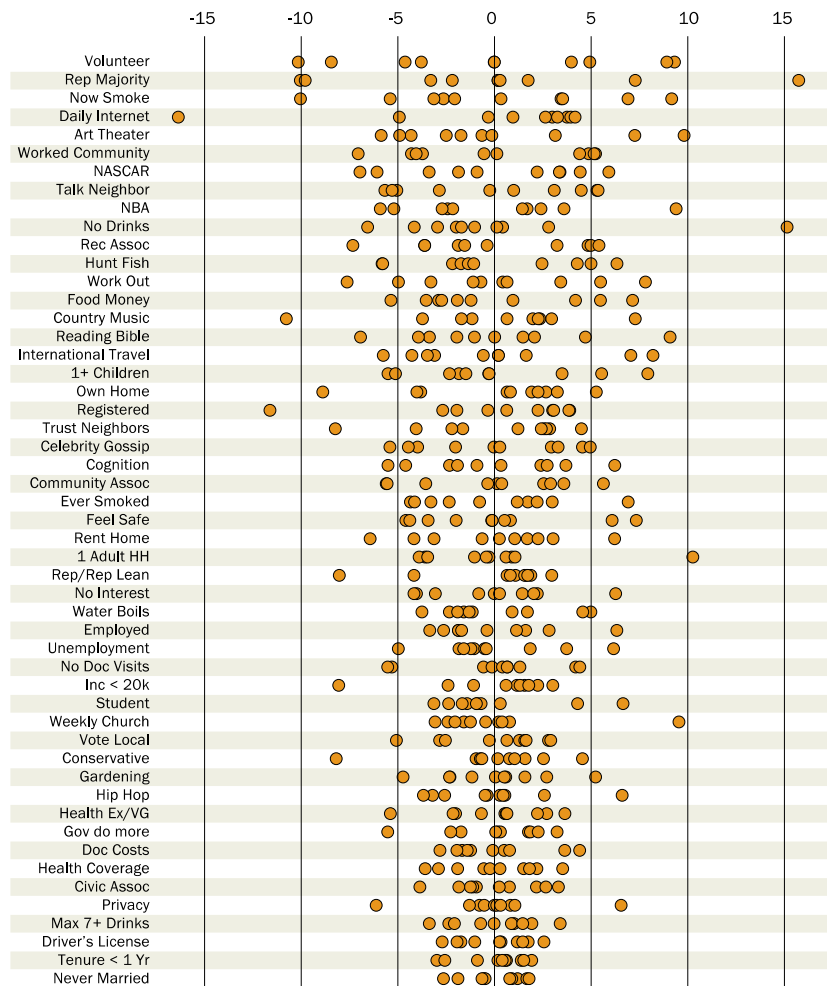
Source: Pew Research Center analysis of nine online nonprobability samples and the Center's American Trends Panel data. See Appendix A for details.
"Evaluating Online Nonprobability Surveys."

PEW RESEARCH CENTER

invade their privacy is densely clustered around 27% except for sample E, the low outlier at 21% and the ATP, the high outlier at 34%.

As we might expect, items closer to zero or 100% tend to display somewhat less variability between samples than items closer to 50%, though several items that fall near 50% exhibit low variability as well. For instance, the grand mean (average across all 10 samples) for the percentage who rate their health as excellent or very good is 47%, with samples differing from that by an average of 2.1 percentage points. The grand mean for the share that believes the government should do more to solve problems is 54% with an average deviation of 1.9 points.

Other than these ceiling and floor effects, there are no clear patterns that explain why some items vary across samples and others do not. Behavioral and attitudinal items are both well represented at both ends of the spectrum, as are potentially sensitive or socially desirable items. We also do not know if these patterns would persist if we were to repeat this exercise, as we lack measures of consistency over repeated surveys from the same sample sources.

5. Variation in online nonprobability survey design

One possible explanation for the variability in estimates across the different nonprobability samples is the range of methods that online sample vendors employ. They differ on recruitment, weighting and everything in between. Even if the impact of each of these differences is small, the cumulative effect could be much larger.

Panel recruitment and survey sampling

Methods to recruit respondents to modern nonprobability surveys differ across vendors. Some use ads in banners, on search engines or on social media. Some allow panelists to sign up directly at the panel vendor's website, while others recruit at the end of other web surveys. Some panel vendors directly recruit customers of partner companies, and some even turn to phone and mail recruitment for hard-to-reach demographic groups.

Vendors also differ on how they draw samples for a particular survey from among their own panelists or other sources available to them. Some draw a sample from their panel in much the same way a probability-based sample is drawn from a sampling frame, with subgroups selected at different rates depending on their propensity to respond and their desired proportion in the final sample. Most online sample providers do not draw samples for individual surveys, but rather invite panelists to an unspecified survey and route them to one of many surveys fielding simultaneously. The outcome of the routing is determined by the respondent's characteristics and algorithms that determine where each respondent is needed most. These routing algorithms make for much more efficient use of sample, but they do imply that a respondent's inclusion in any particular survey depends to some extent on what other surveys are fielding at the same time. These routing algorithms vary from provider to provider and their effects have received very little study by survey methodologists.

As previously discussed, sample providers typically apply some form of quota sampling during data collection to achieve a pre-specified distribution on some set of variables. Most panel vendors set quotas on some combination of age, gender and Census region. However, they differ on which of those variables they use as well as the categories into which responses are grouped. For example, one panel vendor might quota on male vs. female and separately on age groups of 18-29, 30-49, 50-64, and 65 and older. Another might have quotas set on the fully crossed age-by-gender categories of male 18-34, female 18-34, male 35-54, female 35-54, male 55 and older, and female 55 and older.

Some online sample vendors are offering more statistically sophisticated sampling techniques that go beyond setting basic quotas. One such approach, propensity score matching, involves assigning each panelist a score based on their likelihood of being in a probability reference sample (e.g., the Current Population Survey) – rather than in a nonprobability sample – given their demographic profile. Quotas are then based on quantiles of this propensity score rather than on specific respondent characteristics. A related technique uses statistical matching to achieve a desired sample composition. Under this approach, the vendor draws a subsample from a large probability sample, such as the CPS, and then looks for members of its own panel who closely resemble each case in the probability subsample on a number of variables. The survey is complete when a suitably close "match" has been identified for every case in the subsample. Both of these methods allow vendors to flexibly incorporate a larger number of respondent characteristics into the selection process than is possible with standard quotas, in theory improving their ability to correct for sources of selection bias.

The sample used for a survey might also not be limited to members of a particular vendor's panel. Sometimes panelists from multiple vendors' panels are sampled for a survey, especially if a low incidence or hard-to-reach group is being targeted. Additionally, some panel vendors offer the option of including "river sample" cases along with panel sample to make up the final survey sample. "River" sample is a term used when internet users are invited to take a survey through an advertisement or webpage without being required to join a panel. In some cases, answering survey questions allows them to access content that they would otherwise have to pay for, in an arrangement known as a "survey wall."

Weighting

Once the survey is out of the field there are also differences across vendors in how the data are weighted in order to be representative of the population of interest. A number of nonprobability sample vendors do not weight their data by default. Their view is that if the sample is properly balanced because of the quotas employed in sampling, weighting is unnecessary. When weights are provided, the technique used varies by vendor (e.g., from iterative proportional fitting or "raking" as a default practice to more sophisticated, generalized regression-based approaches for some custom surveys), as do the variables on which the data are weighted.

Strategies to increase quality: Incentives, monitoring, verification

Vendors also differ on the incentives they offer individuals to join their panel and/or respond to surveys. None of the vendors we tested offer direct monetary incentives for completing surveys. The more common approach is to incentivize panel members with points, which can be redeemed for consumer goods like gift cards and airline miles, as well as for cash. Other vendors offer

drawings or donations to charity. Some offer incentives only if a panelist qualifies for and completes a survey, while others offer incentives even to panelists who are sampled for a particular survey and complete screening questions but ultimately do not qualify for the survey.

Finally, each vendor has its own set of quality control measures, which can range from the simple to the complex. These measures may be implemented at the survey level, at the respondent level or at a combination of the two. They may include monitoring for speeding (when respondents answer questions rapidly and without actually considering the question) and straightlining (when respondents simply select the same answer choice to every question), as well as trap questions that check to make sure respondents are reading the questions carefully. They may also regulate the frequency with which panelists can be invited to take surveys or the frequency with which they can respond to surveys at their own initiative. Most panels are double opt-in, meaning potential panelists first enter an email address and then respond to an email sent from the panel provider in order to confirm the email account. Depending on the vendor, other quality control features include IP address validation and digital fingerprinting, which guards against a single person having multiple accounts in a given panel.

Acknowledgements

Appendix A: Survey methodology

The American Trends Panel (ATP), created by Pew Research Center, is a national panel of randomly selected U.S. adults living in households. The panel is being managed by Abt SRBI.

Members of the American Trends Panel were originally recruited from the 2014 Political Polarization and Typology Survey, a large (n=10,013) national landline and cellphone random-digit-dial (RDD) survey conducted Jan. 23 to March 16, 2014, in English and Spanish. At the end of that survey, respondents were invited to join the panel. The invitation was extended to all respondents who use the internet (from any location) and a random subsample of respondents who do not use the internet.[15]

Of the 10,013 adults interviewed, 9,809 were invited to take part in the panel. A total of 5,338 agreed to participate and provided either a mailing address or an email address to which a welcome packet, a monetary incentive and future survey invitations could be sent. Panelists receive a small monetary incentive after participating in each wave of the survey.

For the ATP, the questions used in this report were asked on different waves of the panel, which are fielded roughly once a month. Estimates for each question are calculated using the respondents to the wave in which it was asked. Some items such as demographics were measured at recruitment and updated periodically, in which case they do not belong to any individual wave. For these kinds of questions, the respondents to Wave 10 were used to produce the estimates in this report.

For all but Wave 5, ATP panelists who self-identify as internet users and who provided an email address participate in the panel via monthly self-administered Web surveys, and those who do not use the internet, do not have an email address or refuse to provide their email address participate via the mail.

Wave 5 featured an experiment comparing telephone and web-based survey administration. For this wave, half of the respondents who are usually interviewed online were surveyed via the telephone. All of the panelists who are usually surveyed by mail were also surveyed by telephone. In order to minimize mode effects, the internet panelists who were surveyed by telephone were excluded from this analysis. The non-internet panelists were retained in order to ensure that the offline population was still represented in the sample. The proportion of the sample with and

---

[15] When data collection for the 2014 Political Polarization and Typology Survey began, non-internet users were subsampled at a rate of 25%, but a decision was made shortly thereafter to invite all non-internet users to join. In total, 83% of non-internet users were invited to join the panel.

without internet access was then corrected as part of the weighting process described in Appendix B.

For the nonprobability surveys included in this report, sample was obtained from the vendors, but the survey was administered using the SurveyMonkey platform in order to ensure that respondents all experienced identical survey instruments. The exception is sample I, which was not able to interface with the SurveyMonkey platform and was administered using vendor I's proprietary survey software.

The field dates and sample sizes for each ATP wave and nonprobability survey are presented in the table below. For the ATP, we report the cumulative response rate that incorporates the response rate for the 2014 Survey of Political Polarization (10.6%) and attrition from panel members who were removed at their request or for inactivity as well as nonresponse to the individual waves. No response rate is provided for the nonprobability surveys as it is not a meaningful metric in the absence of random selection from a population frame.

Summary design and outcome metrics

Sample	Interviews	Field dates	Cumulative response rate	95% margin of error[2]
ATP: W5	1,857	July 7-August 4, 2014	3.7[1]	3.0
ATP: W6	3,278	August 11-September 3, 2014	3.6	2.3
ATP: W7	3,154	September 9-October 3, 2014	3.5	2.7
ATP: W9	3,212	November 17-December 15, 2014	3.5	2.3
ATP: W10	3,147	March 10-April 6, 2015	3.4	2.4
A	1,022	February 25, 2015	N/A	3.5
B	1,049	February 26-March 3, 2015	N/A	4.2
C	1,178	February 25-27, 2015	N/A	3.8
D	1,005	February 25-27, 2015	N/A	3.6
E	1,022	February 24-March 8, 2015	N/A	5.2
F	1,008	February 25-26, 2015	N/A	4.4
G	1,010	October 1-6, 2015	N/A	4.7
H	1,007	October 2-8, 2015	N/A	4.3
I	1,000	August 19-31, 2015	N/A	4.3

[1] The cumulative response rate for Wave 5 is for the full sample.

[2] The benchmarking analysis in this report indicates that these 95% margins of error underestimate the actual margins of error at least for some survey estimates, such as those related to civic or political engagement.

PEW RESEARCH CENTER

Precision estimates

Precision estimates for this study were computed using the Taylor series approximation. The estimates treat the nonprobability samples as if they were drawn as simple random samples from the population. This is not an accurate description of the sampling mechanisms, but it allows us to attempt to quantify and compare the variability observed in these samples. The precision estimates for the nonprobability samples account for the increase in variance due to weighting (design effect).

The precision estimates for the ATP reflect both the actual sample design and the raking adjustments. Specifically, the ATP weights reflect the differential probabilities of selection for the recruitment telephone survey, a propensity model to adjust for differential likelihood of joining and participation in the ATP, as well as differential nonresponse to the individual ATP wave.

The precision calculations assume that the survey estimates are approximately unbiased, which was shown in the benchmarking analysis to be a flawed assumption. It is important to bear in mind that these precision statements only reflect sampling error and do not account for other sources of error such as noncoverage, nonresponse, or measurement error.

Appendix B: Weighting

The ATP and nonprobability surveys included in this report were weighted according to a modified version of the standard weighting procedure used for the ATP.

Each wave of the ATP used in this report was weighted in a multi-step process that begins with a base weight incorporating the respondents' original survey selection probability and the fact that some panelists were subsampled for invitation to the panel. Next, an adjustment was made for the fact that the propensity to join the panel and remain an active panelist varied across different groups in the sample.

Both the ATP and the nonprobability surveys were then weighted using an iterative technique that matches gender, age, education, race, Hispanic origin and region to parameters from the U.S. Census Bureau's 2013 American Community Survey. Population density is weighted to match the 2010 U.S. Decennial Census. Telephone service is weighted to estimates of telephone coverage for 2014 that were projected from the July-December 2013 National Health Interview Survey.

The ATP was also adjusted for internet access using as a parameter a measure from the 2014 Survey of Political Polarization. Because the nonprobability surveys do not include any respondents who do not have access to the internet, they cannot be weighted on this dimension.

The standard ATP weighting process adjusts party identification to match the three most recent Pew Research Center general public telephone surveys; however, this was not done for this report in order to allow analysis of variation in party identification across panels.

Some of the panels provided weights for their surveys. For each panel, we used the set of weights that yielded the lowest average absolute bias in the benchmarking analysis. Sample I was the only sample where the vendor-supplied weights outperformed the ATP-style weights. As such, all analyses in this report use the vendor-supplied weights for sample I and the ATP-style weights for all other nonprobability samples.

Appendix C: Weighted estimates

Weighted estimates for non-demographic questions

% of respondents in the reference category

Description	Bench-mark	ATP[9]	A	B	C	D	E	F	G	H	I
Talk with neighbors a few times a week or more	39[3]	51	50	43	49	51	40	40	45	46	40
Have worked with other people from your neighborhood	8[5]	33	33	24	33	33	28	24	28	24	21
Trust all/most people in your neighborhood	51[3]	52	56	57	55	58	45	56	52	56	50
Very safe from crime when walking in your neighborhood after dark	-	46	37	39	39	39	45	34	34	39	35
Worry a lot that computers and technology are being used to invade your privacy	-	34	28	27	27	28	21	27	27	28	26
Interested in country music	-	31	36	28	29	32	18	31	31	27	25
Interested in reading the Bible	-	39	34	23	28	31	26	26	30	29	32
Interested in hunting or fishing	-	27	25	19	24	26	15	19	20	15	20
Interested in working out	-	50	41	41	42	45	47	34	37	43	39
Interested in Hip Hop	-	17	20	17	17	24	14	16	17	14	13
Interested in art and theater	-	38	30	27	31	34	41	25	29	26	29
Interested in celebrity news and gossip	-	18	25	22	27	27	18	25	22	20	17
Interested in gardening	-	43	38	39	40	41	39	36	33	36	37
Interested in international travel	-	37	24	28	29	30	36	23	26	29	25
Interested in NASCAR	-	12	21	17	18	19	8	18	14	13	9
Interested in NBA	-	17	23	22	17	29	14	21	17	21	14
Interested in none of the above	-	5	5	10	8	6	9	11	15	9	11
Prefer complicated problems	-	35	35	28	33	36	39	30	27	31	32
Participated in a school group, neighborhood, or community assoc. in the past 12 mon.	13[3]	22	26	16	23	23	20	14	20	20	14
Participated in a service or civic org. in the past 12 mon.	6[3]	14	15	12	16	16	13	12	12	11	9
Participated in a sports or recreation org. in the past 12 mon.	9[3]	20	21	16	21	22	13	14	15	13	9
Volunteered in past 12 mon.	25[5]	58	52	44	53	48	57	40	48	44	38

Description	Bench-mark	ATP[9]	A	B	C	D	E	F	G	H	I
Usually access the internet every day over the last year	62[4]	69	89	86	88	89	89	88	85	88	80
Usually use the internet 6 or more hours per day	-	-	46	36	42	45	38	40	41	34	31
At least occasionally access the internet on a cell phone, tablet, or other mobile handheld device	-	-	71	64	72	73	77	66	65	70	68
In general health is excellent/very good	61[8]	46	50	49	47	47	45	44	47	49	41
Needed to see a doctor but could not because of the cost in the past 12 mon.	-	27	24	22	23	28	22	24	22	21	20
Zero doctor visits during the past 12 mon.	-	14	20	24	21	19	14	24	20	20	20
Didn't drink any alcoholic beverages during the past 30 days	-	41	40	38	39	36	43	41	57	43	45
Had 7 or more drinks during the past 30 days	-	7	9	10	8	11	9	9	4	5	6
Have smoked at least 100 cigarettes in your entire life	-	47	46	47	42	51	40	46	41	40	44
Now smoke cigarettes every day	13[8]	15	24	21	18	26	7	21	15	14	12
Government should do more to solve problems	-	48	54	52	56	56	56	54	57	54	52
Republicans currently have a majority in both the House and the Senate (Correct answer)	-	78	62	64	60	62	69	62	52	52	59
Always vote in local elections	32[3]	35	40	40	41	41	41	43	37	43	37
Contacted or visited a public official in the last 12 mon. to express your opinion	-	-	23	20	24	25	33	22	22	22	28
Registered to vote at current address	69[6]	76	73	77	71	76	76	74	71	77	62
Republican/Republican Lean	-	42	42	42	43	44	33	43	43	43	37
Very conservative/ Conservative	-	32	34	34	33	37	25	32	34	32	35
Water boils at a lower temperature in Denver than Los Angeles (Correct answer)	-	34	28	27	28	30	34	25	27	28	31
Antibiotics will kill viruses as well as bacteria- False (Correct answer)	-	-	54	57	58	54	74	54	55	59	60
Have not had enough money to buy food your family needed in the last year	-	28	34	26	33	36	24	27	30	25	26
Household received any state or federal unemployment compensation in the last year	4[2]	5	12	9	14	16	9	9	9	8	8

Description	Bench-mark	ATP[9]	A	B	C	D	E	F	G	H	I
Currently have a valid driver's license	86[7]	86	83	85	86	84	85	86	83	87	82
Attend religious services more than once a week/once a week	-	36	27	24	25	28	24	25	27	26	26
Have any kind of health care coverage	86[1]	84	84	81	85	86	87	82	83	88	86
Never married	30[1]	25	26	29	29	27	29	29	30	27	27
Employed	-	57	49	48	50	48	52	49	52	54	49
Student	-	15	7	6	9	7	13	5	7	8	7
One-adult household	19[1]	20	25	25	25	23	20	24	34	21	24
One or more children	35[1]	27	35	29	33	37	24	27	24	29	28
Own the place where you are living	-	49	54	61	60	60	61	59	59	63	54
Rent the place where you are living	-	28	41	35	36	36	32	34	37	31	38
Have lived at this address 1 year or less	15[1]	17	15	15	15	16	15	12	16	12	14
Have lived at this address 5 years or more	-	51	56	61	56	56	62	62	60	61	60
Household's combined annual income $20,000 or less	16[2]	22	22	22	20	23	18	21	22	13	24

[1]ACS 2014
[2]CPS ASEC (March 2015)
[3]CPS Civic Engagement Supplement (November 2013)
[4]CPS Internet Supplement (July 2013)
[5]CPS Volunteer Supplement (September 2013)
[6]CPS Voting Supplement (November 2014)
[7]FHWA 2014
[8]NHIS 2014
[9]American Trends Panel Waves 5/7/10

PEW RESEARCH CENTER

Appendix D: Benchmarks

Sources and details for benchmarks

Benchmark item	Source	Question text	Response category	Benchmark estimate (%)	Notes
Talk with neighbors basically every day or a few times a week	CPS Civic Engagement Supplement (Nov 2013)	How often did you talk with any of your neighbors—basically every day, a few times a week, a few times a month, once a month, less than once a month, or not at all?	Basically every day/ A few times a week	39	
Have worked with other people from your neighborhood for your community	CPS Volunteer Supplement (Sep 2013)	Since September 1st, 2013 have you worked with people in your neighborhood to fix or improve something?	Yes	8	
Have participated in a school group, neighborhood, or community association in the past 12 months	CPS Civic Engagement Supplement (Nov 2013)	Have you participated in any of these groups during the last 12 months, that is since November 2012: A school group, neighborhood, or community association such as PTA or neighborhood watch group?	Yes	13	
Have participated in a service or civic organization in the past 12 months	CPS Civic Engagement Supplement (Nov 2013)	Have you participated in any of these groups during the last 12 months, that is since November 2012: A service or civic organization such as American Legion or Lions Club?	Yes	6	
Have participated in a sports or recreation organization in the past 12 months	CPS Civic Engagement Supplement (Nov 2013)	Have you participated in any of these groups during the last 12 months, that is since November 2012: A sports or recreation organization such as a soccer club or tennis club?	Yes	9	

Sources and details for benchmarks (continued)

| Benchmark item	Source	Question text	Response category	Benchmark estimate	Notes
Usually accessed the internet everyday over the last year	CPS Computer and Internet Use Supplement (July 2013)	How often did you USUALLY access the Internet over the last year? Consider time spent on the Internet from any computer or mobile device at home, work, or any other location. Did you usually access the Internet...Every Day, More than once a week but not every day, once a week, less than once a month or never?	Every day	62	According to the CPS, daily internet use grew by 1.8 percentage points between 2011 and 2013. In order to account for continued growth past 2013, the benchmark value used in this study is the 2013 CPS figure plus an additional 1.8 percentage points. In reality, the actual change over time during this period may not have been linear, but our approach stays as close as possible to the CPS data that was available.
In general health is excellent/very good	National Health Interview Survey (2014)	Would you say your health in general is excellent, very good, good, fair, or poor?	Excellent/ Very good	61	
Now smoke cigarettes every day	National Health Interview Survey (2014)	Do you NOW smoke cigarettes every day, some days or not at all?	Every day	13	
Always vote in local elections	CPS Civic Engagement Supplement (Nov 2013)	Do you always vote in local elections, do you sometimes vote, do you rarely vote, or do you never vote?	Always vote	32	
Registered to vote at current address	CPS Voting and Registration Supplement (Nov 2014)	In any election, some people are not able to vote because they are sick or busy or have some other reason, and others do not want to vote. Did you vote in the election held on Tuesday, November 4, 2014? (IF NO) Were you registered to vote in the November 4, 2014 election?	Voted in the election or registered to vote if did not vote	69	This estimate uses the adjustment recommend in Hur and Achen (2013) to correct for bias resulting from the fact that item nonrespondents are treated as not having voted in the CPS. Adjustment factors for 2014 can be found at: http://www.electproject.org/home/voter-turnout/cps-methodology
Household received any state or federal unemployment compensation in the last year	CPS Annual Social and Economic Supplement (Mar 2015)	At any time during 2014 did you receive state or federal unemployment compensation?	Anyone in the household received state or federal unemployment compensation	4	The variable used to produce this estimate is a recode that aggregates responses for individuals up to the household level.
Currently have a valid driver's license	Federal Highway Administration (2014)	n/a	Number of adults 18+ with driver's license from FHWA divided by the total number of adults from 2014 ACS	86	Data used to calculate this figure is available at: https://www.fhwa.dot.gov/policyinformation/statistics/2014/dl20.cfm

Sources and details for benchmarks (continued)

|				
Has health care coverage	American Community Survey (2014)	Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans?	Selected at least one type of health coverage	86
Never married	American Community Survey (2014)	What is this person's marital status?	Never married or under 15 years old	30
One-adult household	American Community Survey (2014)	n/a	Sum of individuals in household 18 or over is equal to 1	19	This figure is calculated by counting the number of adults enumerated in each ACS household.
One or more children	American Community Survey (2014)	n/a	Sum of individuals in household under 18 is greater than or equal to 1	35	This figure is calculated by counting the number of children under 18 in each ACS household.
Have lived at this address 1 year or less	American Community Survey (2014)	Did this person live in this house or apartment 1 year ago?	Yes	15
Household's combined annual income $20,000 or less	CPS Annual Social and Economic Supplement (Mar 2015)	Which category represents the total combined income of all members of this FAMILY during the past 12 months?	Income less than $20,000	16

Notes: The survey also included several questions measured in the Behavioral Risk Factor Surveillance Survey (BRFSS), but those were not considered reliable population benchmarks given limitations in the BRFSS design. One question about the number of doctor visits in the last year (Q0021) matches a question in the National Health Interview Survey; however, it was excluded from the benchmarking analysis because differences in the preceding questionnaire content and administration were deemed likely to produce substantial context effects.

PEW RESEARCH CENTER

Appendix E: Absolute bias on benchmarks

This appendix contains the estimated absolute bias on 20 benchmarks for all of the samples evaluated in this report. Average absolute bias is calculated by subtracting the benchmark value from the weighted point estimate and taking the absolute value.

Sample I

Sample H

Sample F

Sample B

ATP

Sample G

Sample E

Sample C

Sample A

Sample D
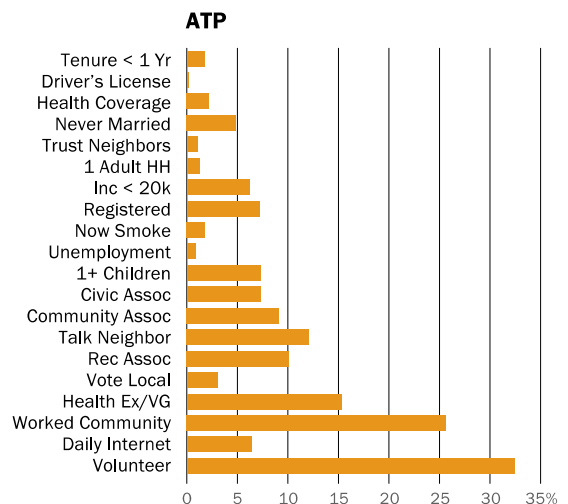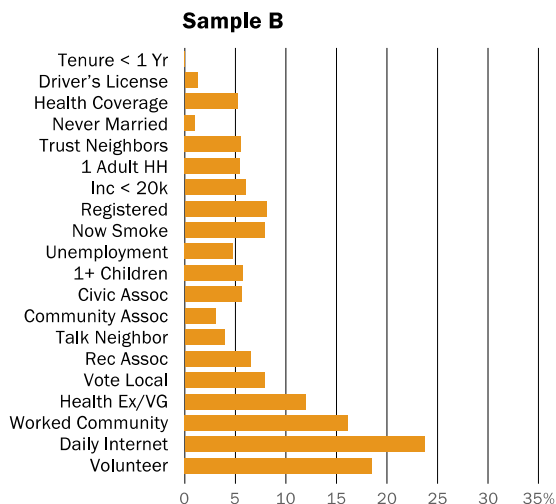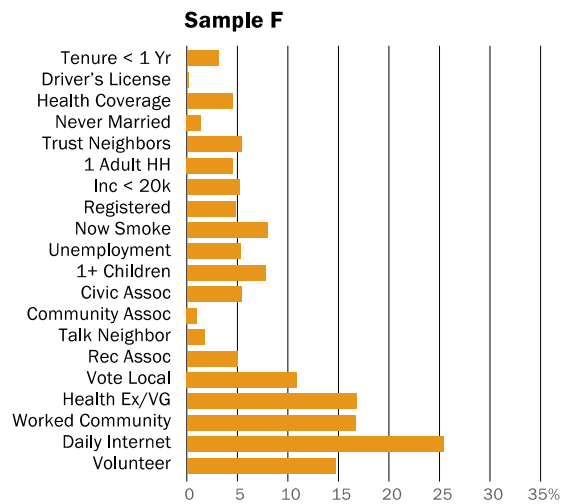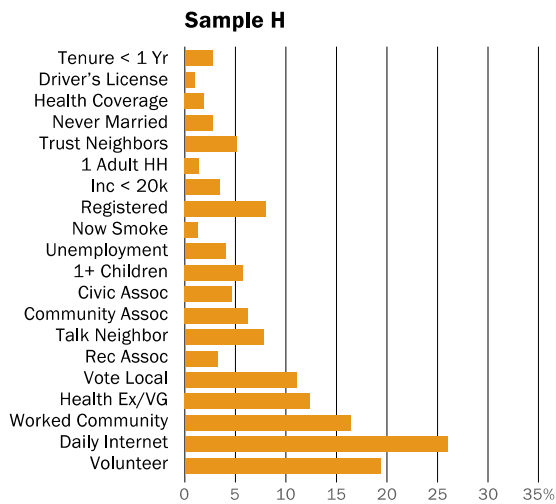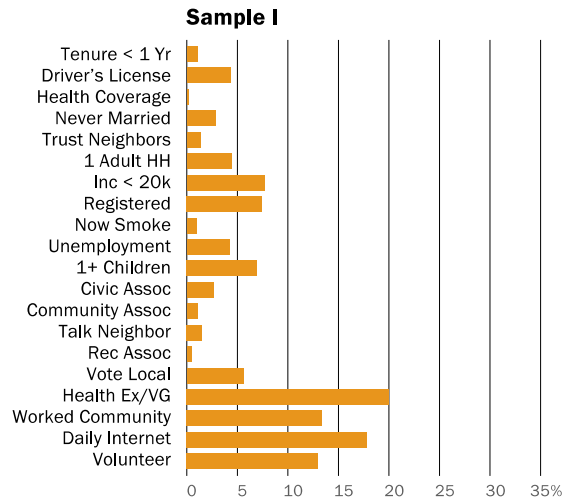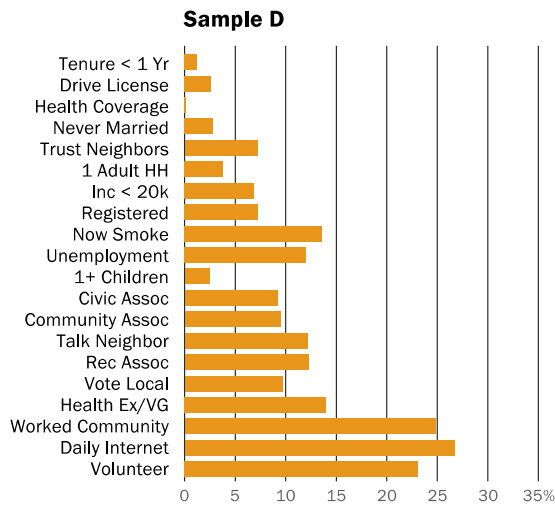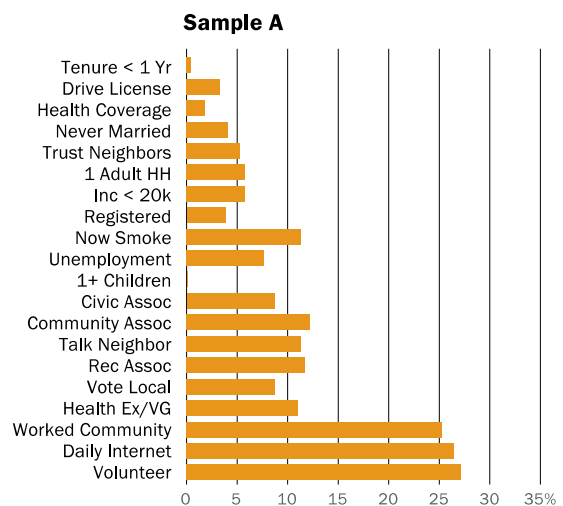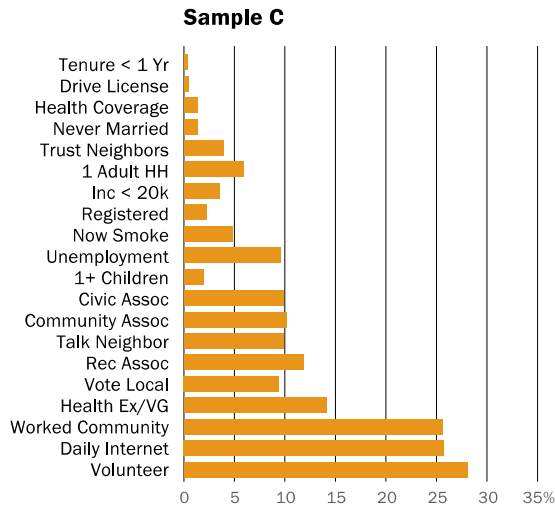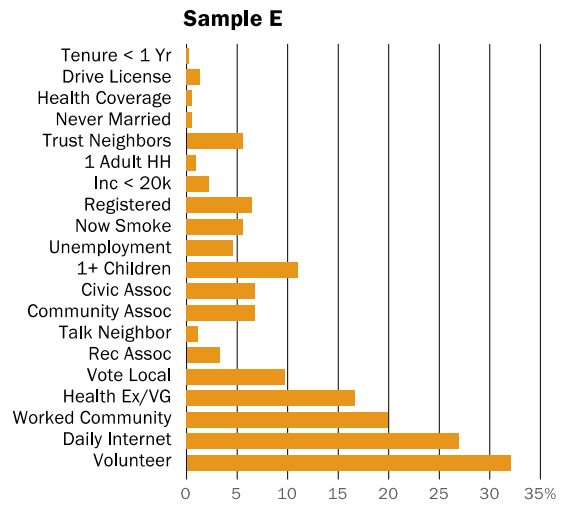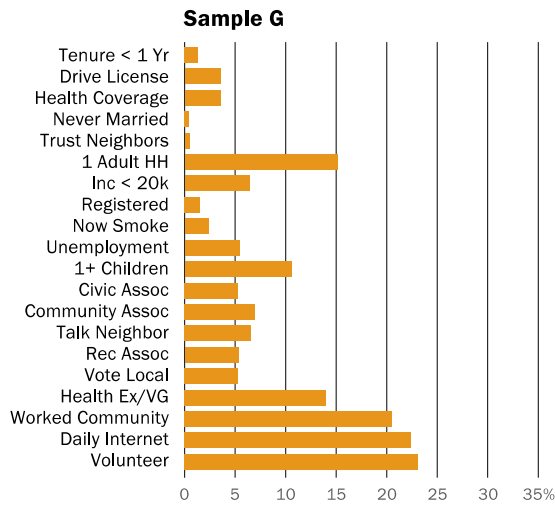
Appendix F: Questionnaire

**PEW RESEARCH CENTER
ONLINE NONPROBABILITY LANDSCAPE STUDY
STUDY OF PEOPLE AND COMMUNITIES
Questionnaire for Programming**

Thank you for participating in the Study of People and Communities.

What is the Study of People and Communities?
The Study of People and Communities is a survey that collects information about how a neighborhood or community affects a person's life.

Who is the sponsor of this study?
This study is sponsored jointly by The Pew Research Center and WESTAT. The Pew Research Center is a nonpartisan organization that provides information to the public on important events and trends in both America and the world. Pew Research does not take policy positions. WESTAT is a nonpartisan survey research firm that conducts statistical research for government and other organizations.

How long will it take to complete this survey?
Most individuals will be able to complete the survey in about 5-8 minutes.

Who will use this information?
Results from this study will be used by researchers and made available to policy makers at all levels of government and private industry. In particular, this survey will describe how individuals are affected by the communities and neighborhoods they live in. This helps policy makers understand how to best serve these individuals, areas, and communities.

How do I know you'll keep my information confidential?
We are required by law to keep your information confidential to the full extent protected by law. After the study is completed, all information used to contact you will be destroyed.

ASK ALL:
Q0001   What do you think is the most important problem facing the country today?

                [PROGRAMMING NOTE: OPEN-END TEXT BOX]

ASK ALL:
Q0002   During a typical month in the past year, how often did you talk with any of your neighbors?

        1   Basically every day
        2   A few times a week
        3   A few times a month
        4   Once a month
        5   Not at all

ASK ALL:
Q0003   In the past year have you worked with other people from your neighborhood to fix a problem or improve a condition in your community or elsewhere?

        1   Yes
        2   No

ASK ALL:
Q0004   How much do you trust the people in your neighborhood? In general, do you trust…

        1   All of the people in your neighborhood
        2   Most of the people in your neighborhood
        3   Some of the people in your neighborhood
        4   None of the people in your neighborhood

ASK ALL:
Q0005   In general, how safe would you say you are from crime when walking in your neighborhood after dark?

        1   Very safe
        2   Somewhat safe
        3   Not too safe
        4   Not at all safe

ASK ALL:
Q0006   How much do you worry that computers and technology are being used to invade your privacy?

        1   A lot
        2   Some
        3   Not much
        4   Not at all

ASK ALL:
Q0007   Below is a list of things that some people are interested in, and others are not. Which of the following are you interested in?

[Check all that apply]

[PROGRAMMING NOTE: RANDOMIZE A-K WITH L AWLAYS LAST]

     a.   Country music
     b.   Reading the Bible
     c.   Hunting or fishing
     d.   Working out (e.g. yoga, cycling, hiking)
     e.   Hip hop
     f.   Art and theater
     g.   Celebrity news and gossip
     h.   Gardening
     i.   International travel
     j.   NASCAR
     k.   NBA basketball
     l.   None of the above **[EXCLUSIVE PUNCH]**

ASK ALL:
Q0008   Where would you rate yourself on the following scale?

[PROGRAMMING NOTE: REVERSE RESPONSE OPTION SCALE FOR RANDOM HALF OF RESPONDENTS]

     1   1 I prefer simple problems rather than those that require a lot of thought
     2   2
     3   3
     4   4
     5   5 I prefer complicated problems that require new solutions and a lot of thought

---

ASK ALL:
Below is a list of types of groups or organizations in which people sometimes participate. Have you participated in any of these groups during the last 12 months, that is since February 2014?

Q0009   A school group, neighborhood, or community association such as PTA or neighborhood watch group?

     1   Yes
     2   No

Q0010   A service or civic organization such as American Legion or Lions Club?

     1   Yes
     2   No

Q0011   A sports or recreation organization such as a soccer club or tennis club?

     1   Yes
     2   No

ASK ALL:

Q0012   We are interested in volunteer activities for which people are not paid, except perhaps expenses. We only want you to include volunteer activities that you did through or for an organization, even if you only did them once in a while. In the last 12 months that is since July of last year, have you done any volunteer activities through or for an organization?

     1   Yes
     2   No

ASK IF HAVE NOT VOLUNTEERED (Q0012=2, MISSING):

Q0013   Sometimes people don't think of activities they do infrequently or activities they do for children's schools or youth organizations as volunteer activities. Since July of last year, have you done any of these types of volunteer activities?

     1   Yes
     2   No

ASK ALL:

Q0014   Next we'd like to ask about computers and the Internet. For the following questions, consider time spent on the Internet from a computer or mobile device at home, work, or any other locations.

     How often did you USUALLY access the Internet over the last year?

     1   Every day
     2   More than oce a week but not every day
     3   Once a week
     4   Once a month
     5   Less than once a month
     6   Never

ASK IF "EVERY DAY" (Q0014=1):

Q0015   How many hours per day do you USUALLY use the Internet, including time spent at work?

     [PROGRAMMING NOTE: OPEN-END TEXT BOX]

ASK IF "MORE THAN ONCE A WEEK", OR "ONCE A WEEK" (Q0014=2-3):

Q0016   How many hours per week do you USUALLY use the Internet, including time spent at work?

     [PROGRAMMING NOTE: OPEN-END TEXT BOX]

ASK IF "ONCE A MONTH", OR "LESS THAN ONCE A MONTH" (Q0014=4-5):
Q0017   How many hours per month do you USUALLY use the Internet, including time spent at work?

          [PROGRAMMING NOTE: OPEN-END TEXT BOX]

ASK ALL:
Q0018   Do you access the Internet on a cell phone, tablet or other mobile handheld device, at least occasionally?

          1   Yes
          2   No

ASK ALL:
Moving onto another topic…

Q0019   Would you say that in general your health is…

          1   Excellent
          2   Very Good
          3   Good
          4   Fair
          5   Poor

ASK ALL:
Q0020   Was there a time in the past 12 months when you needed to see a doctor but could not because of the cost?
          1   Yes
          2   No

ASK ALL:
Q0021   During the past 12 months, how many times have you seen a doctor or other health care professional about your own health at a doctor's office, a clinic, or some other place? Do not include times you were hospitalized overnight, visits to hospital emergency rooms, home visits, dental visits, or telephone calls.

          [PROGRAMMING NOTE: OPEN-END TEXT BOX; ALLOW ANSWER RANGE OF 0-500 visits]

          SOFT PROMPT TEXT: *"Please enter number of visits in the box. If you would like to skip this question, click Next."*

ASK ALL:
Q0022   During the past 30 days, how many days did you have at least one drink of any alcoholic beverages such as beer, wine, a malt beverage, or liquor?

          [PROGRAMMING NOTE: OPEN END SHORT NUMERIC TEXT FIELD; INTEGERS 0-30. NUMERIC FIELD ONLY, DO NOT ALLOW TEXT.]

          Error message if outside range: "Please enter a number between 0 and 30"

ASK ALL:
Q0023 **During the past 30 days, what is the largest number of drinks you had on any occasion?**

[PROGRAMMING NOTE: OPEN END SHORT NUMERIC TEXT FIELD; RANGE 1-30. NUMERIC FIELD ONLY, DO NOT ALLOW TEXT.]

[PROGRAMMING NOTE: PROMPT TO DISPLAY THIS IF ENTRY IS GREATER THAN 30 AND ALLOW TO CONTINUE]
Please confirm your entry. We are interested in the largest number of drinks on any one occasion. Click Next to continue.

ASK ALL:
Q0024   Have you smoked at least 100 cigarettes in your entire life?
    1   Yes
    2   No

ASK IF HAS SMOKED 100 CIGARETTES (Q0024=1):
Q0025 Do you NOW smoke cigarettes…

    1   Every day
    2   Some days
    3   Not at all

ASK ALL:
We have a few questions about elections and politics.

ASK ALL:
Q0026   What ONE WORD best describes your impression of politics today?

[PROGRAMMING NOTE: OPEN-END TEXT BOX]

ASK ALL:
Q0027   Which statement comes closer to your views, even if neither is exactly right?
[PROGRAMMING NOTE: RANDOMIZE RESPONSE OPTIONS]

    1   Government should do more to solve problems
    2   Government is doing too many things better left to businesses and individuals

---

ASK ALL:
Q0028   Thinking about Congress, do Republicans currently have a majority in the House of Representatives, the Senate, both the House and Senate, or neither the House nor the Senate?

[PROGRAMMING NOTE: RANDOMIZE RESPONSE OPTIONS]

1   The House of Representatives
2   The Senate
3   Both the House and Senate
4   Neither the House nor Senate
5

ASK ALL:
Q0029   The next question is about LOCAL elections, such as for mayor or a school board. Do you…

1   Always vote in local elections
2   Sometimes vote in local elections
3   Rarely vote in local elections
4   Never vote in local elections

ASK ALL:
Q0030   In the past 12 months, have you contacted or visited a public official—at any level of government—to express your opinion?
1   Yes
2   No

ASK ALL:
Q0031   Are you registered to vote at your current address?

1   Yes, I am absolutely certain I am registered to vote at my current address
2   Maybe, I probably am registered to vote at my current address, but there is a chance that registration has lapsed
3   No, I am not registered to vote at my current address

ASK ALL:
Q0032   In politics today, do you consider yourself a:

1   Republican
2   Democrat
3   Independent
4   Something else, please specify **[OPEN-END; TEXT BOX]**:

ASK IF INDEP/SOMETHING ELSE (Q0031=3 or 4) OR MISSING:
Q0033   As of today do you lean more to…
1   The Republican Party
2   The Democratic Party

ASK ALL:
Q0034   In general, would you describe your political views as…

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

ASK ALL:
Moving onto other topics…

Q0035   Denver, CO is a higher altitude than is Los Angeles, CA. Which of these statements is correct? (We'd like your best guess)

     [PROGRAMMING NOTE: RANDOMIZE RESPONSE OPTIONS]

     1   Water boils at a lower temperature in Denver than Los Angeles
     2   Water boils at a higher temperature in Denver than Los Angeles
     3   Water boils at the same temperature in both Denver and Los Angeles

ASK ALL:
Q0036   Please indicate if the following statement is true or false: Antibiotics will kill viruses as well as bacteria.

     1   True
     2   False
     3   Don't know

ASK ALL:
We are interested in how people have been doing financially during the past year.
Q0037   Have there been times during the last year when you did not have enough money to buy food your family needed?
     1   Yes
     2   No

ASK ALL:
Q0038   At any time in the last year did you or anyone in your household receive any State or Federal unemployment compensation?
     1   Yes
     2   No

ASK ALL:
Q0039   Is there at least one telephone INSIDE your home that is currently working and is not a cell phone?
     1   Yes
     2   No

ASK ALL:
Q0040   Do you or anyone in your family have a working cell phone?
     1   Yes
     2   No

ASK ALL:

Q0041   Do you currently have a valid driver's license, or not?

       1   Yes

       2   No

ASK ALL:

And finally, a few questions about yourself and your household.

Q0042   What is your gender?

       1   Female

       2   Male

ASK ALL:

Q0043   What is your age?

       [PROGRAMMING NOTE: Numeric text box, 5 characters wide, range 18-120]

       _____years

ASK ALL:

Q0044   Are you of Hispanic, Latino, or Spanish origin, such as Mexican, Puerto Rican or Cuban?

       1   Yes, Hispanic or Latino

       2   No, not Hispanic or Latino

ASK ALL:

Q0045   Which of the following describes your race?

       [You can select as many as apply]

       1   White

       2   Black of African-American

       3   Asian

       4   American Indian or Alaska Native

       5   Native Hawaiian or other Pacific Islanders

       6   Some other race, specify:_____

ASK ALL:

Q0046   Aside from weddings and funerals, how often do you attend religious services?

       1   More than once a week

       2   Once a week

       3   Once or twice a month

       4   A few times a year

       5   Seldom

       6   Never

ASK ALL:
Q0047   Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare or Indian Health Service?

    1   Yes, I have health care coverage
    2   No, I do not have health care coverage

ASK ALL:
Q0048   Are you currently…
    1   Married
    2   Divorced
    3   Separated
    4   Widowed
    5   Never Married
    6   Living with a Partner

ASK ALL:
Q0049   Are you currently…

    [Please select all that apply]

    1   Employed for wages
    2   Self-employed
    3   Out of work for less than 1 year
    4   Out of work for more than 1 year
    5   A stay-at-home parent or spouse
    6   A student
    7   Retired
    8   Unable to work

ASK ALL:
Q0050   What is the highest grade or year of school you completed?

    1   Never attended school or only attended kindergarten
    2   Grades 1 through 8 (Elementary School)
    3   Grades 9 through 11 (Some High School)
    4   Grade 12 or GED (High School Graduate)
    5   Completed some college
    6   Completed technical school
    7   Associate degree
    8   Bachelor's degree
    9   Completed some postgraduate
    10   Master's degree
    11   Ph.D., law, or medical degree
    12   Other advanced degree beyond a Master's degree

ASK ALL:
Q0051   How many adults, ages 18 and older, including yourself, live in your household?

    [PROGRAMMING NOTE: Numeric text box, 5 characters wide, range 0-20]

    _____

ASK ALL:
Q0052   And how many children younger than 18 years of age live in your household? (Please fill in zero "0" if no children)

[PROGRAMMING NOTE: Numeric text box, 5 characters wide, range 0-20]

_____

ASK ALL:
Q0053   Do you own or rent the place where you are living?

1   Own
2   Rent
3   Other, describe below**: [OPEN-END**; **TEXT BOX]**

ASK ALL:
Q0054   How long have you lived at this address?
1   1 year or less
2   Less than 5 years, but more than 1 year
3   5 years or more

ASK ALL:
Q0055   What is your zip code?

[PROGRAMMING NOTE: Numeric text box, 5 characters wide, range 0-99999]

_____

ASK ALL:
Q0056   What is your household's combined annual income?

1   Less than $10,000
2      $10,000 to $19,999
3      $20,000 to $39,999
4      $40,000 to $59,999
5      $60,000 to $79,999
6      $80,000 to $99,999
7      $100,000 to $149,999
8      $150,000 to $249,999
9      $250,000 or more