Comparative Evaluation of PPC Non-Probability Sample

Evan Fehsenfeld, Evan Charles Lewitus, Steven Kull

Executive Summary

Currently, one of the most central questions in survey research is how well online opt-in samples (i.e. non-probability samples) can accurately represent a population, and how they compare with probability-based samples surveyed online, or by telephone or mail, which are widely viewed as the "gold standard".

Studies have been conducted in which multiple probability-based and opt-in sampling firms were asked to conduct the same survey, which included questions about key population variables. The samples' responses were then compared to federal benchmarks for those population variables derived from high-quality federal surveys with very large samples ranging from tens of thousands to over three million US residents and thus with very low margins of error. The amount that each population variable in a sample deviates from federal benchmarks was calculated. Each sample's overall error rate was then calculated by taking an average of those deviations.

The most prominent of such studies were by MacInnis et al. (2018) and the Pew Research Center (2016, 2023).

Each study found that survey firms using opt-in sampling consistently produce samples that are less representative than firms using probability-based sampling. Each study also found that there is much more variability between the samples obtained by opt-in sampling firms than between those obtained by probability-sampling firms. In nearly all cases opt-in samples do not do as well as the probability-based samples, though there have been a couple exceptions. (Sample I in Pew Research Center, 2016; Panel 2 in Pennay et al., 2018)

As The University of Maryland's Program for Public Consultation (PPC) has increased its use of opt-in sampling, it has sought to determine how the samples it obtains–given the methods it uses for stratification (quotas and weighting) and quality assurance–compare to those obtained by other firms. Thus, PPC conducted its own benchmark study. To enable comparisons, the study was modeled after the MacInnis et al. and Pew Research Center studies.

PPC fielded a national survey with a sample of 3,019 adults drawn from three large opt-in panels: Cint, Prodege, and Dynata. The survey was fielded April 1-4, 2024. It was available in both English and Spanish.

The survey included questions from the studies by MacInnis et al. (2018) and Pew Research Center (2023): 10 were what most consider primary demographics (e.g. age, race), and 25 secondary variables that include demographic and behavioral variables (e.g. marital status, employment status, volunteerism). (Appendixes A and B)

The first test was to determine how representative the sample was before applying the standard weights of primary demographics, which are used by nearly all firms in weighting their sample (e.g. age, race, education). The sample's unweighted primary demographics were compared to federal benchmarks, using the same variables and method of measuring error as used by MacInnis et al.

PPC's analyses of 10 primary demographic variables found that its opt-in sample had an overall error rate lower than all of the opt-in samples in the MacInnis et al. study. Relative to the probability-based samples PPC's sample varied depending on the methodology used relative to historical demographic changes. One methodology produced an error rate slightly higher (less than half a percentage point), while the other produced an error rate a bit lower than both probability-samples.

The next question is how well the sample, after it has been weighted, compares to federal baseline benchmarks. Because the primary variables are already incorporated into the sample through weighting, benchmarks based on secondary variables are used which include demographic (e.g. marital status) and non-demographic variables (e.g. employment). This test was modeled after Pew Research Center's 2023 study, replicating their question wording, benchmark sources, and pre-stratification and weighting schema. However, 3 of the 28 variables used in Pew's study were not included in the survey: two involved Covid-19 and their surveys were fielded at the height of the Covid-19 pandemic in 2021, and another (retirement accounts) had unreliable benchmark data according to Pew. Thus, an analysis of 25 secondary variables was conducted.

PPC's analysis of the 25 secondary demographic and non-demographic variables found that its opt-in sample had an error rate of 4.3 percentage points. Compared to the samples from Pew's study, this is lower than the other opt-in samples by 0.6 to 2.6 percentage points, and higher than the probability-based samples by 1.5 to 2.2 points.

After conducting the initial analysis, PPC found evidence that six of the variables had potentially unreliable benchmarks, due to large discrepancies between federal surveys, or between federal surveys and administrative data from federal agencies. Thus, a second analysis was done excluding those variables.

The analysis of 19 secondary variables, found an error rate for PPC's sample of 3.4 points, lower than the other opt-in samples by 0.7 to 2.1 points, and higher than the probability samples by 1.1 to 1.4 points.

One of the conclusions of the Pew study was that a key factor in lowering the quality of opt-in samples was the presence of "bogus" respondents. Pew defined as "bogus" any respondent who claimed to have received three out of four economic benefits, as well as those who answered "Yes" to at least 10 of the 16 Yes/No questions.

To test how many "bogus" respondents PPC's sample had, who had made it past the quality control measures, another analysis was conducted modeled after Pew's 2023 study. Just 4.4% of PPC's sample was classified as bogus, which is 3.6 percentage points lower than for the opt-in samples in Pew's study, on average, but 3.4 points higher than the probability-based samples, on average.

Pew also concluded that one of the biggest sources of "bogus" respondents in opt-in samples are among those who report being Hispanic or 18-29 years old. PPC's sample had substantially lower levels of "bogus" respondents than the opt-in samples in Pew's study, among those who identified as Hispanic by 12.3 points, and 18-29 years old by 9.0 points.

In conclusion, the sample that PPC developed, drawn from three opt-in panels, using nearly all of its own pre-stratification methods and quality control measures, was:

- **before weighting, in terms of primary demographic benchmarks, better than all of the opt-in samples and around the same as the probability-based samples in the MacInnis et al. study;**
- **after weighting, in terms of secondary variables (demographic and behavioral), better than all of the opt-in samples and slightly worse than the probability-based samples in the Pew study;**
- **in terms of tests to identify "bogus" respondents, better than the opt-in samples and slightly worse than the probability-based samples in the Pew study.**

Introduction

Currently, one of the biggest questions in survey research is how well probability-based sampling (PS) and non-probability-based sampling (NPS), can obtain samples that accurately represent a population. Samples obtained through PS are widely viewed as the "gold standard".

Since PPC is moving to survey populations using online panels, whose members are obtained using non-probability methods, it sought to test how well the sampling methods it uses can obtain a sample that is representative of the general population, and how that sample compares to samples obtained by other firms using NPS and PS.

A substantial number of studies have been conducted over the last two decades to test the accuracy of samples obtained using PS and NPS, by various firms. (Pasek and Krosnick, 2010; Pew Research Center, 2016 and 2023; Dutwin and Buskirk, 2017; Pennay et al., 2018; MacInnis et al., 2018; MacInnis et al., 2020).

In these studies, multiple firms using PS and NPS conducted the same survey, which included questions about key population variables. The samples' aggregate responses were then compared to federal benchmarks for those same population variables, derived from high-quality federal surveys with very large samples ranging from tens of thousands to over three million US residents and thus with very low margins of error. This included the Census' American Community Survey (ACS), Current Population Survey Annual Social and Economic Supplement (CPS ASEC), and the Voting Supplement, as well as the National Health Interview Survey. The amount that each population variable in a sample deviated from federal benchmarks was then calculated, and then an overall error rate was determined. The method of calculating the overall error rate differs depending on the study.

These studies tested the accuracy of:

- the unweighted sample on primary demographic variables
- the weighted samples on secondary and non-demographic variables

Nearly all samples obtained using NPS were less accurate along both measures than samples obtained using PS.

There are strong indications that variations in methods used for pre-stratification, completion quotas, post-stratification weighting and quality assurance can significantly impact the representativeness of samples obtained using NPS. Samples obtained from firms using NPS had a much greater variation in overall error rate, and a much higher range of errors, compared to samples obtained by PS firms. Thus, while nearly all of the NPS firms performed worse, two of the above studies found NPS firms that were more accurate than the sample(s) obtained using PS (Sample I in Pew Research Center, 2016; Panel 2 in Pennay et al., 2018).

Another factor affecting the quality of a sample is the use of quality assurance methods. Survey research has found that NPS surveys tend to suffer from higher rates of "bogus" respondents than do PS surveys. This includes respondents who answer questions dishonestly (to qualify or even to just be mischievous) or speed through the survey without reading the questions, as well as respondents who are actually bots. (Yeager et al., 2011; Griffen et al., 2021; Pew Research Center, 2022; ABC News, 2024)

Design

In order to test the accuracy of a general population sample that PPC can obtain using NPS and its own pre-stratification, completion quotas, and quality assurance methods, a study was constructed in which PPC fielded a national survey using NPS. PPC conducted three tests on the sample obtained, modeled after studies by MacInnis et al. (2018) and Pew Research Center (2023).

The first test on the sample was to determine how well the unweighted sample matched federal benchmarks for primary demographic variables. To do this, PPC used the methods employed by MacInnis et al. on the unweighted benchmarks in 2018, which compared samples obtained by firms using PS, NPS and a mix of both.

The statistical method employed to analyze primary demographic variables was the root mean squared error (RMSE) approach, which is often used to test the magnitude of errors on the unweighted data. (Yeager et al., 2011) The reasoning for this approach is that the error rates generated by RMSE are more sensitive to errors of higher magnitude, and is thus a more appropriate approach to evaluate overall representativeness.

The RMSE was obtained by computing the squared error for each measure, i.e. the square of the deviation between the sample responses and the most current benchmark baseline for each primary demographic variable. Then, the mean squared error was calculated, by taking the sum of the squared errors across variables, and dividing it by the number of measures. Lastly, the square root of the mean squared error was calculated.

The second test conducted was to determine, after weighting the sample to primary demographics, the accuracy of the sample relative to secondary variables (demographic and behavioral) according to federal benchmarks. PPC used the methods employed by Pew Research Center in 2023, which compared six samples weighted to primary variables: three obtained from PS firms and three from NPS firms. Each sample was analyzed in terms of its variation from 28 secondary variable benchmarks. The error rate for each sample was obtained by taking the average of the absolute errors for each variable.

The third test conducted was to assess the quality of the sample in terms of “bogus” respondents. This is a test of the general quality of the sample and to some extent the effect of the quality assurance methods that eliminate certain respondents. PPC replicated an analysis by Pew using their definition of a "bogus" respondent, by calculating the percentage of respondents who answered certain questions dishonestly or mischievously.

Methodology

PPC obtained a sample of 3,019 respondents for a national survey recruited from the general adult population and drawn from three large NPS panels: Cint, Prodege, and Dynata. The survey was fielded April 1-4, 2024, and was available in both English and Spanish.

PPC designed the questionnaire and programmed it online in the Alchemer platform, where respondents with a link can take the survey via computer or mobile phone. Survey responses were collected directly on the Alchemer platform, which adheres to the European Union’s General Data Protection Regulation policies for data privacy and security.

Sample collection was managed by QuantifyAI with consultation from PPC. Respondents were invited to participate via email invitation, push notification from the sample management company, or SMS among some cell phone users. Respondents were offered an incentive to participate in the survey, which averaged two dollars, in cash or an equivalent such as reward points. Incentives were administered by the panel companies. The overall response rate was 7.4%.

Consistent with industry-wide practices, PPC requested that the sample be pre-stratified to meet primary demographic benchmarks for the general adult population, replicating the schema Pew used in their 2023 study. (Appendix C) To meet the pre-stratification quotas, sample recruitment accounted for differing response rates among demographic groups, in order to reduce the errors for the final primary demographic benchmarks. Thus, demographic groups that tend to have lower response rates were over-invited, including people of color, young people, and those with lower education. Respondents were required to answer all primary demographic questions (Appendix A) so there was no missing data for these benchmarks.

Sample collection also involved the industry-wide practice of completion quotas to meet pre-stratification requirements, which disqualify respondents based on their primary demographic characteristics once the benchmark for a particular demographic has been met. Pew required that all of the firms in their 2023 study pre-stratify their samples according to the same criteria, but did not indicate whether the sample was simply pre-stratified in the sample recruitment phase, or also included completion quotas. PPC reached out to the authors of the 2023 Pew study to clarify this, but did not receive any response.

Quality control measures were included in the survey to disqualify respondents from completing the survey if they were engaging in dishonest, mischievous behavior, or simply not paying attention. This included two attention-check questions and a speed limit. Among the sample who completed the survey, 117 respondents were removed because they failed to answer at least half of the questions or provided answers to a quality control question incorrectly (e.g. answered that they had bought a private jet in the past week).

Findings

Test 1: Evaluation of Unweighted Sample Relative to Primary Demographics

The first test sought to evaluate how PPC's unweighted NPS sample performed relative to primary demographic benchmarks, obtained from the Census' ACS 2022 and CPS ASEC 2023. PPC modeled its analysis after that conducted by MacInnis et al. (2018), which used RMSE to determine each sample's overall error rate. (Yeager et al., 2011)

The MacInnis et al. analysis included in its analysis ten variables from seven primary demographics:

- Age (30-49)
- Gender (Female)
- Income ($20,000 - $49,999)
- Region (South)
- Education (High School diploma or equivalent)
- Home ownership (Yes, owned)
- Race and ethnicity:
    - (Non-Hispanic)
    - (White, including Hispanic)
    - (Not Black or African American)
    - (Not other race)

The variables used in the MacInnis et al. analysis were selected because they were the modal variable for each demographic (i.e. the largest group).

However, this method of selecting variables posed a problem for PPC's analysis, because the modal variables for two of the demographics had changed since the surveys were fielded for the MacInnis et al. study, back in 2012: income and education. Thus, PPC was faced with the dilemma of whether to use in its analysis the exact variable for each demographic variable analyzed by MacInnis et al., or to use the current modal variables for income and education. PPC decided to run both analyses.

The first analysis used the *current modal variable* for each of the demographic categories. This resulted in two of the variables being different from those used in the MacInnis et al. study: the modal variable for income changed from $20,000 to $49,999, to $150,000 or higher; and for education it changed from "High School diploma or equivalent", to "some college.".

This analysis resulted in an RMSE for PPC's sample of 4.4, which is lower than each of the samples from NPS firms by an average of 1.6 points. It is higher than both of the samples from PS firms by an average of 0.3 points.

Accuracy metrics for PS, combined, and NPS surveys of primary demographics. *Using current modal variables for the analysis of PPC's sample*											
Root Mean Squared Error											
MacInnis: Probability		MacInnis: Mixed		MacInnis: Non-probability						PPC: Non-probability
Phone	Internet	1	2	1	2	3	4	5	6	
4.3	3.9	6.0	4.9	5.9	5.5	5.0	5.7	4.8	9.0	4.4

The second analysis used the *exact variables* used by the MacInnis et al. study. This resulted in an RMSE for PPC's sample of 2.9 points, which is lower than each of the samples from NPS firms in MacInnis et al. by an average of 3.1 points. It is also lower than both of the samples from PS firms by an average of 1.2 points.

Accuracy metrics for PS, combined, and NPS surveys of primary demographics. *Using the exact variables from the MacInnis et al. analysis*										
Root Mean Squared Error										
MacInnis: Probability		MacInnis: Mixed		MacInnis: Non-probability						PPC: Non-probability
Phone	Internet	1	2	1	2	3	4	5	6	
4.3	3.9	6.0	4.9	5.9	5.5	5.0	5.7	4.8	9.0	2.9

The drop in the RMSE obtained when PPC switched from using the current modal variables to using the exact variables used by MacInnis et al. is due entirely to the change in the income variable. The income variable used in the first analysis ($150,000 or higher) had an error rate over eight points higher than the income variable used in the second analysis ($20,000 to $49,999): 11.6 vs 3.2 points. This drop in error rate was only partly offset by an increase in the error rate for the education variable: from 0.4 points for "some college" to 4.1 points for "High School diploma or equivalent" (used in the second analysis).

The reason for the large variation in error rates between the income variables is likely due to the fact that income is the only demographic by which PPC did not pre-stratify its sample. This was because PPC replicated how Pew pre-stratified their sample; and they did not pre-stratify by income.

MacInnis et al. noted that many of the survey research firms used both pre-stratification and completion quotas for their weight variables and that this could improve, but not eliminate, the accuracy of the firms' samples. PPC's pre-stratification methods, along with its quality control measures, may explain its sample's higher level of accuracy in this study, with the exception of the income variable.

Test 2: Evaluation of Weighted Sample Relative to Secondary Variables

The second test sought to evaluate how the sample, after post-stratification by primary demographic variables (i.e. weighting), performed relative to secondary demographic and non-demographic variables.

In 2023, Pew conducted a study comparing the accuracy of six weighted samples from three firms using PS and three firms using NPS, relative to 28 secondary and non-demographic variables. Each sample was pre-stratified and weighted using the same schema. The overall error rate for each sample was determined by the average absolute error, which is calculated by taking the absolute deviation of the sample's responses from the benchmark data for each variable, and then averaging those deviations.

This study found that the weighted non-probability samples had an average absolute error over twice as large as the probability-based samples (5.8 vs 2.6 points). (2023) These results differ substantially from Pew's original study from 2016, which found that non-probability samples from nine different NPS firms had an average absolute error rate just 0.3 points higher, on average, than Pew's own probability-based sample. (2016) However, the range of errors between the opt-in samples was high, with some doing substantially worse than Pew's own sample.

To test the accuracy of PPC's weighted opt-in sample relative to secondary variables, a test was modeled after Pew's 2023 study, but altered slightly by asking 25 of the 28 benchmark questions used by Pew (Appendix B).

To ensure an apples-to-apples comparison, PPC replicated Pew's:

- question wording,
- sources for benchmark data, using baselines from the most recent fieldings (Census' ACS 2022, CPS ASEC 2023 and Voting Supplement 2023; and the NHIS 2022 and 2023)
- pre- and post-stratification schema (excluding frequency of internet usage) (Appendix C),
- method for calculating the sample's error rate: the average absolute error.

PPC made the decision to remove frequency of internet usage as a post-stratification weight. Because samples from NPS firms are recruited exclusively online, and take surveys online, the percent of respondents in NPS surveys who report "rarely or never" using the internet will always be substantially less than national rates. Thus, including internet frequency in the weighting for NPS surveys results in weights as high as 30, as PPC found out when it initially included internet usage in its weighting schema. Furthermore, we have not found any other survey research firm that uses internet usage as a weighting variable for a general adult population sample.

PPC also chose not to test two variables related to Covid-19, since Pew asked these questions during the Covid-19 peak in 2021, and it was no longer a typical and persistent health issue in 2024.

PPC also excluded the variable pertaining to retirement accounts as Pew recognized that this benchmark was unreliable. Pew noted that the benchmark data likely undercounts the actual percent of adults with retirement accounts, citing analyses by the Social Security Administration (2021) and the Employee Benefits Research Institute (Copeland, 2019). Another study has also found substantial undercounting by the Census (ASPPA, 2018).

Since PPC was able to replicate Pew's analysis without these 3 variables, and recalculate their findings using their supplemental materials, we were still able to make an apples to apples comparison of the results. The analysis with all 25 benchmarks was conducted and is discussed below.

Analysis of All 25 Benchmarks

PPC's sample obtained an error rate of 4.3 percentage points for the 25 benchmarks. This is lower than the samples from NPS firms for those benchmarks by 0.6 to 2.6 percentage points, and higher than samples from PS firms by 1.5 to 2.2 points.

Accuracy metrics for PS and NPS surveys on 25 secondary and non-demographic variables.							
Average absolute error							
Pew: Probability				Pew: Non-probability			PPC: Non-probability
1	2	3		1	2	3	
2.1	2.8	2.6		6.9	6.1	4.9	4.3

Analysis of 19 Benchmarks

For reasons discussed in greater detail below, PPC also determined, after data collection, that the benchmark data for 6 of the 25 variables were unreliable due to inconsistencies between various federal benchmark surveys and administrative data. Thus, this analysis excludes these 6 variables and was conducted with 19 of the 25 benchmarks.

To make an apples to apples comparison, PPC ensured that the other samples were adjusted accordingly. Using Pew's supplementary materials the samples were recalculated to produce the Average Absolute Errors of the 19 benchmarks for each of the six samples.

Among the 19 secondary and non-demographic variables, PPC's sample had an average absolute error rate of 3.4 percentage points. This is lower than the three samples from NPS firms for those benchmarks by 0.7 to 2.1 points, and higher than the samples from PS firms by 1.1 to 1.4 points.

Accuracy metrics for PS and NPS surveys on 19 secondary and non-demographic variables.							
Average absolute error							
Pew: Probability				Pew: Non-probability			PPC: Non-probability
1	2	3		1	2	3	
2.0	2.3	2.2		5.5	4.6	4.1	3.4

Rationale for Excluding Six Secondary Variables

PPC's alternate analysis of secondary and non-demographic variables excluded 6 of the 25 variables included in the survey, due to concerns that the benchmark baselines for each were unreliable, due to inconsistencies between federal surveys, or between federal surveys and administrative data from federal agencies.

The first three variables excluded were for participation in means-tested government programs: household SNAP participation, unemployment compensation, and Social Security. All samples–probability as well as non probability–showed substantially higher rates of participation than that found by the Census' CPS ASEC, which was the benchmark used by Pew in their analysis.

This led PPC to investigate those variables more closely and ultimately conclude that the problem may not be with the samples obtained, but with the Census benchmark data for those variables. Numerous studies have found that the Census, and other large federal surveys, substantially undercount participation in means-tested government benefit programs. A report authored by two Census Bureau researchers acknowledged this undercounting, stating that, "a large body of research has documented substantial measurement error in survey reports of income and program participation[...] Fox et al. (2017), Shantz and Fox (2018), Meyer and Mittag (2018), Meyer et al. (2018), among others, have shown high rates of under-reporting of the receipt of means-tested program benefits, including in the CPS ASEC, the Survey of Income and Program Participation (SIPP), and the American Community Survey (ACS)." (Rothbaum et al., 2022)

The problem of undercounting means-tested program participation has existed for at least two decades, as uncovered by studies on participation in SNAP (Parker, 2011), Medicaid (Pascale et al., 2009), and the full range of programs including SSI, TANF, SNAP and Medicaid. (Meyer et al., 2009; Wheaton, 2009) A recent study by the Federal Reserve researchers found that the Census's CPS ASEC substantially underestimated the number receiving unemployment insurance. (Larrimore, et al., 2023)

Looking at the most recent SNAP participation data from the CPS ASEC, for which administrative data exists, PPC confirmed that undercounting is still occurring: The USDA reported that 17.2% of households participated in SNAP in 2023 (USDA, 2024), while the CPS ASEC 2023 found 10.6% (var. HFOODSP)

The other three variables excluded from the main analysis were e-cigarette/vape usage, cigarette usage, and food allergies. PPC found large discrepancies in baselines for these variables between different federal surveys.

Rates of e-cigarette usage differed by a factor of up to three between federal surveys. The CDC's National Health Interview Survey (NHIS), which was the benchmark used in Pew's study, found that 16.7% report ever having used an e-cigarette (var. ECIGEV_A); SAMHSA's National Survey on Drug Use and Health (NSDUH) found 24.4% have used an e-cigarette in their lifetime (Table 2.2b); and the CPS Tobacco Use Supplement found just 8.4% (var. PEJ1A3_5)

Rates of cigarette smoking also differ substantially between surveys. The NHIS 2022 found 64.2% reported never smoking 100 cigarettes (SMKEV_A), while the CPS Tobacco Supplement found 74.5%. (var. PEA1)

Rates of food allergy prevalence differ by a factor of about 1.5. The CDC's NHIS found that 6.2% of adults have a food allergy (Zablotsky et al., 2023), while the NIH's National Institute of Allergy and Infectious Diseases found 10.8% (Gupta et al, 2019).

Because administrative data does not exist for these variables, no conclusion can be made about the true baselines. For this reason, PPC excluded these variables.

Test 3: Prevalence of "Bogus" Respondents

Another major concern in online survey research is the potential for "bogus respondents," generally defined as respondents who respond without reading the questions, who answer dishonestly, or who are not real respondents but rather bots (Griffin et al., 2021; Cornesse and Blom, 2023; ABC News, 2024).

A comprehensive study by Pew Research Center found that opt-in panels tend to have at least four times as many "bogus" respondents as probability-based samples. (2020). In evaluating the samples of NPS firms a key question is how well their quality control measures eliminate or reduce "bogus" respondents.

PPC Standard Quality Control Methods for Eliminating "bogus" Respondents

PPC includes in its surveys a number of methods for disqualifying "bogus" respondents which were applied to the sample for this study. These include questions to determine whether respondents are paying attention, questions to determine whether they are responding dishonestly ("mischievous respondents", see Kurtz, 2014; Cimpian et al., 2018; Pew Research Center, 2022), and speed limits. These are all common practices among many survey research firms.

PPC disqualified respondents who answered the following attention-check questions incorrectly: "Please select the option that does not fit [Tuesday, Friday, April, Sunday]," and "Please select the option *somewhat convincing* [Somewhat unconvincing, Very convincing; Somewhat convincing; Very unconvincing]

PPC also asked respondents, “Which of the following did you do in the past week? Select all that apply” and presented four unrealistic options (e.g. “purchased a private jet”), two realistic options (e.g. “watched TV”) and a “none of the above” option. This question was taken from a Pew study of “bogus” respondents among non-probability samples (2022). Respondents were disqualified for choosing any of the unrealistic options.

Lastly, PPC disqualified respondents that completed the first 15 questions within 60 seconds, which indicates they were speeding through the survey.

Tests of Effectiveness for Eliminating "Bogus" Respondents

Despite the quality control methods used to disqualify “bogus” respondents within the survey, it is always possible that some "bogus" respondents completed the survey. To test how well PPC’s quality control methods worked, two analyses were conducted, replicating measures of suspected "bogus" respondents used by Pew in their 2023 study.

One analysis looked at those who reported receiving at least three out of four economic benefits: Social Security, food stamps (SNAP), unemployment compensation, and workers’ compensation. The rationale for the analysis, according to Pew, is that few respondents meet this criteria. According to the Census’ CPS ASEC only 0.1% receive at least three of those benefits. Thus, nearly all respondents who meet this criteria are not answering honestly. As Pew notes, “it is highly unlikely that the few individuals who do fit that description are massively overrepresented in online opt-in samples.” (2023)

In Pew’s analysis of six weighted samples, three from NPS firms and three from PS firms, they found in the non-probability samples a range of 6 to 9% of respondents claiming to receive at least three of the four economic benefits. In comparison, just 1% met this criteria in the three probability-based surveys. PPC replicated this analysis on its sample and found that 3.2% of the total sample met this criteria.

Pew determined that the higher rates in the non-probability samples were primarily driven by respondents who reported being Hispanic (19% met this criteria) or 18-29 year olds (15%). In the PPC study these numbers were lower–5.6% of Hispanics and 3.8% of 18-29 year olds.

Percentage that reported receiving at least 3 out 4 economic benefits

	Pew: Range of PS	Pew: Range of NPS	PPC: NPS
All	1 to 2%	6 to 9%	3.2%
Hispanic	1 to 2%	16 to 19%	5.6%
18-29	1 to 2%	15%	3.8%

Pew conducted another analysis, which defined "bogus" respondents as those that answered Yes to at least 10 of the 16 Yes/No secondary and non-demographic questions found in the questionnaire. These included the initial 4 economic benefits already tested, as well as certain behaviors (such as smoking, volunteering, and vaping), medical issues (such as hypertension and allergies), work status (such as having a job (3 questions) and union membership), speaking another language, citizenship, and whether the respondent is a parent.

Pew's rationale is that those answering Yes to so many questions are more likely to be engaged in straight-lining or satisficing, both of which are indications that respondents are not answering honestly.

Pew found that the non-probability samples had an average of 8% of respondents meeting that criteria, while the probability-based samples had an average of just 1%. Again, Pew determined that the higher rate among the NPS surveys was driven by Hispanics (19%) and 18-29 year olds (15%).

PPC replicated this analysis and found that 4.4% of its sample answered met that criteria, including 6.7% of Hispanics and 6.0% of those 18-29.

Percentage that said Yes to 10 or more of the 16 Yes/No questions			
	Pew: Average of PS	Pew: Average of NPS	PPC
All	1%	8%	4.4%
Hispanic	1%	19%	6.7%
18-29	2%	15%	6%

The analysis of the four economic benefits is likely a better measure of "bogus" activity among respondents than the analysis of all of the Yes/No secondary variables. While it is unlikely for a person to say Yes to at least 10 of the 16 Yes/No questions asked, it's not improbable enough to give proof they were not answering honestly. Enough honest respondents can meet that criteria and get illegitimately labeled as "bogus", to make the analysis unreliable as a measure of "bogus" respondents. However, it is nearly impossible for a person to have received at least three out of those four benefits within a year, as demonstrated by the Census' CPS ASEC finding just 0.1% meeting this criteria. Thus, very few non-bogus respondents are likely swept up in this criteria, making it the better measure.

Appendix A: Primary Demographic Variables

Source	Demo	Category comparison	Bench-mark	Unweighted percent	Absolute error
ACS 2022	Age	30-49	33.2	32.2	1.0
ACS 2022	Gender	Female	50.8	52.8	2.0
CPS 2022	Region	South	38.4	39.7	1.3
ACS 2022	Race/Ethnicity	Non-hispanic	82.7	83.2	0.5
ACS 2022	Race/Ethnicity	White, incl. Hispanic	73.3	76.7	3.4
ACS 2022	Race/Ethnicity	Not Black/African America	88.2	88.1	0.1
ACS 2022	Race/Ethnicity	Not other race	90.2	95.4	5.2
ACS 2022	Education	Some College*	29.4	29.0	0.4
ACS 2022	Education	High School degree**	27.2	31.5	4.3
CPS ASCE 2023	Household Income	More than $150,000*	21.1	9.5	11.6
CPS ASCE 2023	Household Income	$20,000 - $49,000**	22	24.9	3.2
ACS 2022	Home ownership	Own home	69.2	66.1	3.1

*Used in first analysis representing the highest modality for each category in the measure for 2022. This was different from the 2012 study.
**Used in secondary analysis, since it represented an exact replication of the measure tested by MacInnes using data from 2012.

Appendix B: Secondary Variables

Source	Question	Combined Response Options	Benchmark	Weighted percent	Average Error
FEC	Did you vote in the 2020 election?	Yes	66.2	70.6	-4.4
		No	33.8	29.5	
		Avg Abs Error			4.4
FEC	[If voted] Who did you vote for?	Donald Trump	46.9	46.2	0.7
		Joe Biden	51.3	51.5	-0.2
		Jo Jorgensen, Libertarian	1.2	1.0	0.2
		Howie Hawkins, Green Party	0.3	0.6	-0.3
		Another candidate	0.4	0.8	-0.4
		Avg Abs Error			0.4
NHIS 2023	Health Insurance	Yes	93.3	89.2	4.1
		No	6.6	10.6	
NHIS 2022	Smoked 100 cigarettes? IF "Yes" how often?	No	64.2	59.9	4.3
		Every day	8.6	16.9	-8.3
		Some days	2.7	6.6	-3.9
		Not at all	21.7	18.3	3.4
		Avg Abs Error			5.0
NHIS 2022	E-cigarettes or vape? IF "Yes" how often?	No	78.2	64.9	13.3
		Every day	3.1	7.7	-4.6
		Some days	2.7	13.2	-10.5
		Not at all	13.3	14	-0.7
		Avg Abs Error			7.3
NHIS 2022	Diagnosed with hypertension?	Yes	32.0	35.8	-3.8
		No	67.9	63.6	
		Avg Abs Error			3.8
NHIS 2021	Diagnosed with food allergy?	Yes	6.2	13.8	-7.6
		No	93.8	85.8	
		Avg Abs Error			7.6

	Last week, work for pay or profit? Ever absent?	Yes	59.7	50.9	8.8
		Absent at least once	1.8	4.7	-2.9
		Did not work last week	38.5	44.2	-5.7
		Avg Abs Error			5.8
CPS Mar ASEC 2023	Union membership?	Yes	5.5	10.2	-4.7
		No	94.5	89.4	
		Avg Abs Error			4.7
ACS 2022	Ever served in the military?	Never served	92.3	88.9	3.4
		Only active duty for training: Reserves or National Guard	1.1	1.3	-0.2
		Now on active duty	0.5	0.9	-0.4
		Not currently on active duty	6.1	8.5	-2.4
		Avg Abs Error			1.6
ACS 2022	Type of home?	A mobile home	5.1	5.6	-0.5
		A one-family detached housing	67.2	60.5	6.7
		A one-family house attached to one or more houses	6.2	9.3	-3.1
		A building with 2 or more apartments	21.4	23.6	-2.2
		Boat, RV, van, etc.	0.1	0.6	-0.5
		Avg Abs Error			2.6
ACS 2022	Home ownership?	Own w/mortgage or loan?	43.6	35.6	8.0
		Own free and clear (w/o mortgage or loan)	25.6	25.4	0.2
		Rented	29.4	37	-7.6
		Occupied w/o payment	1.5	1.8	-0.3
		Avg Abs Error			4.0
	Where did I live 1 year ago?	This house	92.2	89.3	2.9

CPS ASEC 2023		Different house in the U.S. or Puerto Rico	7.4	9.7	-2.3
		No, outside U.S.	0.4	0.7	-0.3
		Avg Abs Error			1.8
ACS 2022	Are you a US citizen?	Yes	92.4	95.0	-2.6
		No		5.0	
		Avg Abs Error			2.6
ACS 2022	Number of vehicles in HH?	None	6.2	13.2	-7.0
		1	23.9	38.5	-14.6
		2	38.7	32.9	5.8
		3	19.2	11	8.2
		4	8.0	2.8	5.2
		5	2.7	1.2	1.5
		6 or more	1.4	0.4	1.0
		Avg Abs Error			6.2
ACS 2022	Speak other languages than English? If Yes, how well do you speak English?	No	78.0	79.6	-1.6
		Very well	12.9		
		Well	4.4	15.9	-3.0
		Not well	3.2	3.8	0.6
		Not at all	1.5	0.5	2.7
		Avg Abs Error		0.1	1.4
Pew's NPORS 2023	Number of people live in the Household? How many are adults?	1	17.0	20.7	-3.7
		2	49.9	33.9	16.0
		3-4	27.4	34.5	-7.1
		5+	3.9	10.8	-6.9
		Avg Abs Error			8.4
ACS 2022	Parent/guardian child under 18?	Yes	25.8	25.3	0.5
		No	74.2		
ACS 2022	How many are children?	No children	72.8	67	5.8
		1 child	12.2	15.4	-3.2
		2 children	9.7	11.6	-1.9
		3-4 children	4.9	5.5	-0.6
		5 or more children	1.6	0.4	1.2
		Avg Abs Error			2.5
	Did you work during 2022?	Yes	64.8	62.2	2.6

CPS ASEC March 2023		No	35.2		
CPS ASEC March 2023	unemployment compensation?	Yes	3.8	7.7	-3.9
		No	96.2	92.1	
		Avg Abs Error			3.9
CPS ASEC March 2023	Worker's Comp?	Yes	0.4	4.2	3.8
		No	99.6	95.7	
		Avg Abs Error			3.8
CPS ASEC Mar 2023	SS payments?	Yes	21.7	28.2	-6.5
		No	78.3	71.5	
		Avg Abs Error			6.5
CPS ASEC Mar 2023	SNAP benefits	Yes	10.6	23.9	-13.3
		No	89.4	75.7	
		Avg Abs Error			13.3

Appendix C: Stratification

Pre-Stratification Quotas

Age x Gender	Men 18-44		23.5
	Women 18-44		22.7
	Men 45-64		15.7
	Women 45-64		15.9
	Men 65+		9.9
	Women 65+		12.2
Race	White/Other/2+ races, non-Hispanic		71.0
	Black NH		11.8
	Hispanic		17.3
Education	HS or less		37.5
	Some college		29.4
	College or more		33.1

Post-Stratification Weighting

Age x Gender	18-29 - Men (1)	51.1%	10.5
	18-29 - Women (2)	48.9%	10.0
	30-49 - Men (3)	50.1%	16.8
	30-49 - Women (4)	49.4%	16.4
	50-64 - Men (5)	49.5%	11.9
	50-64 - Women (6)	50.1%	12.2
	65 and over - Men (7)	44.9%	10.0
	65 and over - Women (8)	55.1%	12.2
Education x Gender	HS or less - Men (1)		19.8
	HS or less - Women (2)		17.9
	Some college - Men (3)		13.8
	Some college - Women (4)		15.5
	College or more - Men (5)		15.5
	College or more - Women (6)		17.5
Education x Age	HS or less - 18-29 (1)		8.4

	HS or less - 30-49 (2)		10.8
	HS or less - 50-64 (3)		9.4
	HS or less - 65+ (4)		8.9
	Some college - 18-29 (5)		7.2
	Some college - 30-49 (6)		9.3
	Some college - 50-64 (7)		6.9
	Some college - 65+ (8)		5.9
	College or more - 18-29 (9)		4.9
	College or more - 30-49 (10)		13.5
	College or more - 50-64 (11)		8.0
	College or more - 65+ (12)		6.7
Race/Ethnicity x Education	White NH - HS or less (1)	33.5%	20.5
	White NH - Some college (2)	27.3%	16.7
	White NH - College or more (3)	39.2%	24.0
	Black NH - HS or less (4)	45.3%	5.3
	Black NH - Some college (5)	29.6%	3.5
	Black NH - College or more (6)	25.0%	2.9
	Hispanic - HS or less (7)	57.7%	10.0
	Hispanic - Some college (8)	23.6%	4.1
	Hispanic - College or more (9)	18.7%	3.2
	Asian NH - HS or less (10)	26.4%	1.6
	Asian NH - Some college (11)	17.6%	1.1
	Asian NH - College or more (12)	56.0%	3.5
	Other NH - HS or less (13)	36.0%	1.3
	Other NH - Some college (14)	30.0%	1.2
	Other NH - College or more (15)	32.2%	1.1
AMONG HISPANIC AND ASIANS: Born in US -- 50 states, DC (excludes Puerto Rico and other territories)	Born inside the US [50 states, DC] (1)	44.2%	9.8
	Born outside the US [incl. Puerto Rico, other US territories] (2)	55.8%	12.3
Years lived in the US	Born in the US		81.6
	Lived in the US:		
	0-10 years		3.9

	11-20 years		3.7
	21+ years		10.8
Region x Metro/Non-metro	Northeast-Metro (1)	92.7%	16.1
	Northeast-Non-metro (2)	7.3%	1.3
	Midwest-Metro (3)	79.5%	16.4
	Midwest-Non-metro (4)	20.5%	4.2
	South-metro (5)	86.2%	33.1
	South-non-metro (6)	13.8%	5.3
	West-metro (7)	91.8%	21.6
	West-non-metro (8)	8.2%	1.9
Volunteerism	Yes		21.0
	No		79.0
Registered to vote	Yes		69%
	No		
Frequency of Internet Usage	Almost constantly		41.7
	Several times a day		43.9
	Daily or several times a week		8.1
	Less often or does not use the internet		6.3
Religious Affiliation	Protestant		41.3
	Catholic		20.7
	Unaffiliated		29.0
	Other		9.0
Partisan affiliation	Republican/Lean Republican		45.0
	Democrat/Lean Democrat		47.0
	Independent (DK/Refused/No lean)		8.0

Bibliography

Cimpian, J. R. et al. (2018). *Bias from potentially mischievous responders on large-scale estimates of lesbian, gay, bisexual, or questioning (LGBQ)–heterosexual youth health disparities*. American journal of public health, 108(S4), S258-S265.

Copeland, C. (2019, May 30). *Current Population Survey: Checking in on the Retirement Plan Participation and Retiree Income Estimates*. Employee Benefit Research Institute.

Cornesse, C., & Blom, A. G. (2023). *Response quality in nonprobability and probability-based online panels*. Sociological Methods & Research, 52(2), 879-908.

Dutwin, D., & Buskirk, T. D. (2017). Apples to Oranges or Gala versus Golden Delicious? *Public Opinion Quarterly*, *81*(S1), 213–239.

Griffin, M., Martino, R. J., LoSchiavo, C., Comer-Carruthers, C., Krause, K. D., Stults, C. B., & Halkitis, P. N. (2021). Ensuring survey research data integrity in the era of internet bots. *Quality and Quantity*, *56*(4), 2841–2852.

Gupta, R. S. et al. (2019). *Prevalence and severity of food allergies among US adults*. JAMA network open, 2(1), e185630-e185630.

Improving the measurement of retirement income of the aged population. (2021). Social Security Administration Research, Statistics, and Policy Analysis.

Is Retirement Plan Participation Underreported? (2018). American Society of Pension Professionals and Actuaries.

Kennedy, C. et al. (2016, May 2). *Evaluating online nonprobability surveys*. Pew Research Center.

Kennedy, C. (2022). *Exploring the Assumption That Online Opt-in Respondents Are Answering in Good Faith*. Pew Research Center.

MacInnis, B., Krosnick, J. A., Ho, A. S., & Cho, M. (2018). The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension. *Public Opinion Quarterly*, *82*(4), 707–744.

Cornesse, C., Krosnick, J. A., Blom, A. G., Dutwin, D., De Leeuw, E. D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., & Wenz, A. (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology*, *8*(1), 4–36

Kurtz, H. (2020, November 19). *"Mischievous responders" may skew Survey-Based studies of teens*. Education Week.

Larrimore, J., Mortenson, J., & Splinter, D. (2023). *Unemployment insurance in survey and administrative data*. Journal of Policy Analysis and Management, 42(2), 571-579.

Mercer, A. and Lau, A. (2023, September 7). *Comparing two types of online survey samples*. Pew Research Center.

Meyer, B. D., Mok, W. K., & Sullivan, J. X. (2009). *The under-reporting of transfers in household surveys: Its nature and consequences* (No. w15181). National Bureau of Economic Research.

Mitchell, T., & Mitchell, T. (2024, April 14). *Assessing the risks to online polls from bogus respondents*. Pew Research Center.

Parker, J. (2011). *SNAP misreporting on the CPS: Does it affect poverty estimates?*. Social, Economic, and Housing Statistics Division US Census Bureau, Working-Paper Number: 2012-1.

Pascale, J. et al. (2009). Medicaid Underreporting in the CPS: Results from a Record Check Study. *Public Opinion Quarterly*, *73*(3), 497–520.

Pennay, D., Neiger, D., Lavrakas, P., & Borg, K. (2018). *The Online Panels Benchmarking Study: A Total Survey Error Comparison of Findings from Probability-Based Surveys and Non-Probability Online Panel Surveys in Australia*. Monash University

Rogers, K. (2024, April 19). *How pollsters are (and aren't) fighting fraud*. ABC News.

Rothbaum, J., et al. (2022, May). *Fixing Errors in a SNAP: Addressing SNAP Underreporting to Evaluate Poverty*. In AEA Papers and Proceedings (Vol. 112, pp. 330-334).

Wheaton, L. (2008, February 6). *Underreporting of Means-Tested transfer programs in the CPS and SIPP*. Urban Institute.

Yeager, D. et al. (2011). *Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Nonprobability Samples*. Public Opinion Quarterly 75:709–47.

Zablotsky, B., Black, L. I., & Akinbami, L. J. (2023). *Diagnosed allergic conditions in children aged 0–17 years: United States, 2021*. Centers for Disease Control.