Worker Classification Knowledge Survey

Volume II - Methodology Report

11/16/2016

Submitted to:
Jonathan Simonetta
U.S. Department of Labor
200 Constitution Avenue, NW
Washington, DC 20210

Submitted by:
Kelly Daley
Stas Kolenikov
Marci Schalk
Mehera Baugher
Julie Pacer
Jacob Klerman
Lauren Dunton

Abt Associates
55 Wheeler Street
Cambridge, MA 02138

1. Overview

This report describes the methods used by Abt Associates in conducting the Worker Classification Knowledge Survey for the Department of Labor (DOL GS10F0086K). The Worker Classification Knowledge Survey is a dual-frame telephone survey that measures American workers' knowledge about their current job classification and their knowledge about the rights and benefits associated with their job status. We conducted this study in support of the Department of Labor's Misclassification Initiative and its efforts to promote fair hiring practices and access to critical workplace benefits, opportunities, and protections.

Drawing samples from both landline and cell phone random-digit dialing (RDD) frames, the survey obtained interviews with a nationally representative sample of 8,503 American workers. Interviews were conducted from August 26, 2014, to March 9, 2015, in English and Spanish. The balance of this report contains five sections. In Section 2, we describe the sample design, including the target population and respondent selection processes. Section 3 summarizes the questionnaire development and pilot survey process. Section 4 describes the data collection protocol, Section 5 describes the weighting procedures, and Section 6 provides final disposition and response rates. Section 7 analyzes nonresponse.

2. Sample Design

This section discusses the sample design, including the target population, sampling methodology, and respondent selection process.

2.1 Population

The target population for the Worker Classification Knowledge Survey was persons age 18 and older living in the United States who have a telephone (landline or cellular) and who had worked for pay or profit at any time in the 30 days before the interview. According to the 2015 National Health Interview Survey (NHIS), 96.9 percent of U.S. adults live in a household with landline or cellular telephone service (Blumberg and Luke 2015). An important focus of the study was determining what workers know about their current employment status and the rights and benefits associated with that status. The "last 30 days" reference period was used to ensure that the survey sufficiently captured recently employed workers. Given economic conditions at the time of survey fielding, there was concern that a "did work for pay last week" reference period would potentially exclude workers of interest for the analysis. By having a longer look-back period, the survey is designed to capture a larger universe of adults who were employed within the past 30 days, rather than the somewhat smaller universe of adults employed only during the previous week.

2.2 Sample Design

The survey featured a national, overlapping dual-frame landline and cell phone RDD design. We completed 2,554 interviews with respondents sampled through the landline frame, and 5,949 interviews with respondents sampled through the cell phone frame, for a total of 8,503 interviews. Numbers for the landline sample were drawn with equal probabilities of selection from active blocks (area code + exchange + two-digit block number) that contain one or more residential directory listings.

The cell phone sample was drawn through a systematic sampling from 1,000 blocks dedicated to cellular service according to the Telcordia database. Survey Sampling International, LLC, provided the landline and cell phone samples according to our specifications.

The cell phone sample design leveraged "recent activity" flags to improve cost-effectiveness. A recent activity flag is a data field appended to each cell phone sample record that indicates whether that number is working (active) or non-working (inactive) based on a real-time database query of telephone records. The recent activity flags allow the survey researcher to identify many nonworking numbers in the selected sample and to remove them before interviewing begins. Leveraging recent activity flags is fast becoming standard practice in RDD surveys because it helps to reduce the amount of time interviewers spend manually dialing nonworking numbers (Dutwin and Malarek 2014, Pew Research Center 2015).

Specifically, the survey used the Marketing Systems Group's Cell-WINS flag. While the accuracy of the Cell-WINS flag is very high, it is not perfect (Dutwin and Malarek 2014), since purging all flagged-inactive numbers from the sample could potentially reduce the population coverage provided by the design. To realize some of the efficiency offered by this technology without reducing coverage, we appended the flag to the cell phone sample and then subsampled the flagged-inactive numbers for interviewing. From August 26 to September 15, 2014, the subsampling rate was 50 percent, and from September 16, 2014, to March 9, 2015, the subsampling rate was reduced to 30 percent. Recent empirical research on Cell-WINS flags has shown that the flags are over 87 percent accurate and that 2 percent or fewer interviews in a normal cell RDD sample are with cases erroneously flagged as inactive (Dutwin and Malarek 2014; Schalk et al. 2015). The cases that have been flagged as inactive are eventually weighted by the inverse of the subsampling rate to maintain an unbiased design. A more detailed discussion of weighting occurs in Section 5 of this document.

2.3 Respondent Screening, Eligibility, and Selection

In the survey screener, interviewers determined whether the household contained at least one eligible adult. An eligible adult was defined as a person 18 years of age or older who had done any work for pay or profit during the previous 30 days. Households reporting no eligible adults screened out as ineligible for the interview. If multiple adults in the household were eligible, a systematic selection procedure was used to select one respondent for the extended interview.

The selection procedure identified for this purpose is a modified version of the method presented in Rizzo, Brick, and Park (2004). Their method employs both probability and quasi-probability selection techniques. To sample one adult in two-adult households, this method randomly selects either the person who completed the screener (i.e., screener respondent) or the other adult.

To select a respondent in a household with more than two adults, the method draws on the “last birthday” and “next birthday” methods (based on Salmon and Nichols 1983). In these methods, the screener respondent is asked which adult in the household had the most recent (last) birthday or has the next (upcoming) birthday, and that person is sampled. Both birthday methods assume a lack of correlation between date of birth and characteristics of the person.

Rizzo et al.’s quasi-random “birthday” method is used to first identify whether the screener respondent is actually the sampled person; otherwise, the other birthday-selected adult is then pursued. By employing the 50/50, non-enumeration selection for two-adult households, fewer steps are used than in other methods, by leveraging the fact that the large majority of households in the United States have two or fewer adults. This method also avoids asking numerous enumeration questions, which some consider invasive and which can lead to higher non-cooperation. The procedure was shown to perform well for the Health Information National Trends Survey (HINTS), a large RDD survey sponsored by the National Cancer Institute (Cantor et al. 2009).

In implementing this selection procedure for the Worker Classification Knowledge Survey, it was necessary to modify it because eligibility is contingent on working status as well as adult age. In other words, even if the respondent is eligible to complete the screener (because he or she is of adult age), that person is not necessarily eligible for the extended interview (because he or she may not have worked within the last 30 days). We addressed this issue by adjusting the screener to collect all of the necessary information.

Another advantage of the Rizzo, Brick, and Park procedure is that it reduces the potential for selection error associated with the last birthday method by limiting its use to a smaller number of cases. We further reduced problems with this approach by randomizing the use of the “last birthday” and the alternative “next birthday” selection procedure. The next/last birthday approach was implemented only for households with three or more eligible adult workers *or* two eligible adult workers, neither of whom was the screener respondent. This sampling method ensures that each eligible adult identified in the screener will have the same probability of being selected for the extended interview.

If the selected respondent was not the adult who responded to the screener, the interviewer asked to speak with the selected respondent in order to administer the interview. If the selected respondent was present and available, the screener respondent handed off the phone. If such a handoff was not possible, the interviewer asked for the date and time of day when the selected respondent would be available. Interviewers also inquired as to the best phone number to use to reach the selected respondent.

While *within*-household selection and resulting handoffs are quite common in landline surveys, they are less likely when dialing cell phones. Traditionally, residential landlines have been viewed as a point of contact for the entire household. Cell phones, by contrast, are commonly viewed as personal devices, though some sharing does occur. Studies have demonstrated that within-household selection procedures can be implemented for cell phone samples though, not surprisingly, response rates are lower when the researchers want the screened person to hand off to another person in the household (American Association for Public Opinion Research (AAPOR) Cell Phone Task Force 2010; Brick et al. 2007). No household respondent-selection method was used with cell phone cases. Interviewers conducted cell phone interviews only with the respondent who answered, when he or she was eligible.

3. Questionnaire Development and Pilot Survey

The questionnaire development process began by conducting background research on key issues and legal standards that shaped the context for the Worker Classification Knowledge Survey. This included focus groups and in-depth interviews with Department of Labor staff. Specifically, the survey development process proceeded in four main phases:

1. Research of relevant survey questions
2. Review and feedback from Technical Working Group (TWG)
3. Cognitive testing
4. Pilot testing

3.1 Research of Relevant Survey Questions

To the extent possible, the Worker Classification Knowledge Survey used questions from previously tested and fielded surveys. These include the Contingent Worker Supplement to the Current Population Survey, The American Time Use Survey, the 2012 Family Medical Leave Act Survey, and the Working Without Laws Survey. This strategy potentially allows comparability with previously collected data. To ensure that new questions adequately capture the range of issues and experiences related to workers' knowledge and classification, we gathered information from various sources recommended by the Department of Labor Wage and Hour Division. These sources included survey research on related topics, state reports on worker classification issues, and descriptions of worker classification tests conducted by state and federal agencies. (The final questionnaire is included as an appendix in Volume I – Technical Report.)

3.2 Literature Reviews, Interviews, and Listening Sessions

Much of the current writing on employment classification is taking place in "gray" literature (e.g., relevant newspaper articles, policy papers), which we reviewed in addition to the peer-reviewed published literature.

We conducted in-person interviews with staff at the Department of Labor to discuss trends in employment classification across industries. These conversations were particularly important for defining legal and other regulatory terms used during the survey. Through these interviews we also gathered stakeholder (employer and worker) perspectives on the motivations for and impact of misclassification that helped to inform the design of two listening sessions with employer and worker representatives. During the listening sessions, we explained the purpose of the research and asked participants to explain, from the perspective of their constituents, what information would be important to gather in this survey and related qualitative research. The listening sessions included nine employee and four employer stakeholder organizations. We targeted our recruitment toward industries in which rates of misclassification are generally thought to be higher, based on guidance provided by DOL. We also included representatives from intermediary employment agencies, such as professional staffing services.

3.3 Technical Working Group Review and Feedback

We engaged a technical working group (TWG) composed of five legal and economic scholars in the areas of employment law and labor economics. The group also included a former DOL employee, an expert in wage and hour law. This group reviewed early versions of the study design and provided written feedback. We convened the group to discuss their ideas about the sample design and study approach. The TWG also reviewed the questionnaire drafts in detail, providing comments and edits on multiple versions.

3.4 Cognitive Testing

Cognitive interviewing is a qualitative pre-testing method used to identify and analyze potential sources of response error in survey questionnaires. This method concentrates on how respondents cognitively process information to form their responses to survey questions. It is an important step in ensuring that respondents understand the question as it was intended by the researcher (Collins 2003). To ensure respondent understanding, we cognitively tested the Worker Classification Knowledge Survey across a range of different types of workers, using purposive sampling and quota-based recruitment.

To recruit respondents for interviews we placed online advertisements on Craigslist inviting workers to participate in a one-hour interview in our Chicago office. When potential respondents called to learn more information, we asked a number of screening questions including gender, employment status (employed currently and whether they were employed by a company or self-employed), industry, occupation, and age. We selected nine respondents including a mix of men and women, self-employed individuals, and workers in industries with higher rates of misclassification.

Cognitive testing took place the week of May 7, 2012. In total, we completed nine cognitive interviews. The respondents are described in Exhibit 3.1 below.

Exhibit 3.1: Description of Cognitive Interview Respondents

Respondent	Sex	Status	Industry or occupation
1	F	Employee	Temp agency
2	F	Employee	Cleaning (maid)
3	M	Employee	Insurance
4	M	Employee	Food service (dishwasher)
5	F	Self-employed	Consultant
6	F	Employee	Security
7	M	Self-employed	Bookkeeper
8	M	Employee	Construction (wall board installer)
9	M	Self-employed	Construction (mason)

Cognitive interviews were conducted in person, using a main interviewer and a secondary interviewer, who recorded answers and took notes. Response challenges to individual questions were noted. For these questions, alternative wording was proposed and accepted by the Department of Labor. For the knowledge questions we noted some hesitance from respondents, and during debriefing interviews learned that on occasion they felt unsure about whether or not they were providing the “correct’ answer. Based upon this feedback, we made some adjustments in both wording and organization.

3.5 Pilot Testing

The Worker Classification Knowledge Survey pilot included 30 completed RDD landline telephone interviews and 30 completed RDD cell phone interviews conducted from June 24 to July 10, 2014.

The pretest sample design allocated 40 percent to the cell phone sample and 60 percent to the landline sample. The original proposal for this study was developed in 2012, and in that plan we had allocated the sample to best balance cost and design considerations, given best estimates of cell phone prevalence at the projected time of fielding. This allocation also met, if not exceeded, the survey industry standard at the time of the original proposal for methodological rigor in terms of the share of interviews allocated to the cell phone sample.[1] While only a few years had elapsed since the original proposal, the shift away from landlines to cell phones had accelerated. Specifically, over this period, the proportion of landline numbers that are working and in-service has plummeted, from 52 percent to 36 percent (Dutwin 2014). The results from the Worker Classification pretest were consistent with—and actually more extreme than—these estimates based on general RDD (random digit dial) surveys.

The pretest included a household roster on both landline and cell frames, requiring the respondent to hand the phone to another person in the household if the roster respondent was not selected for the interview. At the design stage, we thought that this hand-off protocol would potentially increase productivity in the cell frame, by reaching the working spouse if the screener respondent was not employed in the cell frame. However, as discussed above, this assumption was not supported by the pretest results, with very few successful hand-offs among cell phones. Instead of screening potential household members for participation in the survey, we recommended removing the household roster and respondent selection (i.e., the “hand-off”) from the cell phone protocol, and instead interviewing the person who answers the cell phone (assuming that person is eligible).

Upon completion of the pilot survey, a report was prepared summarizing the findings. This analysis highlighted the need to clarify the response categories for a question related to deductions and a question asking for reasons why respondents’ jobs are not covered under certain laws. Based on this analysis the necessary changes in the survey instrument were proposed and approved before launching the main study.

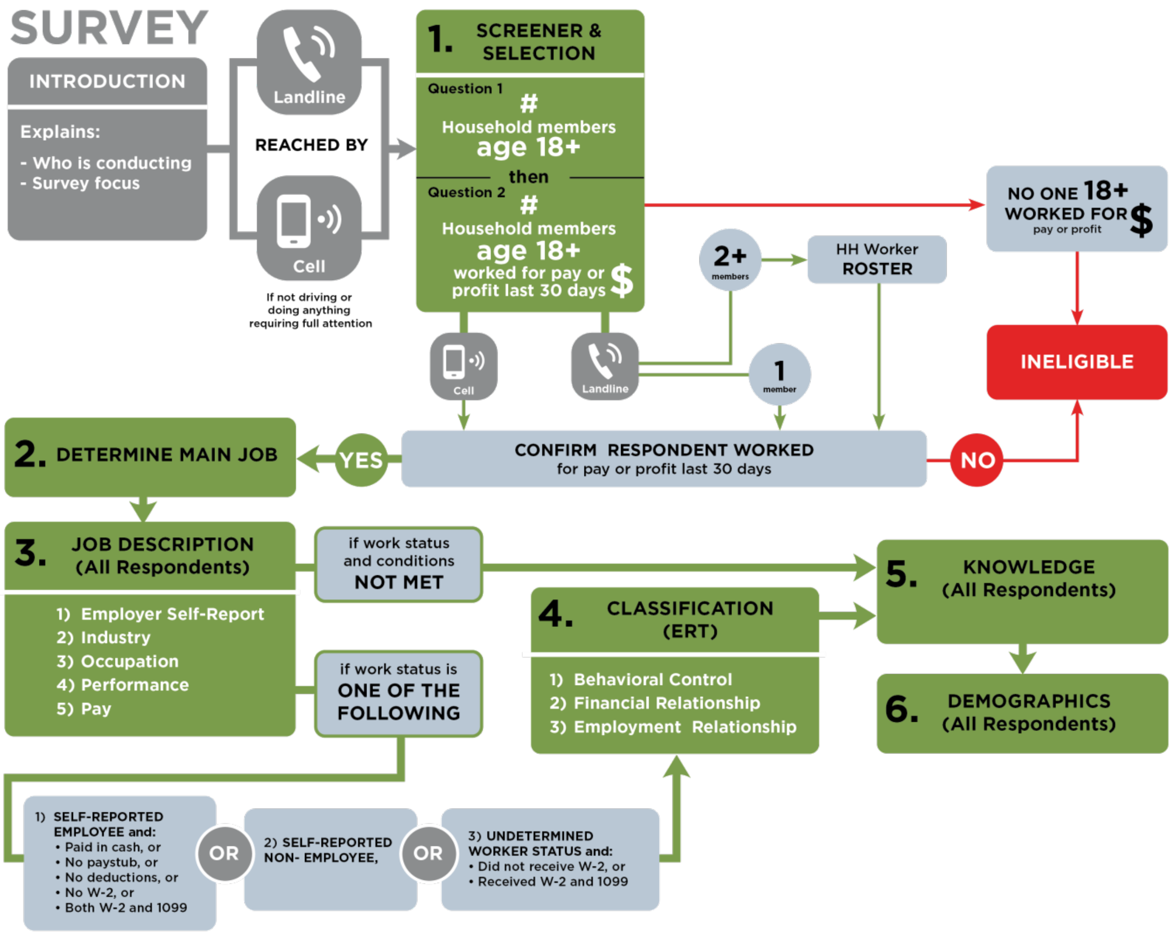
---

[1] In 2012 the CDC’s Behavioral Risk Factor Surveillance System averaged an allocation of 20% cell phones and 80% landlines across the 50 state surveys. The nation’s largest state-level health survey (UCLA’s California Health Interview Survey) was allocated to 20% cell phones and 80% landlines. National surveys conducted for Stanford University’s Global Climate and Energy Project were allocated to 25% cell phones and 75% landlines. National surveys conducted for the Pew Research Center were allocated to 40% cell phones and 60% landlines.

4. Data Collection Protocol

Exhibit 4.1 presents a high-level graphical representation of the flow of the data collection protocol, from the introduction, to screening, to the extended interview.

Exhibit 4.1: Data Collection Protocol



The recruitment protocol featured a maximum of 10 call attempts to landline numbers, and a maximum of 8 call attempts to cell phone numbers. In a small fraction of cases, where we had determined there was an eligible adult, a few additional calls were made in the effort to complete the interview. Interviewers attempted refusal conversion only on soft refusal cases and only in the landline sample.

To maximize the response rate, we implemented a resting protocol where portions of the sample were deactivated at various points in the field period. Following Sangster (2012), we assumed that resting the non-contact sample at the mid-point of the call design in both frames would improve our ability to subsequently reach a respondent to complete an interview or screen out. After five non-contacts were recorded for landline cases, we suspended dialing on that number for two weeks. After the two weeks, we resumed the call design. For cell phone cases, the two-week break was implemented after four non-contacts were recorded.

Interviewers placed phone calls from 5:00 p.m. to 9:00 p.m. on weekdays, from 10:00 a.m. to 6:00 p.m. on Saturdays, and from noon to 9:00 p.m. on Sundays. Consistent with FCC regulations, we made calls to landline telephone numbers using an auto dialer and manually dialed all calls to cell phone numbers. Interviewers dialed telephone numbers until contact was established with a respondent associated with the number, or until the telephone number was determined to be nonresidential or out of service. Interviewers verified that the cell phone respondent was an adult who was not currently driving a vehicle before administering the screener. Because we were using a random-digit dial telephone sample, we did not elect to send advance letters. Addresses would have had to be identified through reverse directory searches, and since a large proportion of the sample were cell phones and a large proportion of the landline sample are usually bad numbers, we decided against this recruitment option. We offered cell sample respondents a postage-paid cash incentive of $10 to compensate them for their time. Providing an incentive to cell phone respondents encourages participation and also offsets any costs they may incur for minutes charged during the survey administration (Oldenick, R.W. and Lambries 2011; Pew Research Center 2015).

We released sample for interviewing in replicates, which are representative subsamples of the larger sample. Using replicates to control the release of sample ensures that complete call procedures are followed for the entire sample. Regional quotas were implemented to ensure that the geographic distribution of the responding sample corresponded to the distribution of the target population.

The average length of a completed interview was 18 minutes in the landline sample and 17 minutes in the cell phone sample. We conducted interviews, in both English and Spanish, from August 26, 2014, to March 9, 2015. (Both versions of the questionnaire appear in an appendix in Volume I of this report.)

A nonresponse follow-up survey (NRFU) was conducted shortly after completion of the Worker Classification Knowledge Survey. The NRFU attempted to interview a subsample of nonrespondents to assess whether nonrespondents had different characteristics than respondents. Details and analysis of nonresponse appear in Section 7.

5. Weighting

Weighting is needed to account for the complex design (as well as other factors) and to support inference to the target population. In this section we describe the weighting procedures and variance calculation of survey estimates.

The full sample weight (labeled as "WEIGHT" in the data file) accounts for the sample selection probabilities, sampling frame overlap, and statistical adjustments for potential under-coverage and nonresponse biases. The full sample weight was computed in several stages. The first stage adjusted for the probability of selection of the telephone number (computed separately by frame and region) and the subsampling of flagged-inactive cell phone numbers.

Next, two successive weighting cell adjustments were performed. The first adjusted for the number of unresolved cases (i.e., unknown if working and residential) estimated to be working, and the second adjusted for nonresponse to the screener interview. In both of these adjustments, the weighting cells were defined by the cross-classification of sampling frame and census region.

A further set of adjustments corrected for multiple chances of selection. The weights for landline cases were multiplied by the number of eligible adults in the household (capped at three) and by the inverse of the number of working voice-use landlines in the household (capped at two). The weights for the cell phone cases were multiplied by the inverse of the number of working cell phones that the respondent had (capped at two). The capping on these adjustments serves to avoid undue variance in the weights. These multiplicity-adjusted weights were then adjusted for the overlap in the landline and cell phone frames (some adults have both kinds of phones and could have been selected in either sample) using a compositing method of frame integration (Hartley 1962, 1974). In this method, a compositing factor between 0 and 1, typically denoted by the Greek letter λ, is selected. The weights of cases that came from the landline frame are multiplied by λ, and those of the cases that came from the cell frame, by $1 - \lambda$ (Lohr 2009).

After the landline and cell phone samples were statistically combined, we performed a weighting cell adjustment for nonresponse to the extended interview. The weighting cells were defined by the cross-classification of sampling frame and the number of eligible workers in the household (capped at three).

The next step was to calibrate the responding sample to benchmark demographic distributions for the target population. This calibration serves to reduce potential nonresponse and non-coverage errors. We used an operation known as raking ratio estimation (Kalton 1983), also known simply as raking, or sample-balancing.

We raked the sample to population control totals for age, gender, education level, race/ethnicity, census region, and household telephone service (i.e., cell phone only, landline only, or having both cell and landline). After examining the distribution of the weights, the maximum weight value was trimmed to equal the median plus six times the interquartile range (Chowdhury, Khare, and Wolter 2007). Trimming serves to avoid undue variance in the survey weights. The full sample weights are scaled to sum to the estimated size of the target population according to the Current Population Survey (CPS).

All of the population control totals, with the exception of telephone service, came from an analysis of the March 2014 CPS Annual Social and Economic Supplement (ASEC), filtered on adults 18 years and older residing in the United States who worked for pay in the past week. The population control total for telephone service was constructed based on national estimates from the 2013 National Health Interview Survey Public Use Microdata file and filtered on adults who currently work for pay in U.S. households with a telephone. These estimates were then updated to reflect national trends in telephone service for all adults from January to June 2014.

5.1 The Population Who Worked in the Last 30 Days Versus in the Past Week

The definition of working adults as defined in the population control totals we used for calibration is slightly different than the target population of the survey, the latter of which requires that adults be employed for pay in the 30 days before the interview. The CPS and the American Community Survey (ACS) both use a reference period of "worked in the last week."

To examine empirically whether this approach influenced our estimates, we asked all respondents whether they had worked in the last week. We used those data to filter the survey sample on respondents who had worked in the past week, and then we created an experimental weight for those cases (n=7,596) following the same protocol as for the full sample weight. We examined 13 key survey questions administered to all respondents, and compared the weighted estimates based on the full sample (n=8,503 who worked in the past 30 days) with the estimates based on the smaller sample (n=7,596 who worked in the past week). The average absolute value of the difference between the weighted estimates was less than 1 percentage point. If, however, meaningful differences had been observed (e.g., an average of more than 1.5 percentage points for a set of key survey estimates), then we would have considered experimental weights as the final survey weights and dropped respondents who had not worked in the past week from the dataset. This is based on the logic that the experimental weights would have been more accurate because the survey target population and the population identifiable in the CPS would be approximately the same.

Based on our analysis, we concluded that the discrepancy between the survey target population and the population represented by the weighting control totals did not represent a meaningful threat to the study data quality. Therefore it was reasonable to proceed using the full sample weight, which includes all 8,503 respondents.

5.2 Variance Estimation

The complex design used for the Worker Classification Knowledge Survey requires proper weighting and variance calculation of the estimates. By default, most statistical software packages will calculate variance assuming that the data are from a simple random sample; doing this would underestimate the variance of estimates produced from the Worker Classification Knowledge Survey's complex sample design. Given the complexity of the sampling design, eligibility screening, nonresponse, and calibration weight adjustment, no explicit variance calculation can be provided for the Worker Classification Knowledge Survey. To accurately estimate variance without jeopardizing data confidentiality and respondent privacy, the dataset provides 100 replicate weights (labeled in the data file as RPL001, …, RPL100) in addition to the full sample weight (WEIGHT). The replicates were computed using the Jackknife Delete Two procedure (Kott 2001). In each of the replicate weights, all weighting steps were repeated using jackknife design replicate weights instead of the base weights, so sampling variability is properly accounted for. For a parameter estimate of interest θ, replicate variance estimation proceeds by repeating the estimation procedure with the r-th set of replicate weights $w_i^{(r)}$, obtaining a point estimate

$\hat{\theta}^{(r)}$ and the variance estimate for $\hat{\theta}$ is then obtained as $V[\hat{\theta}] = \frac{n-2}{2C_n^2}\sum_{r=1}^{100}(\hat{\theta}^{(r)} - \theta)^2$ (Kott 2001; Kolenikov 2010). Complex survey estimation software that supports replicate weights needs to be used, such as SAS, Stata, or R. When analyzing data from the Worker Classification Knowledge Survey dataset, if the single full weight (WEIGHT) is used as a frequency weight in place of using the replicate weights, without specifying the sampling design, the variability of estimates will be underestimated due to the aforementioned incorrect assumption that the sample is a simple random sample.[2]

[2] Please refer to Technical Documentation for an illustration of how the Worker Classification Knowledge Survey can be analyzed to produce valid variance estimates using SAS/STAT 9.2 or higher, using the dataset of 8,503 completed interviews.

6. Final Dispositions and Response Rates

Exhibit 6.1 below reports the final dispositions of all sampled telephone numbers dialed for the survey. Many numbers dialed were determined to be working and residential, but no screener was completed. Such numbers are coded as "unknown eligibility."

Exhibit 6.1: Final Dispositions by Sample

	AAPOR disposition code	Landline sample	Cell sample	Total combined sample
Interview (Category 1)				
Complete	1.000	2,554	5,949	8,503
Partial	1.200	293	821	1,114
Eligible, non-interview (Category 2)				
Refusal and breakoff	2.100	110	193	303
Unknown eligibility, non-interview (Category 3)				
Always busy	3.120	1,287	1,786	3,073
No answer	3.130	14,520	3,088	17,608
Call blocking	3.150	49	128	177
No screener completed: no live contact made	3.211	10,567	36,381	46,948
No screener completed: live contact made	3.212	17,793	21,703	39,496
Other: "cell phone" disposition used in error	3.910	n/a	26	26
Other: cell case physically or mentally unable	3.920	n/a	527	527
Other: cell case language problem	3.930	n/a	1,342	1,342
Not eligible (Category 4)				
Fax/data line	4.200	6,260	156	6,416
Non-working/disconnect	4.300	128,390	22,753	151,143
Temporarily out of service	4.330	3,084	1,202	4,286
Cell phone	4.420	54	0	54
Business, government office, other organizations	4.510	8,913	3,791	12,704
No eligible respondent	4.700	5,814	8,684	14,498
Child phone (under 18 years of age)	4.900	90	1,988	2,078
Total phone numbers used		199,778	110,518	310,296

Exhibit 6.2 presents call outcomes for the screener and extended surveys. The table displays the total number for each screener result, and displays the resultant estimated eligibility rate and response rate by sample. The overall AAPOR response rate 3 (RR3) is 18.9 percent for the landline sample and 18.5 percent for the cell sample.

Exhibit 6.2: Response Rates, by Sample

Call outcomes and response rates	Landline sample	Cell sample
Screener status [AAPOR disposition codes]*		
Ineligible for screener (4.20–4.51)	146,701	27,902
Eligible screener respondent (1.0,1.2,2.1,4.7,4.9)	8,861	17,635
Eligible screener nonrespondent (3.21–3.93)	28,360	59,979
Eligibility undetermined (3.12–3.15)	15,856	5,002
Total	199,778	110,518
Screener "e" (estimated rate of eligibility)	20.2%	73.6%
Screener response rate (AAPOR RR3)	21.9%	21.7%
Extended survey status		
Completion	2,554	5,949
Partial	293	821
Nonrespondent	110	193
Extended survey response rate (AAPOR RR1)	86.4%	85.4%
Extended survey response rate (AAPOR RR2)	96.3%	97.2%
Overall response rate		
AAPOR RR3	18.9%	18.5%
AAPOR RR4	21.1%	21.1%

*AAPOR, 2015.

The Worker Classification Knowledge Survey is the first of its kind, and so its response rates cannot be compared to any historical response rates. As an alternative, we researched studies with similar sample designs and sponsorship, noting that many factors contribute to response rate, including the field period, call design, survey topic, and length. A 2012 DOL study of adult workers using a similar overlapping dual-frame design yielded an AAPOR response rate 3 of 11.2 percent (Gallup 2013). In adherence to Office of Management and Budget regulations on response rates under 80 percent, we conducted an empirical analysis of nonresponse, described below.

7. Analysis of Nonresponse

Most standard statistical procedures assume that the sample data represent the underlying population. In survey research, this assumption is violated when unmitigated nonresponse bias exists in the data. Nonresponse bias arises when different population groups respond at different rates and the differences in response propensity are associated with one or more of the survey outcomes (National Research Council 2013).

When assessing the risk from nonresponse bias, two key properties are particularly relevant. First, nonresponse bias can be negligible for some survey estimates and large for other estimates. In other words, nonresponse bias is an estimate-specific phenomenon. Nonresponse bias varies over estimates within a survey as a function of whether the likelihood of survey participation is related to the variable underlying the estimate (Bethlehem 2002; Groves and Peytcheva 2008). A second, closely related property of nonresponse is that it has been shown to be a rather poor indicator of survey data quality. In his examination of a set of 30 studies, Groves (2006) found that response rates “explain” only about 11 percent of the variation in different estimates of nonresponse bias. This suggests that just because the response rate is low, it would be incorrect to conclude that the survey estimates are therefore not accurate. In fact, several studies have shown that surveys with relatively low response rates can still produce highly accurate estimates when compared to benchmark data (Keeter et al. 2000, 2006; Merkle and Edelman 2002).

Since we do not observe responses for nonrespondents, a direct comparison of respondents and nonrespondents is impossible. Instead, the nonresponse analysis for the Worker Classification Knowledge Survey uses four conventional proxy analyses:

1. ***Nonresponse Follow-up Survey and Comparative Analysis of NRFU and Main Sample Responses.*** An explicit re-contact of nonrespondents was undertaken using a very short instrument consisting of key Worker Classification Knowledge Survey variables. The goal of the NRFU is to provide insight into whether the nonrespondents differ from the respondents on the characteristics of interest (e.g., work experiences and benefits).

2. ***Comparative Analysis of Easier to Reach vs. Harder to Reach Respondents.*** This analysis evaluates whether there were differences in key outcomes across the spectrum of the contact effort. This assumes that the late responders act as proxies for nonrespondents in the sense of being more difficult to reach or to convince to participate in the survey. Support for this “continuum of resistance” model is inconsistent (Lin and Schaeffer 1995, Montaquila et al. 2007), but it can still be a useful framework for assessing the relationship between level of effort and nonresponse bias.

3. ***Estimating Response Propensity Models.*** Response propensity is the theoretical probability that a sampled unit will respond to the survey request (Little 1986, Groves and Couper 1998, Olson 2006). Many respondent characteristics can influence response propensity. The response propensity model allows the researcher to identify the most powerful predictors of response when all available predictors are tested simultaneously.

4. ***Comparison to CPS.*** Finally, we compare the population represented by the Worker Classification Knowledge Survey with the population represented by the CPS. The CPS relies more heavily on face-to-face contact, and consequently it is a higher response rate federal survey.

While these analyses rely on imperfect assumptions, all are standard techniques for assessing potential nonresponse error. The first three techniques above analyze only a subset of all nonrespondents to the survey. The NRFU analysis relies on the NRFU participants as proxies for nonrespondents; the level of effort analysis relies on the “harder-to-reach” respondents as proxies for nonrespondents; the response propensity model captures only variation between the screened extended interview respondents and the screened extended interview nonrespondents. The fourth technique—comparison to external benchmarks—depends on the availability and comparability of the external data source. Specifically, not all of the key survey variables in the Worker Classification Knowledge Survey are collected in the CPS nor are the questions worded identically. Further, while the CPS provides the best available estimates for the comparison measures, the CPS estimates may themselves contain some level of error.

In sum, no single nonresponse analysis for this study can be definitive because the true scores of the nonrespondents are not known. That said, by using several different methodologies (nonresponse follow-up analysis, easy-to-reach versus hard-to-reach comparisons, response propensity models, and comparisons of estimates to external benchmarks), we can draw meaningful conclusions about the level of risk to survey estimates from nonresponse bias. Results from each of these analyses are described in further detail below.

7.1 Nonresponse Follow-Up Survey (NRFU)

A nonresponse follow-up survey was conducted shortly after the Worker Classification Knowledge Survey was completed. The NRFU attempted to interview a subsample of nonrespondents to the survey in order to assess whether they had different characteristics than respondents and, if so, whether any differences remained after controlling for major weighting cells (e.g., within race and education grouping).

A total of 2,958 nonrespondents to the Worker Classification Knowledge Survey were selected to be re-contacted in the NRFU based on the stratified sample design shown in Exhibit 7.1. All eligible nonrespondents from the landline sample who had already started the household screener were selected to be re-contacted for the NRFU. All nonresponding cases from the cell sample were eligible to be selected for the NRFU with the exception of hard refusals and second soft refusals. Of the eligible cell sample nonrespondents, all screened cases were selected to be re-contacted in the NRFU, while un-screened cell cases were subsampled at lower rates. The difference in selection rules for the two frames follows recommended industry protocol as outlined by the American Association for Public Opinion Research (AAPOR 2008) and also discussed by Triplett et al. (2002).

Exhibit 7.1: Sample for Worker Classification Knowledge Survey NRFU Survey

Call disposition* (at end of main field period, prior to NRFU)	Total cases available	NRFU subsampling rate	NRFU sample
Cell phone sample			
2.0 Eligible non-interview	820	100.00%	820
3.0 Unknown eligibility, non-interview	55,434	3.22%	1,785
Landline sample			
2.0 Eligible non-interview	353	100.00%	353
Total			2,958

*AAPOR Final Disposition Codes for RDD Telephone Surveys

The NRFU was conducted via CATI and featured a shortened version of the Worker Classification Knowledge Survey instrument. All NRFU cases were offered a $20 post-paid remuneration for completing the interview. The NRFU was conducted over the period of March 18 to April 1, 2015. From the sample of 2,958, a total of 209 interviews were completed with adults who had worked for pay in the past 30 days. Among the 2,958 adults, 97 were determined to be ineligible (18 in the landline sample and 79 in the cell sample). A total of 69 completed NRFU cases were from the landline sample and 140 cases came from the cell phone sample.

Exhibit 7.2 compares unweighted estimates based on all 8,503 extended (i.e., non-NRFU) interview respondents to the Worker Classification Knowledge Survey with unweighted estimates based on the 209 eligible respondents reached in the NRFU. With the exception of the types of paycheck deductions, the results suggest no major differences between the nonrespondents reached in the NRFU and the main survey respondents. Respondents in the main survey were significantly more likely to report having Social Security and Medicare taxes deducted from their pay compared to NRFU respondents (87.0 percent vs. 80.0 percent, p=0.00253). These results suggest that workers who do not have these taxes deducted from their paycheck may have been less likely to respond to the original survey, and hence underrepresented in it.

Exhibit 7.2: Characteristics of Main Survey Respondents vs. NRFU Respondents

Characteristic	Main survey respondents	NRFU respondents	Difference of proportions *p*-value[a]
Social Security and Medicare taxes (FICA) deducted from pay (% Yes)	87%	80%	0.00253**
Federal, state or local taxes deducted from pay (%Yes)	90%	86%	0.1145
Main employer is private for-profit company	59%	55%	0.2116
Usually work more than 35 hours per week at main job	77%	74%	0.2385
Receive or have access to a paystub or document listing pay and deductions (%Yes)	87%	85%	0.3891
Main job is temporary (% Yes)	10%	10%	0.8419
Two or more paid jobs in last 30 days (including part-time, weekend or evening work)	12%	11%	0.9500
Number of interviews (minimum for items shown)	8,503	209	

[a] p-values correspond to a two-sided difference of proportions test. *** p<.001, ** p<.01, * p<.05.

7.2 Comparison of Easier to Reach Versus Harder to Reach Respondents

The second technique that we used to assess the risk of nonresponse bias is an analysis of differential response by the level of recruitment difficulty. Here we compare the employment-related characteristics of respondents who were easy to reach with those of respondents who were harder to reach. (We provide working definitions of these concepts below.) In this analysis, the level of difficulty in reaching the respondent is based on three dimensions: (1) ease of "contactability" as defined by the number of calls required to complete the interview; (2) amenability as defined by whether or not the case was a converted refusal; and (3) a hybrid metric combining number of call attempts and converted refusal status. Just over half (55.3 percent) of the 8,503 extended interview respondents completed the interview on the first, second, or third call. The remainder (44.7 percent) required at least four calls, with a maximum of up to

12 calls. A small fraction of respondents (n=298, or 3.5 percent of the full responding sample) were converted refusals.[3] In such cases, 11.7 percent of the responding landline sample were converted refusals. Some 45.4 percent of respondents either required four or more calls or were converted refusals. These cases are referred to as "hard-to-reach" in this analysis. Respondents who never refused and completed the interview in three or fewer calls are referred to as the "easy-to-reach." The rationale for this grouping was to have subsamples of approximately equal size that would provide the greatest power to find the differences between the easier-to-reach and harder-to-reach respondents on the continuum of resistance.

Exhibit 7.3 presents several characteristics for these various groups. In this table, each respondent is represented three times according to number of attempts they required, whether or not they ever refused, and whether they were easy or hard to reach based on the hybrid metric. The "hard-to-reach category" included 3,863 respondents, of whom 298 were refusal conversions and 3,565 respondents with 4+ call attempts but no refusal conversion. The 3,798 figure refers to cases that had 4+ call attempts (both with and without refusal conversion.) The minimum sample size is the sample size accounting for potential item missing data, with missing data rates varying across estimates. The only significant finding is that respondents from households that had previously refused to participate in the main survey were more likely to refuse or answer "don't know" when asked how they usually receive pay from their main job, compared to those who never refused to participate (4.0 percent vs. 1.2 percent, chi-square p <.01). Responses to this question were not affected by the number of call attempts or the hybrid measure.

We also analyzed how these groups compared with respect to the number of jobs worked in the past 30 days, number of hours usually worked each week, type of employer at their main job, whether their main job was temporary, and how certain they were about their worker classification status (self-employed or employee). On all of these measures, responses did not vary significantly based on the level of difficulty in reaching respondents.

Exhibit 7.3: Survey Estimates by Level of Contact Effort

	Contact attempts		Refusal behavior		Hybrid	
Survey estimate	3 or fewer attempts	4 or more attempts	Never refused	Converted refusals	Easy to reach	Hard to reach
Number of paid jobs in last 30 days (including part-time, weekend or evening work)						
One paid job	88.3%	88.4%	88.2%	90.9%	88.2%	88.4%
Two or more paid jobs	11.7%	11.6%	11.7%	8.7%	11.8%	11.5%
Don't know/refused	0.1%	0.1%	0.1%	0.3%	0.1%	0.1%
Number of hours usually worked per week at main job						
Less than 35 hours per week	23.4%	21.4%	22.6%	19.8%	23.5%	21.4%
35 or more hours per week	76.4%	78.4%	77.2%	80.2%	76.4%	78.5%
Don't know/refused	0.2%	0.2%	0.2%	0.0%	0.2%	0.2%
Type of employer at main job						
Government	19.1%	17.8%	18.6%	17.8%	19.1%	17.9%

[3] In some of these cases, the refusal may have come from the screener respondent rather than the extended interview respondent, if these happened to be different people.

Survey estimate	Contact attempts		Refusal behavior		Hybrid	
	3 or fewer attempts	4 or more attempts	Never refused	Converted refusals	Easy to reach	Hard to reach
Private for-profit company	58.8%	60.1%	59.5%	55.4%	58.8%	60.0%
Non-profit organization	9.4%	9.1%	9.2%	9.7%	9.4%	9.1%
Self-employed	10.7%	10.8%	10.5%	16.1%	10.6%	10.9%
Other	1.0%	1.0%	1.0%	0.7%	1.1%	1.0%
Don't Know	1.1%	1.1%	1.1%	0.3%	1.1%	1.1%
Usually receive pay from main job as:						
Company check	28.5%	30.7%	29.6%	26.2%	28.7%	30.4%
Personal check	3.0%	3.1%	3.0%	3.4%	3.0%	3.0%
Cash	2.0%	2.2%	2.1%	2.7%	2.0%	2.2%
Direct deposit	62.4%	59.9%	61.3%	60.7%	62.2%	60.1%
Other	2.9%	2.9%	2.9%	3.0%	3.0%	2.9%
Don't know/refused	1.3%	1.4%	1.2% **	4.0% **	1.2%	1.5%
Main job is temporary:						
Yes	10.5%	10.5%	10.5%	9.4%	10.5%	10.4%
No	88.9%	88.8%	88.8%	90.6%	88.8%	88.9%
Don't know/refused	0.7%	0.7%	0.7%	0.0%	0.7%	0.7%
Certainty about being self-employed at main job						
Very/somewhat certain	91.3%	91.6%	91.2%	95.2%	91.1%	91.8%
Not too/not at all certain	6.7%	6.6%	6.9%	2.4%	6.9%	6.4%
Don't know/refused	2.0%	1.9%	1.9%	2.4%	2.0%	1.8%
Certainty about being employee at main job						
Very/somewhat certain	97.0%	96.5%	96.7%	97.6%	96.9%	96.6%
Not too/not at all certain	2.0%	2.4%	2.2%	1.6%	2.0%	2.3%
Don't know/refused	1.1%	1.1%	1.1%	0.8%	1.1%	1.1%
Minimum sample size	4,705	3,798	8,205	298	4,640	3,863

Source: The Worker Classification Knowledge Survey; figures are unweighted.

*** p<.001, ** p<.01, * p<.05.

7.3 Response Propensity Modeling

In this analysis, the primary research question is whether the number of workers in the household is associated with response propensity, especially when controlling for factors included in the weighting protocol. If this factor shows a significant association with response to the extended interview (after controlling for other factors), this would be evidence of possible nonresponse bias. If, however, these factors do not have a significant effect, this suggests that the weighting adjustments are likely to have been effective in reducing nonresponse bias. In these analyses, we condition on the sample selected, and do not attempt to generalize the nonresponse to the target population, as nonresponse is specific to the particular ways the survey field operation was conducted. We thus used unweighted estimates that may differ from those in the primary analytical study report.

In order for a response propensity model to be informative, the researcher must know the values for respondents and nonrespondents on one or more predictors of survey response. In RDD surveys, propensity models are often quite limited, because little information is generally known for the nonrespondents. This is the case for the screener component of the Worker Classification Knowledge Survey. The only types of variables known for the nonrespondents to the screener are sampling frame, region, and level of effort.

A richer model can be constructed to model the propensity to respond to the extended interview for the households that have completed the screener, as this model could additionally include the number of workers, which is known for both respondents and nonrespondents to the extended interview. It should be noted that such modeling does not shed light on nonresponse occurring between the initial contact and completion of the screener.

A logistic regression was used to model response to the extended interview conditional upon completion of the screener. The results are presented in Exhibit 7.4.[4] The explanatory power of the model is fairly low; the area under curve statistic is just 0.536 (values higher than 0.8 are usually considered indicative of good explanatory power). The strongest predictors of response to the extended interview are the sampling frame and region of residence. Landline RDD sample cases who completed the screener were significantly more likely to complete the extended interview than cases drawn from the cell phone RDD sample. Compared to screened households from the West region of the United States, screened households from the Northeast region were significantly less likely to respond to the extended interview, while screened households from the Midwest were significantly more likely. The number of workers in the household was not a significant predictor of the likelihood of the screened household completing the extended interview.

4 Interviewing effort variables, such as the number of call attempts and an indicator for converted refusal cases, were intentionally excluded from this model because they are endogenous and also because a significant association with the outcome being modeled would not communicate any information about the potential risk to survey estimates from nonresponse bias. Interviewing effort variables are considered separately in the analysis of easier-to-reach versus harder-to-reach cases.

Exhibit 7.4: Logistic Regression Estimating the Probability of Response to the Extended Interview Conditional on Completion of the Screener

Parameter	Estimate	s.e.	Wald X^2	p-value	
Intercept	1.366	0.096	199.57	<0.001	***
Sampling frame = cell RDD	-0.073	0.028	6.82	0.009	**
Region = Northeast	-0.158	0.047	11.30	0.001	**
Region = Midwest	0.142	0.048	8.71	0.003	**
Region = South	-0.052	0.039	1.82	0.178	
Number of workers in the household (log)	0.140	0.101	1.94	0.164	
Model diagnostics					
Area under ROC curve (c)	0.536				
-2 Log Likelihood	10,124.4				
Sample Size	10,480				

Source: The Worker Classification Knowledge Survey; figures are unweighted.

Reference groups for categorical variables: sampling frame (landline RDD),region (West).

*** p<.001, ** p<.01, * p<.05.

In terms of potential nonresponse bias, these findings do not represent cause for concern, because the weighting protocol addresses the integration of the sampling frames, and it includes a post-stratification to region of residence, which minimizes the risk of nonresponse bias associated with this effect.

7.4 Comparison to External Benchmarks

Specifically, we compare the weighted final respondent estimates from the Worker Classification Knowledge Survey to those from the CPS. The CPS is considered to be a "gold standard" survey due to its rigorous protocol (i.e., area-probability sampling with in-person interviewing) and high response rate. The strength of this approach is that the benchmark survey (CPS) is well known to be a high-quality federal survey, and so obtaining similar estimates would give some confidence about the Worker Classification Knowledge Survey (Groves 2006). In order to match the target population of the Worker Classification Knowledge Survey targets as closely as possible, CPS weighted estimates were computed based on the population of adults aged 18 and older with a telephone who were employed for pay within the past week (either at work or absent). Five variables identified in the CPS were also administered in the Worker Classification Knowledge Survey but not used in the weighting protocol.[5] These variables are marital status, household income, labor union membership, size of employer, and whether the person attended or was enrolled in a high school, college, or university in the past week. The weighted estimates from both surveys are presented in Exhibit 7.5.

5 Several demographic variables such as age, gender, education, and race/ethnicity are measured in both the CPS and the Worker Classification Knowledge Survey. These variables were intentionally excluded from this analysis, however, because they were included in the raking ratio calibration for the Worker Classification Knowledge Survey weights. In other words, the Worker Classification Knowledge Survey was statistically adjusted to match external benchmarks on these measures, and so comparing those weighted characteristics to the CPS would not be informative about the risk of nonresponse bias.

Exhibit 7.5: Estimates from the Main Worker Classification Knowledge Survey and the Current Population Survey

Characteristic	2014 CPS ASEC	Worker Classification Knowledge Survey	Difference, CPS versus WCKS	*p*-value
Marital status				
Currently married	53.8%	52.7%	1.1%	0.117
Not married	46.2%	47.3%	-1.1%	0.117
Household income past 12 months (in current figures)				
Less than $30,000	12.0%	19.0%	-7.0%***	0.000
$30,000 to $74,999	35.5%	35.3%	0.3%	0.723
$75,000 and higher	52.5%	45.8%	6.7%***	0.000
Labor union membership				
Yes	11.6%	11.0%	0.6%	0.272
No	88.4%	89.0%	-0.6%	0.272
Total number of employees who work for your employer				
Under 10 employees	19.5%	20.1%	-0.6%	0.304
10–49 employees	14.8%	20.1%	-5.3%***	0.000
50–99 employees	7.0%	8.5%	-1.5%***	0.000
100–499 employees	12.8%	17.3%	-4.5%***	0.000
500+ employees	45.9%	34.1%	11.8%***	0.000
Last week attended or enrolled in a high school, college, or university				
Yes	9.9%	11.3%	-1.3%**	0.003
No	90.1%	88.7%	1.3%**	0.003

Sources: Worker Classification Knowledge Survey and March 2014 CPS. Estimates from both surveys are weighted. Estimates exclude item nonresponse. Totals may not sum to 100 due to rounding.

*** $p<.001$, ** $p<.01$, * $p<.05$.

The weighted Worker Classification Knowledge Survey estimate of household income suggests that adults with household income less than $30,000 are over-represented in the Worker Classification Knowledge Survey compared to the CPS (19.0 percent versus 12.0 percent), while those with household income of $75,000 and higher are under-represented (45.8 percent versus 52.5 percent). Differences in marital status are not significantly different. The Worker Classification Knowledge Survey also over-represents adults who attended or were enrolled in a high school, college, or university in the past week compared to estimates from the CPS (11.3 percent versus 9.9 percent).

The weighted Worker Classification Knowledge Survey estimate for percentage of adults belonging to a labor union and the percentage of adults who work for small businesses is highly similar to the adult population estimates from the CPS. However, adults who worked for mid-size organizations with 10 to 499 employees tended to be over-represented in the Worker Classification Knowledge Survey compared to the CPS. The largest discrepancy in weighted estimates between the Worker Classification Knowledge Survey and the CPS was observed in adults who worked for employers with 500 or more employees (34.1

percent versus 45.9 percent). Therefore, the potential for bias exists in the survey due to the under-representation of adults who work for the largest employers. However, differences in the question wording, mode of administration, and population coverage between the gold standard and target survey may confound the comparison. Establishment size is based on self-report both in CPS and in the Worker Classification Knowledge Survey, and is not validated against establishment-level records. In light of these considerations, results from external comparisons must be interpreted with caution.

7.5 Nonresponse Analysis Conclusion

We presented several comparisons aimed at establishing whether the respondents to the Worker Classification Knowledge Survey differed from nonrespondents. An additional contact effort with a shorter instrument (NRFU) showed that out of seven indicators compared, the successfully interviewed nonrespondents to the main survey were less likely (at 5 percent significance level) to have Social Security and Medicare taxes (FICA) deducted from pay. Comparison of harder-to-reach respondents vs. easier-to-reach respondents showed that out of seven variables with 27 categories, only one category ("Don't know/Refused" on the means of payment) showed significant differences between these two groups. The response propensity analysis with frame, region, and the screener information (number of working adults in the household) revealed response differences by frame and region (which were controlled for during the weighting process), but not by the employment-related variable. Comparison of worker characteristics between the main Worker Classification Knowledge survey and a higher-contact-effort survey, the CPS, found no differences in marital status, union membership, and employment in small firms, but detected that respondents to the Worker Classification Knowledge Survey resided in household with higher incomes, were more likely to be employed in mid-size companies (10 to 499 employees), were less likely to be employed in large companies (500+ employees), and were more likely to attend school or college. Overall, these results indicate a moderate risk of nonresponse biases for analyses related to the household income and company size.

References

American Association for Public Opinion Research. (2014). *AAPOR terms and conditions for transparency certification.* Retrieved from http://transparency.aapor.org/docs/TransparencyTermsAndConditions.pdf.

American Association for Public Opinion Research. (2010). *New considerations for survey researchers when planning and conducting RDD telephone surveys in the U.S. with respondents reached via cell phone numbers.* Retrieved from https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/2010AAPORCellPhoneTFReport.pdf

American Association for Public Opinion Research (2008). *Guidelines and considerations for survey researchers when planning and conducting RDD and other telephone surveys in the U.S. with respondents reached via cell phone numbers*. Retrieved from https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/Final_AAPOR_Cell_Phone_TF_report_041208.pdf

Bethlehem, J. G. (2002). Weighting nonresponse adjustments based on auxiliary information. In Groves, R., Dillman, D., Eltinge, J., & Little, R. (Eds.), *Survey nonresponse* (pp. 275–288). New York: Wiley.

Blumberg, Stephen J., and Julian V. Luke. *Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January–June 2015*. Division of Health Interview Statistics, National Center for Health Statistics.

Brick, J. M., Brick, P., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2007). Cell phone survey feasibility in the U.S.: Sampling and calling cell numbers versus landline numbers. *Public Opinion Quarterly*, *71*(1), 23–29.

Bureau of Labor Statistics, U.S. Department of Labor. (2014). *American Time Use Survey*. Retrieved from: http://www.bls.gov/tus/

Cantor, D., Coa, K., Crystal-Mansour, S., Davis, T., Dipko, S., & Sigman, R. (2009, February). *Health Information National Trends Survey final report.* New York: Westat. Retrieved from http://hints.cancer.gov/docs/HINTS2007FinalReport.pdf

Chowdhury, S., M. Khare, and K. Wolter. (2007). Weight Trimming in the National Immunization Survey. *Proceedings- American Statistical Association -Cd-Rom Edition*- 2651-2658. https://www.researchgate.net/publication/260348642_Weight_Trimming_in_the_National_Immunization_Survey

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*, 229–238.

Dutwin, D. (2014). "The Future of Landline and Cell Phone Telephone Surveys in the U.S." Presented at the Annual Meeting of the American Association for Public Opinion Research.

Dutwin, D., & Malarek, D. (2014). Recent activity flags for cellular samples. *Survey Practice 7*(1). Retrieved from http://www.surveypractice.org/index.php/SurveyPractice/article/view/57/html

Gallup. (2013). *Survey of public opinion of U.S. worker rights.* Retrieved from http://www.dol.gov/asp/evaluation/completed-studies/Workers_Rights_Access_Assertion_and_Knowledge_Study/APPENDICES_workers_rights_access_assertion_knowledge_study.pdf.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview studies.* New York: Wiley.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Hartley, H. O. (1974). Multiple Frame Methodology and Selected Applications. *Sankhya, Ser. C, 36,* 99–118.

Hartley, H. O. (1962). Multiple frame surveys. *Proceedings of the Social Statistics Section of the American Statistical Association*. 203–206.

Kalton G. (1983). *Compensating for Missing Survey Data* . Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly, 70*(5), 759–779.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*(2), 125-148.

Kolenikov, S. (2010). Resampling variance estimation for complex survey data. *Stata Journal* 10: 165–199.

Kott, P. S. (2001). The delete-a-group jackknife. *Journal of Official Statistics*, *17*(4), 521-526

Lavrakas, P. J., & the AAPOR Cell Phone Task Force. (2010). *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Cell Phone Numbers,* available at http://www.aapor.org.

Lin, I. F., & Schaeffer, N. C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly*, *59*(2), 236–258.

Little, R. J. (1986). Survey nonresponse adjustments. *International Statistical Review*, *54*(1), 3.

Lohr, S. L. (2009). Multiple-Frame Surveys, In Handbook of Statistics: Sample Surveys: Design, Methods and Applications, Vol. 29A, pp. 71-88, doi:10.1016/s0169-7161(08)00004-7. Elsevier.

Merkle, D. M. & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse.* New York: Wiley.

Montaquila, J. M, Brick, J. M, Hagedorn, M. C., Kennedy, C., & Keeter, S. (2007). Aspects of nonresponse bias in RDD telephone surveys. In J. M. Lepkowski et al. (Eds.), *Advances in Telephone Survey Methodology* (pp. 561–586). New York: Wiley.

National Research Council. (2013). *Nonresponse in social science surveys: A research agenda*. Washington, DC: The National Academies Press. doi:10.17226/18293

Oldendick, R. W. and Lambries, D. N. (2011). Incentives for Cell Phone Only Users: What Difference Do They Make? Survey Practice, vol. 4 (1).

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, *70*(5), 737–758.

Pew Research Center (July 2015). Is It Necessary to Reimburse Cellphone Respondents? Technical report available from http://www.pewresearch.org/files/2015/07/2015-07-06_cell-reimbursement-experimentation.pdf

Polson, Diana, James DeFilippis, and Annette Bernhardt. (2011). *Working Without Laws in New York City*. Challenge (05775132) 54, no. 2: 80-108.

Rizzo, L., Brick, J. M., & Park, I. (2004). A minimally intrusive method for sampling persons in random digit dial surveys. *Public Opinion Quarterly*, *68*, 267–274.

Salmon, C. T., & Nichols, J. S. (1983). The next-birthday method of respondent selection. *Public Opinion Quarterly*, *47*, 270–76.

Sangster, R. L. (2012). *Do current methods used to improve response to telephone surveys reduce nonresponse bias?* Washington, DC: Bureau of Labor Statistics, Office of Survey Methods Research. Retrieved from http://www.bls.gov/ore/pdf/st030290.pdf

Schalk, M., McGeeney, K., Kennedy, C., Turakhia, C., & Williams, D. (2015). *Implications for weighted survey estimates when flagged-inactive cell phones are excluded or subsampled.* Presented at the Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

Triplett, T., Wang, K., Safir, A., Steinbach, R., & Pratt, S. (2002). Using a short followup survey to compare respondents and nonrespondents. *ASA Proceedings of the Joint Statistical Meetings* (pp. 3496–3501).

U.S. Census Bureau, Department of Commerce. (2015). *Current Population Survey Contingent Worker Supplement*. Retrieved from: http://catalog.data.gov/dataset/current-population-survey-contingent-worker-supplement.