

## WORKER CLASSIFICATION INFORMATION COLLECTION

### B. Collections of Information Employing Statistical Methods

- 1. Describe (including numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.**

Persons aged 18 or older who live in the United States, have a telephone (landline or cellular) and who were employed for pay in the 30 days prior to the interview will constitute the known respondent universe from which the sample for the Worker Classification Survey will be taken. According to the 2011 National Health Interview Survey (NHIS), 98.2% of U.S. adults live in a household with landline or cellular telephone service (Blumberg and Luke, 2012). According to our analysis of the March 2011 Current Population Survey Annual Social and Economic Supplement (CPS ASEC), 57.7% of the adult population “did work last week, and 60.6% worked in the past 12 months. Exact figures for the “30 day” population are not available because the CPS and the American Community Survey (ACS) both use the “last week” reference period. Based on these figures, however, Abt Associates (the government contractor) can estimate that the total size of the eligible respondent universe is between 132,907,287 and 139,669,156 adults. The lower bound is computed as the product of the total number of adults in the United States according to the 2010 Decennial Census (234,564,071), the proportion of U.S. adults with a landline or cell phone according to the January-June 2011 NHIS estimates (98.2%), and the proportion of U.S. adults who did work for pay last week according to the 2011 CPS-ASEC (57.7%). The upper bound is computed in exactly the same manner, except that the last factor is the proportion of U.S. adults who did work for pay in the last 12 months according to the 2011 CPS-ASEC (60.6%) rather than “last week.”

**Exhibit 1. Respondent Universe and Sample Size for the Worker Classification Survey**

<b>Number of persons in the universe covered by the data collection</b>	<b>Landline RDD sample</b>	<b>Cellular RDD sample</b>	<b>Total sample size</b>
Between 132,907,287–139,669,156	6,000	4,000	10,000

The “last 30 days” reference period is used to ensure that the survey sufficiently captures recently employed workers. Given current economic conditions, 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.

One drawback of the 30 day reference period is it differs from the “last week” reference period used in the Current Population Survey (CPS) and the American Community Survey (ACS). This is relevant because the CPS is the best available benchmark dataset for deriving control totals needed for the nonresponse weighting adjustment in the Worker Classification Survey. As a result of this discrepancy, the control totals from the CPS will describe a slightly different and somewhat smaller target population than the population sampled and interviewed in the Worker Classification Survey. The implications of this discrepancy depend on the number and characteristics of adults who would report “yes” to working in

the past 30 days and “no” to working during the previous week. Considering the relatively small difference in the reference periods (approximately 21 days), it is reasonable to assume that the demographic distributions (e.g., gender, age, race/ethnicity) for the “worked in the past 30 days” population would be highly similar to the demographic distributions for the “worked last week” population.

An alternative source for the benchmarks that also collects detailed employment information is present in the Survey of Income and Program Participation (SIPP). This survey collects information on employment in the 4-months reference period preceding the interview, with specific questions on the dates of the job status transitions, if any, and thus allows bridging the methodological gaps in the reference period definitions. Therefore, Abt plans to use the both CPS and SIPP as sources of the nonresponse adjustment benchmarks. An additional methodological challenge in correctly analyzing the data from the Worker Classification Survey will be accounting for sampling uncertainty in the CPS and SIPP figures used as benchmarks. Since the sample sizes of CPS (n=60,000 HHs per month) and SIPP (n=40,000 HHs in a panel) are not much larger than that of Worker Classification Survey (n=10,000), sampling variability in the benchmarks will have a non-trivial effect on sampling variability of the calibrated estimates. This effect can be accounted for by specially developed replicate weights.

In order to gain some empirical leverage on the issue of the 30 day vs. 1 week reference period, Abt plans to administer the “worked last week” item from the CPS in the extended Worker Classification Survey interview. Collecting these data will facilitate an analysis in which a “worked last week” screen rather than the “worked in the past 30 days” screen is simulated. Specifically, a second set of experimental weights will be developed in which the respondents reporting not having worked last week are excluded, and weights are created based just on those working last week. The weighted survey estimates based on the full survey sample will then be compared to the experimental weighted estimates based on the “worked last week” respondents. Abt expects to observe minimal differences between these two sets of estimates. If, however, meaningful differences are observed (e.g., an average of more than 1.5 percentage points for a set of key survey estimates), then consideration will be given to using the experimental weights as the final survey weights and dropping the respondents who did not work last week from the dataset. This is based on the logic that the experimental weights may be more accurate because the survey target population and the population identifiable in the CPS would be the same.

Respondents will be sampled through a dual frame, landline and cellular random digit dialing (RDD) telephone design. Abt plans to complete 6,000 interviews with respondents sampled through the landline frame and 4,000 interviews with respondents sampled through the cellular frame, for a total of 10,000 interviews. Numbers for the landline sample will be drawn with equal probabilities from active blocks (area code + exchange + two-digit block number) that contain one or more residential directory listings. The cellular sample will be drawn through systematic sampling from 1,000-blocks dedicated to cellular service according to the Telcordia database. The sample will not be stratified beyond the type of phone line (landline/cellular) because the population does not cluster geographically by employment status in a way that could be leveraged to increase the efficiency of the design.

In the survey screener, interviewers will determine whether the household contains at least one eligible adult. An eligible adult is defined as a person 18 years of age or older who did work for pay during the previous 30 days. Households reporting no eligible adults will screen out as ineligible for the extended interview. For households reporting at least one eligible adult, the extended survey respondent will be randomly selected among all of the eligible adults identified in the screener. Within the household, each

eligible adult identified in the screener will have the same probability of being selected 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). This procedure leverages the fact that the large majority of households in the United States have two or fewer adults, and so asking numerous invasive questions can generally be avoided when randomly selecting an adult survey respondent. The procedure was shown to perform well for the Health Information National Trends Survey (HINTS), a large RDD study sponsored by the National Cancer Institute. In implementing it for the Worker Classification Survey, it was necessary to modify the procedure to account for the fact that eligibility is contingent upon working status as well as adult status. In other words, just because the respondent is eligible to complete the screener, they are not necessarily eligible for the extended interview. Abt addresses this issue by adjusting the screener to collect all of the necessary information as shown in the Worker Classification Survey in Attachment C.

Another advantage of the Rizzo, Brick, Park procedure is that it reduces the potential for selection error associated with the last birthday method by limiting its use to a small fraction of cases. Abt will further reduce problems with this approach by randomizing the use of the “last birthday” and the “next birthday” selection procedure. In the Worker Classification Survey, the next/last birthday approach will be implemented only for households with three or more eligible adult workers *or* two eligible adult workers, neither of whom is the screener respondent. Based on the national incidence (from the 2010 American Community Survey) of such households with three or more adults and multiple workers, the proportion of extended interview respondents selected via the last birthday question is expected to be less than 18%. All other extended interview respondents will be selected with certainty (when there is only one eligible adult in the household) or with 50% probability and selected based on a random number generator in the CATI software (when there are two eligible adults).

If the selected extended interview respondent is not the adult who responded to the screener, the interviewer will ask to speak with the selected respondent before administering the extended interview. If the selected respondent is present and available, the screener respondent would simply hand off the phone to the selected respondent. If such a handoff is not possible, the interviewer will ask for the date and time of day when the selected respondent will be available. Interviewers will also inquire as to the best phone number to reach the selected respondent. This procedure will be implemented for both the landline and cell phone samples.

While *within* household selection and resulting handoffs are quite common in landline surveys, they are less common in cell phone surveys. 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 trying to handoff to another person in the household (AAPOR, 2010). Abt proposes to screen all adult household members in the cell sample as well as the landline sample in order to maximize the incidence rate of cases in which an eligible adult is reached. Abt views the challenge of a handoff as more manageable than the inefficiency of excluding an adult worker in a household simply because their spouse or partner’s cell phone was sampled and not their own (for example).

The Worker Classification Survey is the first of its kind, and so there are no historical response rates specific to this data collection. In lieu of historical response rates, Abt compiled figures from studies with

similar sample designs and sponsorship. The surveys and associated response rates (AAPOR (3)) are presented in Exhibit 2. There are several aspects of Exhibit 2 that are worth noting:

- Combined (landline + cell sample) response rates are rare for dual frame RDD surveys. Abt found no response rates of this nature.
- There are essentially no cell sample response rates available for national dual frame RDD surveys conducted for a federal sponsor. The National Immunization Survey cell pilot could be considered an exception.
- The national dual frame surveys conducted for the Pew Research Center provide the only national production (not pilot) cell sample response rates that Abt could find. Pew is a nonpartisan “fact tank” rather than a government sponsor, but they do have a reputation for methodological rigor and transparency within the survey industry.<sup>1</sup> If one incorporates the Pew data, then the cell sample average for the table is 12%. The landline sample average is 25% for all years in the table, though it is 19% based on the landline samples fielded since 2008.

**Exhibit 2. Summary of Response Rates for Random Digit Dial Surveys with Similarity to the Worker Classification Survey**

Survey	Response Rate (AAPOR 3)
<b>Landline RDD</b>	
2005 Health Information National Trends Survey (HINTS)	21%
2007 National Household Education Surveys (NHES: school readiness)	41%
2007 National Household Education Surveys (NHES: parent and family)	39%
2009 California Health Interview Survey (CHIS: adult)	18%
2009 National Household Transportation Survey (NHTS)	20%
2010 Behavioral Risk Factor Surveillance System (BRFSS) national median	36%
2011 September Generations Survey for the Pew Research Center	11%
2011 August Political Survey for the Pew Research Center	11%
<b>Cell RDD</b>	
2008 National Immunization Survey (NIS) Cell Phone Pilot Study	21%
2009 California Health Interview Survey (CHIS: adult)	11%
2011 September Generations Survey for the Pew Research Center	7%
2011 August Political Survey for the Pew Research Center	7%

Abt would note that averages are of somewhat limited value in this instance because the surveys in Exhibit 2 differ from each other in content, burden, calling protocols, and other important aspects. That said, they do serve as a guidepost for what can realistically be achieved under the general survey design.

Applying the findings from this literature review of government-sponsored (where possible) dual frame RDD surveys, the estimate for the response rate for the Worker Classification Survey would be approximately 20% for the landline sample and 18% for the cell sample. Every effort, within the specifications of the study, will be made to exceed this expectation. Relative to the Pew surveys, which provide the only recent national comparisons, the Worker Classification Survey is expected to yield a higher response rate because of the longer field period, more rigorous calling protocol, and federal sponsorship.

**2. Describe the procedures for the collection of information, including:**

- **Statistical methodology for stratification and sample selection;**

<sup>1</sup> The two lead survey researchers at the Pew Research Center were both elected president of the American Association for Public Opinion Research (in 1994 and 2011).

- **Estimation procedure;**
- **Degree of accuracy needed for the purpose described in the justification;**
- **Unusual problems requiring specialized sampling procedures; and**
- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

### **Worker Classification Survey**

In addition to a base weight reflecting the probability of selection, weights will include: 1) a sampling frame integration weight; 2) a non-response adjustment for people who are included on the household roster but do not complete the interview; and 3) a post-stratification adjustment to independent population controls by age, gender, education, race, Hispanic ethnicity, region, and employment status. Estimates of sampling variance reflecting the variance in the weights will be made using replication methods. The weighting and estimation procedures are explained in detail in Attachment D.

Regarding the accuracy of estimates, the design effect should be relatively small because the final sample design does not incorporate any oversampling of workers with certain characteristics. That said, several aspects of the design will contribute to variance in the final weights. Adults living in cell phone only households will be somewhat underrepresented in the survey (approximately 20% of the sample but 34% of the population (Blumberg and Luke 2012)) and, thus, will need to be weighted up. In addition, the random selection of one eligible adult worker in the household and the mixing factor applied to respondents in dual service households (cell and landline) will increase the design effect slightly. Finally, the weighting will serve to adjust for differential nonresponse across key demographic groups, such as those defined by race, ethnicity, age, gender, education, and region. Taking into consideration all of these weighting adjustments, Abt anticipates a design effect of approximately 1.40.

In Exhibit 3 Abt presents a precision estimation for a national proportion of 50% based on self-employed workers and based on non-self-employed workers. In creating Exhibit 3, an estimated proportion of 50% was used because this provides the most conservative results in terms of survey precision for estimated proportions. Also, this is a new survey and so there are no historical point estimates or variance estimates that can be used to inform the precision of the Worker Classification Survey.

**Exhibit 3. Precision of Subgroup Estimates for Self-Employed and Not Self-Employed Workers in the Worker Classification Survey for Estimated Proportions of p=50%**

	Self-employed Workers	Not Self-employed Workers
Estimated proportion ( <i>p</i> )	50%	50%
Design effect ( <i>Deff</i> )	1.40	1.40
Expected nominal size ( <i>n</i> )	1,100	8,900
se( <i>p</i> ) assuming SRS <sup>2</sup> (%)	1.508	0.530
se( <i>p</i> ) actual <sup>3</sup> (%)	1.784	0.627

As shown in the third row of Exhibit 3, the sample design is expected to yield approximately 1,100 interviews with self-employed workers and the remaining 8,900 interviews with those who are not self-employed. Given that no oversampling will be performed, Abt expects an estimated design effect to be roughly the same for estimates based on both of these groups. The fourth row in Exhibit 3 shows the standard error (se) for the survey estimate without taking into account the design effect from weighting. The estimated standard errors (assuming p=50%) is 1.508% for self-employed respondents and 0.530% for those who are not self-employed. If one were interested in the margin of sampling error for a 95% confidence interval, those margins would be ±2.96% for self-employed respondents and ±1.04% for estimates based on those who are not self-employed. The fifth row, by contrast, presents the expected standard error for the estimated proportion using complex survey software. The corrected (that is, using complex survey software) standard errors (for an estimated 50%) are 1.784% for the self-employed and 0.627% for those who are not self-employed.

In Exhibit 4 Abt presents a similarly structured precision estimation for a national proportion of 50% based on workers who are likely misclassified versus those who are likely correctly classified. Likely misclassified workers will be identified in the survey based upon responses to questions about how they are paid (e.g., in cash), whether or not they receive a paystub, whether or not they have deductions, the nature of their employment, etc. It is important to note that the incidence of misclassification is expected to be somewhat higher among self-employed workers than among workers who are not self-employed. That said, not all self-employed workers are likely to be misclassified. Some of the likely misclassified workers represented by the first column of Exhibit 4 will be self-employed but others will not be. Likewise, some of the correctly classified workers represented by the second column of Exhibit 4 will be self-employed while others will not be. In other words, these two characteristics are assumed to be correlated but not completely overlapping. For additional detail on these subgroup definitions, please refer to the criteria for detailed classification questions outlined in the Worker Classification Survey (Attachment C).<sup>4</sup>

<sup>2</sup>  $se_p^{SRS} = \sqrt{pq/n}$  where  $q=1-p$

<sup>3</sup>  $se_p^{actual} = \sqrt{\left[ Deff \left( \frac{pq}{n} \right) \right]}$

<sup>4</sup> Note that the definitions of the subgroups provided above are for the data collection efforts. For analysis, Abt will use the data from the detailed questions to determine which of the “possibly misclassified” respondents are “likely misclassified” versus “correctly classified”.

**Exhibit 4. Precision of Subgroup Estimates for Likely Misclassified Workers and Workers Not Likely to be Misclassified in the Worker Classification Survey for Estimated Proportions of p=50%**

	Likely Misclassified Worker	Correctly Classified Worker
Estimated proportion ( <i>p</i> )	50%	50%
Design effect ( <i>Deff</i> )	1.40	1.40
Expected nominal size ( <i>n</i> )	500	9,500
se( <i>p</i> ) assuming SRS <sup>5</sup> (%)	2.236	0.513
se( <i>p</i> ) actual <sup>6</sup> (%)	2.646	0.607

The sample design is expected to yield approximately 500 interviews with likely misclassified workers, and 9,500 interviews with correctly classified workers. Classification status will be estimated through the combination of answers to survey questions about the nature of the work and duties performed, including:

1. The degree of control exercised by the “employer” or contractor;
2. The extent of the relative investments of the [alleged] worker and employer;
3. The degree to which the worker’s opportunity for profit and loss is determined by the "employer";
4. The skill and initiative required in performing the job; and
5. The permanency of the worker/employer relationship.

The fourth row in Exhibit 4 shows the standard error (se) for the survey estimate without taking into account the design effect from weighting. The estimated standard errors (assuming p=50%) is 2.236% for likely misclassified respondents and 0.513% for respondents who are likely correctly classified. If one were interested in the margin of sampling error for a 95% confidence interval, those margins would be ±4.38% for likely misclassified respondents, and ±1.01% for respondents who are likely correctly classified. The fifth row, by contrast, presents the expected standard error for the estimated proportion using complex survey software. The corrected (that is, using complex survey software) standard errors (for an estimated 50%) are 2.646% for the likely misclassified, and 0.607% for the correctly classified.

These precision levels, facilitated by a relatively large overall sample size (n=10,000), are expected to support comparisons between these groups, as well as comparisons within the groups. Exhibit 5 presents the magnitude of differences that Abt expects to be able to detect with 80% power at an alpha 0.05 level. For example, the first comparison (i.e., column A) in Exhibit 5 is between self-employed respondents and respondents who are not self-employed. After accounting for the expected design effect, Abt anticipates effective sample sizes of roughly n=786 and n=6,357, respectively for these two groups. A two-group  $\chi^2$  test with a 0.05 two-sided significance level will have 80% power to detect the difference between a smaller proportion of 20.0% and a larger proportion of 24.5% (i.e., a difference of 4.5 percentage points)

---

<sup>5</sup>  $se_p^{SRS} = \sqrt{pq/n}$  where  $q=1-p$

<sup>6</sup>  $se_p^{actual} = \sqrt{\left[ Deff \left( \frac{pq}{n} \right) \right]}$

based on these effective sample sizes. Prior to data collection, it is difficult to gauge the level of power for comparisons within the likely misclassified group (e.g., experiences of likely misclassified men versus likely misclassified women) because there are no firm estimates for the actual size of the misclassified population in the U.S. Indeed, this lack of information is one key motivation for the survey. Based on the best available data, Abt expects the survey to yield approximately 500 interviews with likely misclassified workers, for an effective sample size of approximately 357. The 80% power for a difference of proportions test between likely misclassified men versus likely misclassified women is present in column B of Exhibit 5 (see a slightly larger difference of 6.7 percentage points due to the smaller size of the 1<sup>st</sup> group).

**Exhibit 5. Statistical Power for Difference of Proportions Tests Comparing Key Survey Subgroups**

	(A) Self-employed vs. Non-self-employed	(B) Misclassified Men vs. Misclassified Women
Test significance level (alpha)	0.05	0.05
Smaller proportion (p1)	20.0%	20.0%
Larger proportion (p2)	24.5%	26.7%
Power ( % ) (1-beta)	80	80
Effective size of 1st group (n1)	786	357
Effective size of 2nd group (n2)	6,357	6,786

To compute these values we solved for the absolute percent of effective difference (p1-p2) to obtain group size (n1) using the normalized z-value parameters of alpha (α) equal to 0.05 and beta (β) of 0.20 (this is for a power of 80%, i.e., 1-β) and incorporating the ratio (r) of the first group (n1) to the second group (n2). These calculations were based on an adaptation of the following formulas from the mathematics of sample size determination based on unequal sample sizes (see Fleiss 2003):

$$n_1 = \frac{\left[ z_{\alpha/2} \sqrt{(r+1) \bar{p} \bar{q}} + z_{\beta} \sqrt{r p_1 q_1 + p_2 q_2} \right]^2}{r (p_2 - p_1)^2}$$

and,

$$n_2 = \frac{n_1}{4} \left( 1 + \sqrt{1 + \frac{2(r+1)}{n_1 r |p_2 - p_1|}} \right)^2$$

where  $\bar{p} = [(p_1 + r p_2)/(r+1)]$ ,  $\bar{q} = 1 - \bar{p}$ .

- Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses.**

**For collections based on sampling, a special justification must be provided for any collection that will not yield “reliable” data that can be generalized to the universe studied.**



## **Data Collection**

Several recruitment strategies will be used to increase the response rate in the 2012 Worker Classification Survey.

1. Interviewers will make 10 attempts to contact landline cases. More calls will be attempted if contact is made with an eligible household but the interviewer is asked to call back later.
2. Interviews will be conducted during various times of the day and seven days a week to increase the likelihood of finding the respondent at home.
3. Respondents will be provided with the option of scheduling the interview at the time that is convenient for them.
4. For soft-refusals, “interview converters” who have extensive training in telephone interviewing and converting non-responders will be used to increase the response rate.

The determination of total number of call attempts for the landline took into account several factors and ultimately represents a balance between time, cost, and courtesy considerations of respondents. Time considerations include not only the total number of weeks in the field period but also the optimal time periods in which to reach respondents. Cost considerations included: a) the expected gain in response from increased call attempts; and b) costs associated with the screening effort. While some studies show gains in response rate with increased attempts, these depend upon several factors, including the topic or salience of the study for the respondent, age of the respondent and other factors. Gains in response rate can decline after 15 call attempts (Triplett 2002). Finally, Abt is highly aware of the reluctance to participate in surveys, the use of call screening devices and reported feeling of harassment among an over-surveyed population. Our decision to limit call attempts to 10 for landlines (and 8 for cell phones, discussed below) reflects our best estimate of the optimal balance of these considerations.

This recruitment design may prove to be overly intrusive for prospective cellular frame respondents; therefore Abt will use an 8-call design for the cellular sample (whether or not a request for callback is made). Our protocol for cell phone calls follows recommended industry protocol as outlined by The American Association for Public Opinion Research (AAPOR 2008) and also discussed by Triplett (2002). Abt will make one conversion attempt on all soft refusals on all landline sample cases. In adherence to the recommendations of the American AAPOR Cell Phone Task Force, Abt will not attempt refusal conversion for soft refusals on cell cases:

Logic and anecdotal evidence to date suggest that refusal conversion attempts to cell phone respondents should be of a limited nature so as to reduce the potential for further agitating [cell phone respondents]. This is in large part a result of likely reaching the same respondent who previously refused rather than reaching some other member of the sampling unit (household), as often is the case when trying to convert refusals in RDD landline surveys. (AAPOR, 2010)

## **Analysis of Non-Response**

Abt will evaluate non-response in the Worker Classification Survey in four ways: 1) a non-response follow-up survey (NRFU), 2) a comparison of easy-to-reach versus harder-to-reach respondents, 3) fitting response propensity models, and 4) comparing survey estimates with external benchmarks. These four analyses are detailed below.

## **1. *Non-Response Follow-up Survey (NRFU) and Comparative Analysis of NRFU and Main Sample Responses***

The NRFU will collect information on employees who fail to respond to the Worker Classification Survey and provide insight into whether the nonrespondents differ from the respondents on the characteristics of interest (e.g., work experiences and benefits). Specifically, interviewers will call back a subsample (n=500) of households that declined the original survey (See Attachments A1 and A2). These interviewers will attempt to recruit an eligible employee to complete a shortened interview featuring a \$20 remuneration. Details on incentives are provided in Part A.

Incentives are a common feature in NRFU surveys because, by definition, the NRFU sample did not cooperate with the original survey, and so a major change in the recruitment protocol is required to elicit cooperation in the NRFU. Zimowski and colleagues (1997) noted in their report to the Federal Highway Administration (FHWA-PL-98-029) that large monetary incentives (e.g., \$20 to \$50) are a common element of NRFU designs for household surveys. For example, Peytchev et al. (2009) documented how a \$20 incentive was used in a successful NRFU to the National Intimate Partner and Sexual Violence Survey for the Centers for Disease Control and Prevention.

In addition, all landline sample cases that can be matched to an address (through reverse lookup) will receive a letter encouraging them to cooperate with the interview. Abt expects to complete approximately 200 NRFU interviews. This will provide a sufficient case base for meaningful nonresponse analysis.

Abt will compare the employment characteristics of Worker Classification Survey respondents with the characteristics of NRFU respondents. This analysis will provide insights about the direction and magnitude of possible nonresponse bias. The NRFU estimates will be compared with both weighted and unweighted estimates from the main survey. Abt will investigate whether any differences remain after controlling for major weighting cells (e.g., within race and education groupings). If weighting variables eliminates any differences, this suggests that the weighting adjustments discussed in Attachment D will reduce nonresponse bias in the final survey estimates. If, however, the differences persist after controlling for weighting variables, then this would be evidence that the weighting may be less effective in reducing non-response bias.

## **2. *Comparative Analysis of Easier to Reach vs. Harder to Reach Respondents.***

The second technique that Abt will use to assess nonresponse bias is an analysis of the level of recruitment difficulty. This analysis will compare the unweighted classification and employment characteristics of respondents who were easy to reach with respondents who were harder to reach. The level of difficulty in reaching a respondent will be defined in terms of the number of call attempts required to complete the interview and whether the case was a converted refusal. In some studies, this is described as an analysis of “early versus late” respondents, though Abt proposes to also explicitly incorporate refusal behavior. If the classification and employment characteristics of the harder-to-reach cases are not significantly different from characteristics of the easy-to-reach cases, this would suggest that survey estimates may not be substantially undermined by non-response bias. The harder-to-reach cases serve as proxies for the non-respondents who never complete the interview. If the harder-to-reach respondents do not differ from the easy-to-reach ones, then presumably the sample members never reached would also not differ from those interviewed. Support for this “continuum of resistance” model is inconsistent (Lin and Schaeffer 1995; Montaquila et al. 2008), but it can still be a useful framework for

assessing the relationship between level of effort and non-response bias. An alternative method to jointly model response propensity and the indicators of interest was presented by Peress (2010).

In the easy-to-reach versus hard-to-reach analysis, Abt will define the easy/hard dimension in three ways: (1) in terms of ease of contactability, as defined by the number of calls required to complete the interview; (2) in terms of amenability, as defined by whether or not the case was a converted refusal; and (3) a in terms of both contactability and amenability, as defined by a hybrid metric combining number of call attempts and converted refusal status. This analysis will provide some evidence as to which, if either, of these two mechanisms may be leading to nonresponse bias in survey estimates.

### **3. *Estimating Response Propensity Models.***

The third technique that Abt will use to assess nonresponse bias is response propensity modeling (Little 1986; Groves and Couper 1998; Olson 2006). Response propensity is the theoretical probability that a sampled unit will respond to the survey request. Many respondent characteristics can influence response propensity. Disentangling these effects requires multivariate modeling.

In order for a response propensity model to be informative, the researcher must know the values for respondents and non-respondents 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 non-respondents. For the Worker Classification Survey, Abt proposes to fit a response propensity model predicting the probability of completing the extended interview conditional on having completed the screener. This analysis will be based only on households for which Abt has a completed screener. The model will include an indicator for sampling frame, an indicator for whether or not the household ever refused the interview, and a log-transformed variable for the number of call attempts made to the household. Our preliminary plan is, thus, for the survey response propensity model to predict survey response using the following variables:

- Respondent gender
- Number of adults in the household
- Number of adults in the household who did work for pay in the last 30 days
- Census region
- Sampling frame (landline RDD or cellular RDD)
- Household refused the interview once
- Number of call attempts made to the household
- Day of week called
  
- Time of day called

The estimated logistic regression model will be used to create summary “response propensity scores” (i.e., the predicted probability from the logistic regression model) that estimate how likely the selected respondent was to participate in the survey, regardless of the actual outcome. Abt will create five groups (response propensity classes) from the response propensity scores. In a well-specified model, respondents and nonrespondents will be similar on the characteristics of interest within each class, and likelihood of survey participation will vary across the classes.

The response propensity model will help us to identify the most powerful predictors of response when all available predictors are tested simultaneously. If employment-related variables show a significant association with response to the extended interview (after controlling for other factors), this would be

evidence of possible non-response bias. If, however, the employment predictors do not have a significant effect, this suggests that the screener non-response adjustment, described in Attachment D, will be effective in reducing nonresponse bias. Similarly, comparisons of the respondent characteristics across the five response propensity classes will also provide insight on which types of screened respondents were most likely to complete the extended interview and which types were less likely to do so.

For the response propensity modeling, Abt plans to condition on contacted households and model the probability of cooperating with the interview. This approach is based on the fact that the Worker Classification Survey features an RDD sample design, which means that there is little information available on non-contacted households. Given this lack of data, models predicting the probability of contact would not be very informative. On the other hand, Abt expects a moderate number of households to complete the screener, but to be lost in an attempt to interview the randomly chosen respondent. In that case, Abt can model nonresponse as a function of information collected in the screener.

#### **4. Comparisons to External Benchmarks.**

The final analyses Abt will conduct for non-response is a comparison of survey estimates to national benchmarks. One limitation of the aforementioned techniques is that they analyze only a subset of all non-respondents to the survey. The NRFU analysis relies on the NRFU participants as proxies for all non-respondents; the level of effort analysis relies on the “harder-to-reach” respondents as proxies for all non-respondents; the response propensity model captures only variation between the screened extended interview respondents and the screened extended interview non-respondents.

One approach for evaluating the total level of nonresponse bias in a survey is to compare the weighted survey estimates with external estimates based on a “gold standard” survey. The “gold standard” survey should feature a more rigorous protocol (e.g., area-probability sampling with in-person interviewing) and a higher response rate than the target survey (the Worker Classification Survey). Critically, the gold standard survey and the target survey must feature one or more questions administered in a highly similar manner. Estimates based on these questions can be compared. By virtue of its more rigorous design, the estimates from the gold standard survey are assumed to contain less non-response bias than those from target survey.

Differences in the question wording or mode of administration, however, may confound the comparison. Differences in population coverage between the gold standard and target survey may also confound the comparison. In light of these considerations, results from external comparisons must be interpreted with caution.

Abt proposes to compare weighted Worker Classification Survey estimates with those from the CPS. Examples of possible analytic variables administered in both surveys are: marital status, average hours worked per week, and labor union membership.

#### *Non-Contact versus Non-Cooperation*

Where possible, Abt will treat non-contact and non-cooperation as two distinct outcomes. Non-contact and non-cooperation are generally considered to reflect two different dimensions on which sample households can be placed (Stinchcombe et al. 1981; Goyder 1987; Groves and Couper 1998; Lynn et al 2002). As noted by Stoop (2005), decomposing non-response into these two different dimensions can be analytically useful in several ways:

- When trying to enhance response rates different measures apply to improving contactability and improving cooperation;
- When comparing surveys over time or across countries different nonresponse rates and a different composition of the non-respondents (non-contacts and refusals) may be confounded with substantive differences;
- When estimating response bias or adjusting for nonresponse, knowledge about the underlying nonresponse mechanism (noncontact, refusal) should be available as contacting and obtaining cooperation are entirely different processes;
- When estimating response bias or adjusting for nonresponse, information on the difficulty of obtaining contact or cooperation is often used assuming that “difficult” respondents are more like final refusers than easy respondents.

Finally, the benchmark comparison analysis is designed to compare survey estimates with external benchmark estimates. The outcomes of interest are NET differences between these two sets of estimates. In this analysis nonresponse must be treated in the aggregate. Decomposing non-contact and non-cooperation is not possible when evaluating estimates based on the responding sample.

#### *Summary of Non-Response Analyses*

While these analyses rely on imperfect assumptions, all are standard techniques for assessing potential non-response error. No single nonresponse analysis for this study can be definitive because the true scores of the non-respondents are not known. That said, by using several different methodologies (non-response follow-up analysis, easy-to-reach versus hard-to-reach comparisons, response propensity models, and comparisons of estimates to external benchmarks), Abt draws some meaningful conclusions about the level of risk to survey estimates from non-response bias. This information may also be helpful in modifying nonresponse weighting adjustments to reduce bias to the extent possible.

#### **4. Describe any tests of procedures or methods to be undertaken.**

**Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.**

Abt conducted cognitive tests on the worker survey with nine volunteer respondents in Chicago. These purposively selected respondents included workers who were potentially misclassified in order to test the applicability of questions on different types of workers (employees and self-employed). The respondents came from a diversity of backgrounds and education levels in order to test applicability of the questions for different types of workers (salaried versus hourly, for example) and to capture the range of possible comprehension issues. The survey included in this package reflects the findings from those interviews. The current version of the Worker Classification Survey is included as Attachment C.

The in-depth interviews are semi-structured, open-ended interviews; testing does not apply.

#### **5. Disclosure Limitation Methods**

A public use file (PUF) for the Worker Classification Survey will be made available after completion of the data collection. Abt will implement a disclosure limitation protocol so that the PUF fully protects

respondent privacy. The risk of disclosure in the Worker Classification Survey is extremely low for the following reasons:

1. **No sampling frame information, contact information, or other personal identifying information will be included in the PUFs.** It will not be possible to link the survey records to administrative data. Each record will have a unique case ID, but that value will be randomly assigned and will carry no information about the record.
2. **No geographic variables will be included.**
3. **The survey is cross-sectional rather than longitudinal, and it does not feature clustering in the sample design.**
4. **The sampling fraction is extremely small.** In the cell RDD and landline RDD frames, the expected sampling fractions are 0.00024 and 0.00068, respectively. Surveys with very small sampling fractions entail a lower risk of disclosure than surveys with larger sampling fractions.
5. **Sample design variables will not be released.** Replicate weights will be provided so that data users can account for the complex nature of the sample design. When replicate weights are provided, it is not necessary to provide sample design variables, such as PSU or stratum.

According to guidelines published by the Federal Committee on Statistical Methodology Report on Statistical Disclosure Methodology (2005) and the National Center for Health Statistics Staff Manual on Confidentiality (2004) these properties of the Worker Classification Survey reduce the risk of disclosure limitation.

Below Abt describes the specific additional steps that will be taken to ensure that the data released in the PUF fully protect respondent privacy. Abt will employ variable suppression, rounding, top-coding, bottom-coding, and other data coarsening as needed so that no identifying values are released in the PUF. Abt prefers these techniques over data swapping because for variables like respondent age, recoding has been shown to improve protection more than random data swapping (Reiter 2005).

Basic demographic variables are often the most susceptible to matching. In order to make sure that no identifying values are released, Abt will make the following manipulations to the Worker Classification Survey dataset. These manipulations are in addition to the disclosure limitation procedures mentioned above.

**QEDUCATION\_1** Abt will collapse the 16 education cells for this question into the following six cells: “Less than 1<sup>st</sup> grade” through “7<sup>th</sup> or 8<sup>th</sup> grade” into “Less Than High School,” “9<sup>th</sup> Grade” through “12<sup>th</sup> Grade No Diploma” into “Some High School,” “High school Grad-Diploma or equivalent (GED) as “High School Graduate,” “Some college but no degree” through “Associate Degree – Academic Program” into “Some College/Associate’s Degree,” “Bachelor’s Degree” as “College Graduate,” and “Master’s Degree” through “Doctorate Degree” into “Graduate School.”

**QINCOME** The variables QINC\_H and QINC\_J will be suppressed (not included in the PUF). These variables detail relatively small income categories. The lowest income classification will, thus, be under \$20,000 and the highest will be \$100,000 or above. Specifically, Abt will bottom-code income. The top code (\$100,000 or above) is not a rare characteristic and will not be manipulated.

QRACE	The “Native Hawaiian or Pacific Islander” cell will be collapsed with the cell for “Some other race.” The incidence of that group is very low (0.3% of the US population), meaning that it could potentially be an identifying variable if used in conjunction with other variables.
QZIP	ZIP code (and all other geographic or personally identifying information) will not be released.
QCELL	The sampling frame variable will not be released.
QSTATE	State of job will not be released.

Screener data (S1 through and S7) collected for household members other than the selected respondent will not be included in the PUF. The main sections of the questionnaire contain a series of questions asking about the start year of the main job and changes in work status (year). Given that this type of information may be known by numerous people in the respondent’s life and some combinations of values may be quite rare, these variables pose a disclosure risk. Abt proposes to suppress all variables containing the year of a job or work status beginning or ending. Instead, Abt will report the duration of the job and/or work status in a specially-constructed variable.

QMAINJOB	Hours per week usually work at main job, reported in the following ranges: 0-20 hours 20-40 hours > 40 hours
QMAINJOB_3	Year began main job
QSELF_NUM	Number of clients in last thirty days
QCHANGE_EE2	Year became an employee
QCHANGE_NE2	Year became self-employed
QFIRMSIZE	Number of persons working for employer/you
QAGE	Respondent age – will be top coded so that larger values are not personal identifiable information.

We will also suppress all variables related to earnings, for example: ERNP, ERNPR, ERNUOT, ERNOTP, ERNOTHD, ERNOTHC. In addition to these pre-identified data edits, Abt will review the final data for rare responses. As necessary, Abt will recode so that no single response category or combination of closely related response categories has an unweighted frequency below five.

**6. Provide the name and telephone number of individuals consulted on statistical aspects of the design, and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.**

Abt SRBI has been contracted to conduct the Worker Classification Survey. The individuals assigned to this project include:

Jacob Klerman, Principal Associate, (617) 520-2613

Kelly Daley, PhD, Vice President, (312) 529-9703  
Alyssa Pozniak, PhD, Associate (617) 520-2455  
Courtney Kennedy, PhD, Vice President, (617) 386-2604

The following individuals were consulted on statistical aspects of the design:

Courtney Kennedy, PhD, Vice President, (617) 386-2604  
Charles DiSogra, PhD, Senior Vice President, (617) 386-4070  
Stanislav Kolenikov, PhD, Senior Survey Statistician, (617) 386-2621

In addition, the Project Officer for DOL is Jonathan Simonetta, (202) 693-5085



## References

- American Association for Public Opinion Research (AAPOR). 2008. "AAPOR Cell Phone Task Force: 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." Available at [http://www.aapor.org/uploads/Final\\_AAPOR\\_Cell\\_Phone\\_TF\\_report\\_041208.pdf](http://www.aapor.org/uploads/Final_AAPOR_Cell_Phone_TF_report_041208.pdf)
- American Association for Public Opinion Research (AAPOR). 2010. "New Considerations for Survey Researchers When Planning and Conducting RDD and Other Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers." Available at [http://aapor.org/AM/Template.cfm?Section=Cell\\_Phone\\_Task\\_Force&Template=/CM/ContentDisplay.cfm&ContentID=2818](http://aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force&Template=/CM/ContentDisplay.cfm&ContentID=2818).
- Blumberg S.J., Luke J.V. 2012. "Wireless substitution: Early release of estimates from the National Health Interview Survey, January–June 2012." National Center for Health Statistics. Available at <http://www.cdc.gov/nchs/nhis.htm>.
- Federal Committee on Statistical Methodology Report on Statistical Disclosure Methodology. 2005. "Statistical Policy Working Paper 22 - Report on Statistical Disclosure Limitation Methodology." Office of Management and Budget. Available at [http://www.fcsm.gov/working-papers/SPWP22\\_rev.pdf](http://www.fcsm.gov/working-papers/SPWP22_rev.pdf).
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: Wiley. pp 69-77.
- Goyder, J. 1987. *The Silent Minority*. Boulder, CO: Westview Press.
- Groves, R. and M. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York, NY: John Wiley & Sons, Inc.
- Lin, I. and N. Shaeffer. 1995. "Using Survey Participants to Estimate the Impact of Nonparticipation." *Public Opinion Quarterly* 59: 236-258.
- Little, R. 1986. "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review* 54:139-57.
- Lynn, P. and P. Clarke. 2002. "Separating Refusal Bias and Non-Contact Bias: Evidence from UK National Surveys." *Journal of the Royal Statistical Society Series D (The Statistician)* 51(3): 319-333.
- Montaquila, J., J. Brick, M. Hagedorn, C. Kennedy, and S. Keeter. 2008. "Aspects of Nonresponse Bias in RDD Telephone Surveys," in *Telephone Survey Methodology*, edited by J. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. Lavrakas, M. Link, R. Sangster. New York, NY: John Wiley & Sons, Inc.
- National Center for Health Statistics. 2004. "Staff Manual on Confidentiality." Available at <http://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf>.
- Olson, K. 2006. "Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias." *Public Opinion Quarterly* 70: 737-758.
- Peress, M. 2010. "Correcting for Survey Nonresponse Using Variable Response Propensity." *Journal of the American Statistical Association*. 105 (492), 1418-1430

- Peytchev, A., R. Baxter, and L. Carley-Baxter. 2009. "Not All Survey Effort is Equal: Reduction of Nonresponse Bias and Nonresponse Error." *Public Opinion Quarterly* 73: 785-806.
- Reiter, J. 2005. "Estimating Risks of Identification Disclosure in Microdata." *Journal of the American Statistical Association* 100: 1103 - 1113.
- 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.
- Stinchcombe, A. L., C. Jones, and P. Sheatsley. 1981. "Nonresponse Bias for Attitude Questions." *Public Opinion Quarterly* 45: 359-379.
- Stoop, I.A.L. 2005. *The Hunt for the Last Respondent*. Social and Cultural Planning Office, The Hague. Netherlands. USA: Transaction Publishers, New Brunswick.
- Triplett, Timothy. 2002. "What Is Gained from Additional Call Attempts and Refusal Conversion and What Are the Cost Implications?" Washington, DC: The Urban Institute. Available at: <http://mywebpages.comcast.net/ttriplett13/tncpap.pdf>.
- United States Census Bureau. 2010. "American Community Survey."
- Zimowski, M., T. Tourangeau, and R. Ghadialy. 1997. "Nonresponse in Household Travel Surveys (FHWA-PL-98-029)," prepared for the Federal Highway Administration.