

Attachment D: Weighting and Estimation Procedures

Employee Survey: Weighting Protocol

The Employee Survey features a national, dual-frame landline and cellular random digit dial (RDD) probability sample design. The landline and cell phone sampling frames overlap, as some employees age 18 years and older have both a residential landline in their household and a cell phone (dual users). While many dual frame RDD surveys treat the cell phone as a personal device, this survey will treat the cell phone as a household device. Interviews will conduct a household roster to identify all eligible workers in both the landline and cell phone samples. The weighting procedures described here account for the overall probability of selection, sampling frame integration, and appropriate non-response and post-stratification ratio adjustments.

The reciprocals of the overall selection probabilities, which will vary depending on whether the employed person is a: 1) leave taker, 2) leave needer, or 3) other employed individual, are referred to as “base weights.” The base weights are the product of several components: 1) the inverse of the selection probability of the sample telephone number; 2) an adjustment for the number of voice-use landline numbers in the household (equal to 0 for cell-only households) and the number of non-business adult-use cell phones in the household (equal to 0 for landline-only households); 3) in the case of employed persons, the inverse of the subsampling rate that is used.

The third component of the base weight consists of two multiplicative factors. The first factor arises from the classification of all eligible employed adults in the household into the three groups listed above. Among the groups present in a household (i.e., groups containing one or more persons), one group will be selected with a known probability determined by the group subsampling probability used in the sample replicate that the telephone number comes from. The second factor arises from the random selection of one person from the sampled group. This factor equals one if the sampled group only contains one person. We may decide to put a cap on the maximum value allowed for the number of persons in the sampled group to avoid extreme design weight values.

Due to the overlap in the landline and cellular RDD frames, a frame integration weight (or “compositing factor”) is needed to combine the two sample components. The frame integration weight will be based on the ratio of the effective sample sizes of the landline and cell phone samples. Specifically, the frame integration weight for dual user (landline and cell) respondents in the landline sample will be the ratio of the effective number of dual service landline cases to the total effective number of dual service cases in both samples. Similarly, the frame integration weight for the dual service cell sample cases will be the

ratio of the effective number of dual service cell sample cases to the total effective number of dual service cases in both samples. Landline-only and cell-phone-only cases will be assigned a frame integration weight of 1. The product of the base weights and the frame integration weights are referred to as “design weights” and will be used as the input weight for the non-response adjustment.

The screener non-response adjustment will need to be calculated at the household level for telephone numbers for which no screening interview is conducted. The adjustment cells will be based on region for both the cell and landline samples. The interview non-response adjustment will account for persons who have been listed on a household roster and selected for the sample, but an interview was not completed. The person-level interview non-response adjustment will be done within specified non-response adjustment weighting classes of persons, and these factors will be applied to the design weights adjusted for household non-response, to compensate for unit non-response. The weighting classes will be defined by the age and gender of the sampled person. Interviewers will attempt to collect these variables on all adults on the roster of the survey. For item non-response on weighting class variables (i.e., age and gender), missing data will be filled using hot deck imputation methodology.¹

The non-response-adjusted weight, w_{gi}^A , for the i -th responding eligible person in weighting class g will be computed as

$$w_{gi}^A = w_{gi}^B \left(\frac{N_{Rg} + N_{Ng}}{N_{Rg}} \right) \quad (1)$$

where w_{gi}^B is the weight that includes the design weight and the household screener non-response adjustment, N_{Rg} is the weighted sum of eligible responding persons in weighting class g , and N_{Ng} is the weighted sum of non-responding persons in weighting class g .

To help reduce possible residual non-response and non-coverage errors, the final estimation weights will also include a post-stratification adjustment to reflect the most recent population information available. The target population for the Employee Survey is adults (age 18 and older) residing in the U.S. who have been employed in the last 12 months. We will compute control totals for this population from the March 2011² Current Population Survey, Annual Social and Economic Supplement (CPS-ASEC) micro datafile.

¹ Specifically, Ellis, Bruce. 2007. “A Consolidated Macro for Iterative Hot Deck Imputation.” Proceedings of the 20th NorthEast SAS Users Group Conference, Baltimore, MD.

² We will compute the population benchmarks from the most up-to-date CPS monthly micro dataset that is publically available. We expect this to be the March 2012 dataset, with a reference period of the January 2011 to December 2011; i.e., the 12 months prior to January 2012. Since the Employee Survey will ask about the previous 12 months and have a field period of January 2012 to May 2012, there will be some, but not exact, overlap with the field period of the CPS dataset. Consequently, there will be a slight misalignment between the “last 12 months”

Specifically, we will use the CPS to estimate the total size of the target population, along with demographics distributions for gender, age, education, race, Hispanic ethnicity, and region. In addition, we will use the 2010 National Health Interview Survey, to compute the distribution for telephone service groups.

The post-strata will be constructed using the relevant demographic variables in the survey dataset. The proposed initial post-strata are as follows and collapsing rules are detailed below:

GENDER	(1=Male, 2=Female)
AGE	(1=18 to 29, 2=30 to 39, 3=40 to 49, 4=50 to 59, 5=60 and above)
EDUCATION	(1=High school graduate/GED or less, 2=Some college or Associate degree, 3=Bachelor’s degree, 4=Master’s, Doctorate, or professional school degree (e.g., MD, DDS, JD))
RACE_ETHNICITY	(1=White only non-Hispanic, 2=Black only non-Hispanic, 3=Asian only non-4=Hispanic, 5=Other race or mixed race non-Hispanic, 6=Hispanic)
REGION	(1=Northeast, 2=Midwest, 3=South, 4=West)
PHONESERVICE	(1=Cell phone only, 2=Landline only, 3=Dual service)

We will fill missing data on these weighting variables using hot deck imputation methodology.³ After the post-strata are created but prior to raking, we will check the distribution of all six dimensions in the survey dataset. If any of the cells defined above contains less than 5% of the unweighted sample, that cell will be collapsed with the most appropriate cell in the dimension. For example, if fewer than 5% of the respondents are Asian only non-Hispanic, then this cell will be collapsed with the cell for other race or mixed race non-Hispanic. The purpose of this collapsing is to avoid excessively large weight values, which reduce the precision of survey estimates. This procedure also helps to ensure that the raking algorithm will converge and generate a solution.

The general approach for making the post-stratification adjustment will be as follows. Let Y_k denote the aggregate number of persons in post-stratum k from the population controls and let

$$\hat{Y}_k = \sum_{i=1}^{n_k} w_i^A \tag{2}$$

used to define the Employee Survey target population and the “last 12 months” for which employment is measured in the CPS dataset. Especially since the primary focus of the survey is not on short-term changes, we do not expect, however, that the misalignment will be large enough to compromise the validity of the CPS data as control totals for this survey.

³ Ellis, *op. cit.*

denote the corresponding estimate from the sample, where w_i^A is the interview non-response-adjusted weight defined previously and n_k is the number of responding persons in post-stratum k . The final weight for person i in post-stratum k will then be computed as

$$w_{ki}^F = w_i^A \left(\frac{Y_k}{\hat{Y}_k} \right). \quad (3)$$

The above adjustment has the effect of forcing the weighted estimate of the aggregate number of employees in a post-stratum to agree with the corresponding independent population control. Given that the sample will be post-stratified to several variables, we plan to use raking ratio estimation to calculate the final weights.

Employer Survey: Weighting Protocol

The Employer Survey sample of business establishments will be drawn from the Dun’s Market Identifiers (DMI) file. In order to make statistically valid estimates from the survey results, it will be necessary to weight the sample data. The weight to be applied to each responding business establishment is a function of the overall probability of selection, and appropriate non-response and post-stratification ratio adjustments. The reciprocals of the overall selection probabilities, which will vary depending on the size of the establishment, are referred to as “base weights.” These weights will produce unbiased estimates if there is no non-response in the survey. Since some non-response is inevitable, adjustment factors will be calculated within specified non-response adjustment weighting classes of establishments, and these factors will be applied to the base weights to compensate for unit non-response. The weighting classes will be created with the goal to minimize the bias due to non-response. Item non-response will be addressed using appropriate imputation methodology. Regression-based imputation is usually appropriate for continuous variables, and hot-deck imputation is often used for categorical variables.

The non-response-adjusted weight, w_{gi}^A , for the i -th responding eligible establishment in weighting class g will be computed as

$$w_{gi}^A = w_{gi}^B \left(\frac{N_{Rg} + N_{Ng}}{N_{Rg}} \right). \quad (4)$$

where w_{gi}^B is the base weight, N_{Rg} is the weighted sum of eligible responding establishments in weighting class g , and N_{Ng} is the weighted sum of eligible non-responding establishments in weighting class g . Sample establishments for whom eligibility status is unknown will be excluded from the above non-response adjustment.

To help reduce possible under-coverage errors in the DMI sampling frame and reduce possible non-response bias, the final estimation weights will also include a post-stratification adjustment to reflect the most recent population information available from the Quarterly Census of Employment and Wages (QCEW). The adjustments will be made within broad classes (post-strata) such as Census Region, broad NAICS category, and size of establishment. These post-strata will be defined based on analysis of the distributions of these variables in the County Business Patterns data.

The post-strata will be constructed using the relevant variables in the survey dataset. The proposed initial post-strata are as follows:

REGION	(1=Northeast, 2=Midwest, 3=South, 4=West)
NAICS_SIZE	(1=NAICS Group I Size 1 to 10, 2=NAICS Group I Size 11 to 249, 3=NAICS Group I Size 250+, 4=NAICS Group II Size 1 to 10, 5=NAICS Group II Size 11 to 249, 6=NAICS Group II Size 250+, 7=NAICS Group III Size 1 to 10, 8=NAICS Group III Size 11 to 249, 9=NAICS Group III Size 250+, 10=NAICS Group IV Size 1 to 10, 11=NAICS Group IV Size 11 to 249, 12=NAICS Group IV Size 250+) See Exhibit 2 in Part B for the NAICS codes represented by each group

We will fill missing data on these weighting variables using hot deck imputation methodology. After the post-strata are created but prior to raking, we will check the distribution of all six dimensions in the survey dataset. If any of the cells defined above contains less than 5% of the unweighted sample, that cell will be collapsed with the most appropriate cell in the dimension.

The general approach for making the post-stratification adjustment will be as follows. Let Y_k denote the aggregate number of employees in post-stratum k as given in the most recent County Business Patterns publication and let

$$\hat{Y}_k = \sum_{i=1}^{n_k} w_{ki}^A y_{ki} \tag{5}$$

denote the corresponding estimate from the sample, where w_{ki}^A is the non-response-adjusted weight defined previously, y_{ki} is the observed number of employees of establishment i in post-stratum k , and n_k is the number of responding establishments in post-stratum k . The final weight for establishment i in post-stratum k will then be computed as

$$w_{ki}^F = w_{ki}^A \left(\frac{Y_k}{\hat{Y}_k} \right). \tag{6}$$

The adjustment above has the effect of forcing the weighted estimate of the aggregate number of employees in a post-stratum to agree with the corresponding County Business Patterns number.

Variance Estimation

An important advantage of probability sampling methods is that they permit the calculation of the sampling errors (variances) associated with the survey estimates, and thus, with reasonably large samples, provide an objective way of measuring the reliability of the survey results. The sampling errors may be calculated directly using analytical variance formulas, or by a replication procedure such as “jackknife” or balanced half-sample replication. Use of the direct variance estimators is relatively straightforward for simple linear estimates such as the “expansion” estimate of a population total, but the variance formulas can be complex for nonlinear statistics, or estimates based on complex sample designs. On the other hand, replication methods provide a relatively simple way of estimating variances. In addition, the impact of the non-response and post-stratification adjustments can be more easily reflected in the variance estimates obtained by replication methods.

Therefore, for both the Employee Survey and the Employer Survey, we propose to calculate the standard errors by using a repeated replication technique, specifically paired unit jackknife repeated replication (JK2). As with any replication method, JK2 involves constructing a number of sub-samples (replicates) from the full sample, and computing the statistic of interest for each replicate.

Specifically, let $\hat{\theta}$ be any estimate from the survey. For example, if $\hat{\theta}$ is an estimate of the proportion of employees who took family leave in the past year, then $\hat{\theta}$ has the following form

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i^F y_i}{\sum_{i=1}^n w_i^F x_i} \tag{7}$$

where w_i^F is the final estimation weight as defined earlier, y_i is the number of employees who took family leave in the past year, and x_i is the number of employees in establishment i . Further, let $\hat{\theta}_{(j)}$ be the corresponding estimate for JK2 replicate j . The estimated variance of $\hat{\theta}$ can then be computed from the formula

$$v(\hat{\theta}) = c \sum_{j=1}^R (\hat{\theta}_{(j)} - \hat{\theta})^2 \tag{8}$$

where the summation extends over all R replicates. Under JK2, the constant, c , in the equation above equals 1.

In the JK2 replication method, adjacent pairs of sampled telephone numbers are treated as having been sampled from the same stratum. Each pair of sampled telephone numbers is treated as an implicit stratum, where each such stratum is defined by the sort order used in the sample selection of telephone numbers.

Under the JK2 method, the number of replicates is equal to the number of variance estimation strata. The choice of the number of variance estimation strata is based on the desire to obtain an adequate number of degrees of freedom to ensure stable estimates of variance while not having so many as to make the cost of computing variance estimates unnecessarily high. Generally, at least 30 degrees of freedom are needed to obtain relatively stable variance estimates. A number greater than 30 is often targeted because there are other factors that reduce the contribution of a replicate to the total number of degrees of freedom, especially for estimates of subgroups.

For each of the 2012 FMLA surveys, we propose to create 80 variance estimation strata. For the Employee Survey, the sampled telephone numbers will be sorted by frame (landline RDD or cellular RDD) and, within each frame, sorted by region. For the Employer Survey, the sampled telephone numbers will be sorted by the 28 strata (establishment size x NAICS group) defined in Part B.

Next, adjacent sampled telephone numbers will be paired to establish initial variance estimation strata (the first two sampled phone numbers will be the first initial stratum, the third and fourth sampled telephone numbers were the second initial stratum, etc). Each telephone number in the pair will be randomly assigned to be either the first or second variance unit within the variance stratum. Each pair will be sequentially assigned to one of 80 final variance estimation strata (the first pair to variance estimation stratum 1, the second to stratum 2, etc.). As a result, each variance stratum will have approximately the same number of telephone numbers. The same process will be followed for each sampling stratum.

Once the variance strata are created, the replicate weights can be created. The full replicate weights are computed by first modifying the full sample base weights. The replicate base weight for replicate k for record i is

$$w_i^{F(k)} = \begin{cases} 2w_i^F, & \text{if } i \text{ is in variance stratum } k \text{ and variance unit 1} \\ 0, & \text{if } i \text{ is in variance stratum } k \text{ and variance unit 2} \\ w_i^F, & \text{if } i \text{ is not in variance stratum } k \end{cases} \quad (8)$$

The same sequence of weighting adjustments used in the full sample weight is then applied to the replicate base weights to create the final replicate weights. As a result, all of the different components of the weighting process are fully reflected in the replicate weights. Use of replicate weights to compute standard errors is straightforward in complex survey software (e.g., SUDAAN, Stata, SAS 9.2, or WesVar).