

Part B:
Supporting Statement
Survey of Working Women
Collection of Information Employing Statistical Methods

B. Collection 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 has been conducted previously, include the actual response rate achieved during the last collection.

The WB is interested in conducting a *Survey of Working Women* (Survey) in order to identify women's current employment issues and challenges and how these issues and challenges relate to job and career decisions, particularly reasons for exiting the workforce. Understanding women's perceptions about the workplace and their participation in the workforce, as well as decision points made at the intersection of work and family obligations, will allow the WB to share valuable information and data with employers, advocates and other stakeholders in an effort to foster greater collaboration and inform policies and practices that meet women's changing needs; and also foster greater public dialogue on these key issues impacting women in today's workforce.

The potential universe for this study consists of adult (18 years of age or older) working women in all 50 states in United States and in the District of Columbia (DC). This will include both full-time and part-time workers, self-employed women and those women who recently were working and opted-out of the workforce for either personal reasons or were laid off/fired. Based on Current Population Survey (CPS) March 2013 estimates, the universe of working women contains approximately 71.8 million women nationwide. The universe counts by census region (Northeast, Midwest, South, and West) based on CPS 2014 estimates are the following: Northeast (13.1 million), Midwest (16.0 million), South (26.5 million) and West (16.2 million). For the purpose of sampling, the geographic region (census region) will be used as a stratification variable (the definition of each region in terms of states is given below in section 2).

Data collection for this Survey will be carried out using a household-based telephone survey. As mentioned above, only working women and those who recently opted-out of the workforce will qualify for this study. This population will not include women without access to a telephone (landline or cell) and those who do not speak English or Spanish well enough to be interviewed.

For this study, a total of 2,700 interviews with workers or recent workers will be completed. This will include three oversamples of specific subgroups of interest of size 500 each. The subgroups of special interest within the group of women are women who recently exited the workforce, and low wage earning women. Additionally, an oversample of 500 working men will be surveyed as a comparison group.

A traditional RDD (Random Digit Dialing) telephone survey will require screening of respondents to reach the target group for this particular survey and hence may not be optimal in terms of sampling efficiency. This will be particularly true for the oversamples where the incidence of the selected subgroup of interest (for example, women who recently off-ramped out of the workforce) may be relatively low. In order to minimize the burden of screening for eligible households for this Survey, 'Recontact sample' obtained from the G1K poll conducted daily by Gallup, the contractor, will be used as the sample source. Further details about the daily G1K poll are given below.

Gallup Daily Tracking Poll (G1K)

Gallup interviews 1,000 adults nationally by telephone for the G1K survey by using a list-assisted RDD telephone data collection methodology. This happens seven days a week and excludes only major holidays. Survey respondents are asked a series of questions associated with well-being, political and economic issues. The survey methods for the G1K study use a dual-frame RDD sampling approach that includes landlines as well as cell phones to reach those in cell-only households, and a random selection method for choosing respondents at random within a sampled household. For the landline part of the telephone sample, a list-assisted telephone sampling method is used. The cell phone sample is selected from the telephone exchanges dedicated to cell phone numbers. Each sampled number is called up to three times to complete an interview.

For the purpose of recruiting for surveys like the proposed Survey of Working Women, respondents completing the G1K are routinely asked if they would be willing to be recontacted. Respondents saying 'yes' to the recontact question in the G1K constitute the 'Recontact Pool.' Phone numbers belonging to working women from the Recontact List will form the sampling frame for this proposed Survey of Working Women. Necessary contact information (name, mailing address) is also collected from these respondents in the G1K survey to be able to reach them later in the survey for which they are recruited.

As in the case of the underlying G1K survey that will be used to generate the 'Recontact Pool,' the telephone samples for the Survey of Working Women will include both landline and cell phones to minimize bias in survey based estimates. Once the sampled household is reached from the 'Recontact' sample, an interview will be attempted with the person who completed the G1K survey and agreed to be recontacted. That person will be interviewed as long as he or she is found eligible for the survey. The eligibility (working, self-employed status or recently opted-out of the workforce) of that respondent will be checked again when they are reached by phone for the Survey of working women.

The methodology, as described above, proposes sampling from the Recontact List as opposed to carrying out an independent RDD telephone data collection methodology. For the traditional RDD approach, bias may be introduced in survey-based estimates due to unit level non-response, i.e., due to non-participation in the RDD telephone survey. The same source of bias is also present in the proposed methodology in the form of non-response bias in the underlying G1K RDD poll. Gallup has conducted non-response bias follow-up studies in the past for studies based on G1K poll. For example, for a study conducted by Gallup for the Census Bureau on perception of federal agencies based on a specific module of questions of the G1K, the non-response bias did not appear to be significant. Necessary non-response adjustment weighting will be carried out for the survey of Working Women to address non-response bias, although it is not likely to be significant.

In addition to the non-response bias discussed above, sampling from the Recontact List may be subject to another potential source of bias because the sampling is proposed from the list of working women who completed the G1K and agreed to be recontacted and not from the entire group of all completed surveys with working women in the G1K. It will, therefore, be necessary to examine the nature and extent of this potential source of bias, if any, and take appropriate steps to minimize bias. For this purpose, important characteristics of the group of all completed surveys with working women in the G1K who did not agree to be recontacted will be compared to those on the Recontact List (those who agreed to be recontacted). If these two groups do not differ significantly in terms of key variables (selected demographic and other relevant variables such as age, race, ethnicity, education, income, geographic region, marital status, category of work), additional steps to remove potential bias due to sampling from the Recontact List may not be necessary. If significant differences are found in terms of some variables, those variables may be chosen to carry out weighting adjustments for removal of the resulting bias due to use of the Recontact List. The potential of non-response due to non-participation of eligible respondents in the survey of Working Women will also be examined and necessary weighting adjustments will be made to minimize that bias.

Following the strategy outlined above, it will be possible to minimize potential bias, if any, due to sampling from the Recontact List. Also, as discussed above, the use of traditional RDD approach would not be cost-effective given the requirements of this study in terms of the number of completed surveys with working women and with other specific subgroups such as low income working women and women who recently exited the workforce. The potential loss in terms of bias, if any, due to use of the Recontact List as the sampling frame for this study is expected to be more than compensated by significant gains in terms of time and cost for sampling of the group of working women and its specific subgroups of relatively low incidence in the general population.

This study has not been conducted previously. The goal will be to maximize the response rate by taking necessary steps as outlined later in this document on "Methods to maximize response rates." The calculation of response rates will be based on AAPOR RR3 definition. The response rate for the Survey based on the Recontact List is expected to be in the 25 to 30 percent range.

2. Describe the procedures for the collection of information including:**-Statistical methodology for stratification and sample selection:**

For the proposed Survey, about 2,700 telephone interviews will be completed nationwide. This will include three selected oversamples (examples described in section B1 above) of 500 respondents each. As described in Section B1, the Recontact List from Gallup's daily G1K survey will be used as the sampling frame for the purpose of sampling. Once the sampled household is reached from the 'Recontact' sample, an interview will be attempted with the person who completed the G1K and agreed to be recontacted for a follow-up survey. A 5 + 5 call design will be employed, i.e., a maximum of five calls will be made to the phone number to reach the specific person we are attempting to contact and up to another five calls will be made to complete the interview with the targeted person. Calls will be made at different times of the day and on different days of the weeks to maximize the potential of contacting the targeted respondent.

The population of working women derived from the Recontact List (the sampling frame) will be geographically stratified into four Census regions (Northeast, Midwest, South, and West). The definition of the four Census regions in terms of states is given below.

Northeast: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, and Pennsylvania.

Midwest: Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.

South: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, and Texas.

West: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, and Washington.

In order to meet the targeted number of interviews for the oversamples, it may also be necessary to sub-stratify each of the Census regions into additional strata based on the variables used to define the subgroups for oversampling. For example, since working women with low wages will be selected for oversampling, sub-stratification based on wage within each Census region will be carried out. Note that necessary information for these stratification variables (income, industry of employment, etc.) will be available for respondents on the Recontact List. Samples will be drawn independently from each stratum following the stratified sample design.

The sample allocation across the sampling strata will depend on the size of strata and the required number of interviews to be generated from each stratum. The goal will be to use proportional allocation to the extent possible across the four census regions (Northeast, Midwest, South and West), i.e. the number of interviews to be completed per region will be roughly in proportion to the size of that region in terms of the estimated number of working women. However, depending on the type of subgroups to be oversampled, the allocation, particularly at the sub-strata level within each region, may have to be based on disproportional allocation. Also, it should be noted that the actual number of completed surveys for each Census region will depend on observed response rates and so they may not exactly match the targeted number of interviews by strata. However, the goal will be to meet those targets to the extent possible by constant monitoring of the response rates and by optimally releasing the sample in a sequential manner throughout the data collection period.

Estimation procedure:

The Survey data will be weighted to generate unbiased survey based estimates. Within each sampling stratum (region), weighting will be carried out to adjust for (i) unequal probability of selection in the sample and (ii) non-response. The sample, as mentioned, will be drawn from the Recontact List derived from the daily G1K. The probability weight component will be

obtained by multiplying the (i) initial weight inherited by these cases from the daily G1K and (ii) the inverse of the selection probability of these cases based on any sub-sampling from the Recontact List. The weighting for the G1K, a dual-frame (landline and cell phone samples) has been completed (details given below later in this section). For non-response adjustments, suitable non-response adjustment cells will be formed based on the sampling stratum and necessary adjustment factors will be derived within each of those adjustment cells. At the next stage, post-stratification weighting will be carried out to project the sample results to known characteristics for the target population. Demographic variables like Age, Race/Ethnicity, and Education will be used as weighting variables. For the oversampled subgroups, additional variables may be included, e.g., wage, presence of children in the household and occupation. Finally, some trimming of weights may be carried out to avoid extreme weights. The target data for post-stratification weighting for demographic variables will be derived from the latest available data from CPS. The rest of this section describes the weighting procedure for the Gallup daily G1K poll.

Gallup Daily Tracking (G1K) Poll Weighting

The weighting for the underlying G1K poll is done following the basic approach described in Kennedy, Courtney (2007): Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys, *Public Opinion Quarterly*, Special Issue 2007, Volume 71 / Number 5: 750-771. In the G1K poll, a dual-frame design is used where dual users (those with access to both landline and cell phones) can be interviewed in either sample. This results in two estimates for the dual users based on the two samples (landline and cell). The two estimates for the dual users are combined and added to the estimates based on landline-only and cell-only population to generate the estimate for the whole population.

For the purpose of sample weighting for the G1K, the composite pre-weight is generated within each weighting class (based on census region and time zones) following Kennedy, Courtney (2007). The weight assigned to the i^{th} respondent in the h^{th} weighting class ($h=1, 2, 3, 4$) will be calculated as follows:

$$W_{(\text{landline},h)} = (N_h/n_{hl})(1/RR_{hl})(n_{cwa}/n_{ll})(\lambda^{IDual}) \quad \text{for landline sample cases} \quad (1)$$

$$W_{(\text{Cell},h)} = (N_{hc}/n_{hc})(1/RR_{hc})(1 - \lambda)^{IDual} \quad \text{for cellular sample cases} \quad (2)$$

where

N_h : size of the landline RDD frame in weighting class h

n_{hl} : sample size from landline frame in weighting class h

RR_{hl} : response rate in weighting class h associated with landline frame

n_{cwa} : number of adults in the sampled household

n_{ll} : number of residential telephone landlines in sampled household

I^{Dual} : indicator variable with value 1 if the respondent is a dual user and value 0 otherwise

N_{hc} : size of the Cell RDD frame in weighting class h

n_{hc} : sample size from Cell frame in weighting class h

RR_{hc} : response rate in weighting class h associated with Cell frame

' λ ' is the "mixing parameter" with a value between 0 and 1. If roughly the same number of dual users is interviewed from both samples (landline and cell) within each census region, then 0.5 will serve as a reasonable approximation to the optimal value for λ . This adjustment of the weights for the dual users based on the value of the mixing parameter ' λ ' is carried out within each census region.

Once the landline and cell samples are combined using the composite weight (equations (1) and (2) above), a post-stratification weighting step is carried out following Kennedy (2007), to simultaneously rake the combined sample to (i) known characteristics of the target population (adults 18 years of age or older) and (ii) an estimated parameter for relative telephone usage (landline-only, cell only, cell mostly, other dual users). The demographic variables used for weighting include Age, Gender, Race, Ethnicity (Hispanic/Non-Hispanic), Education and population density. The target numbers for post-stratification weighting are obtained from the latest available Current Population Survey (CPS) data. The collapsing of categories for post-stratification weighting may become necessary where the sample sizes are relatively small. The target numbers for the relative telephone usage parameter are based on the latest estimates

from NHIS (National Health Interview Survey). After post-stratification weighting, the distribution of the final weights is examined and trimming of extreme weights, if any, is carried out if necessary to minimize the effect of large weights on variance of estimates.

Degree of accuracy needed for the purpose described in the justification:

For the Survey, 2,700 telephone interviews, including 500 interviews for each of the three oversamples will be completed. The survey estimates of unknown population parameters (for example, population proportions) based on a sample size of 2,700 will have a precision (margin of error) of about ± 1.9 percentage points at 95% level of significance. This is under the assumption of no design effect and also under the most conservative assumption that the unknown population proportion is around 50%. The margin of error (MOE) for estimating the unknown population proportion 'P' at the 95% confidence level can be derived based on the following formula:

$$\text{MOE} = 1.96 * \sqrt{P*(1-P)/n} \text{ where "n" is the sample size (i.e. the number of completed surveys).}$$

For the Survey, the total sample size (2,700) will include three oversamples and therefore may be subject to a relatively higher design effect. A design effect of 2, for example, will result in effective sample size of 1,350 and a margin of error around $\pm 2.7\%$ at 95% confidence level. For each of the oversample subgroups with about 500 completed interviews, an estimate for an unknown population proportion will have margin of error around ± 4.4 points ignoring any design effect. With an anticipated design effect of about 1.25, the precision will be around ± 4.9 percentage points. Hence, the accuracy and reliability of the information collected in this survey will be adequate for its intended uses. The sampling error of estimates for this survey will be computed using special software (like SUDAAN) that calculates standard errors of estimates by

taking into account the complexity, if any, in the sample design and the resulting set of unequal sample weights.

The necessary sample size for a two-sample proportion test (one-tailed test) can be derived as follows:

$$n = \left[\left\{ z_{(1-\alpha)} \text{SQRT}(2p^*q^*) + z_{(1-\beta)} \text{SQRT}(p_1q_1 + p_2q_2) \right\} / \{p_2 - p_1\} \right]^2 (3)$$

where

n: sample size (number of completed surveys) required per group to achieve the desired statistical power

$z_{(1-\alpha)}$, $z_{(1-\beta)}$ are the normal abscissas that correspond to the respective probabilities

p_1 , p_2 are the two proportions in the two-sample test

and p^* is the simple average of p_1 and p_2 and $q^* = 1 - p^*$.

For example, the required sample size, ignoring any design effect, will be around 310 per group (top and bottom halves) with $\beta=.2$ (i.e., with 80% power), $\alpha=.05$ (i.e., with 5% level of significance), and $p_1=.55$ and $p_2=0.45$. The sample size requirement is highest when p_1 and p_2 are around 50% and so, to be most conservative, those values (.55 and .45) of p_1 and p_2 were chosen. The proposed sample size will therefore meet the sample size requirements for estimation and testing statistical hypotheses not only at the national level but also for a variety of subgroups that may be of special interest in this study.

-Unusual problems requiring specialized sampling procedures: (e.g., special hard to reach populations, bias toward landline versus cell phone respondents, populations that need to be reached via other methods such as those who do not use telephones for religious reasons, large non-English speaking populations expected to be surveyed but only English questionnaires available, exclusion of elderly using computer response only, etc.)

Note: For surveys with particularly low response rates and a substantial suspicion of non-response bias, it may be necessary to collect an additional sub-sample of completed surveys from non-respondents in order to confirm if non-response bias is present in the sample and make adjustments if appropriate.

Unusual problems requiring specialized sampling procedures are not anticipated at this time. If response rates fall below the expected levels, additional sample will be released to generate the targeted number of surveys. However, all necessary steps to maximize response rates will be taken throughout the data collection period and hence such situations are not anticipated.

-Any use of periodic (less frequent than annual) data collection cycles to reduce burden:

The burden on any sampled respondent will be low and hence use of less frequent data collection cycle is not considered necessary. This survey is also not intended to be recurrent.

- 1. 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.**

Methods to maximize response rates—Necessary steps will be taken to maximize response rates. Gallup will use a comprehensive plan that focuses on (1) a call design that will ensure call attempts are made at different times of the day and different days of the week to maximize contact rates, (2) conducting an extensive interviewer briefing prior to the field period that educates interviewers about the content of the survey as well as how to handle reluctance and refusals, (3) having strong supervision that will ensure that high-quality data are collected throughout the field period, and (4) using troubleshooting teams to attack specific data collection problems that may occur during the field period. A 5 + 5 call design will be employed, i.e., a maximum of five calls will be made on the phone number to reach the specific person we are attempting to contact and up to another five calls will be made to complete the interview with that selected person.

Issues of Non-Response—Survey based estimates will be weighted to minimize any potential bias, including any bias that may be associated with unit level non-response. All estimates will be weighted to reduce bias and it will be possible to calculate the sampling error associated with any subgroup estimate in order to ensure that the accuracy and reliability is adequate for intended uses of any such estimate. Based on experience from conducting similar surveys previously, the extent of missing data at the item level for the telephone based survey is expected to be minimal. We, therefore, do not anticipate using any imputation procedure to handle item-level missing data for this Survey. The goal will be to further minimize item level missing data by using a properly designed survey with necessary instructions.

Non-response bias Study and analysis—Non-response bias analyses will be conducted to examine the non-response pattern and identify potential sources of non-response bias. Non-response bias associated with estimates consists of two factors—the amount of non-response and the difference in the estimate between the groups of respondents and non-respondents.

Bias may, therefore, be caused by significant differences in estimates between respondents and non-respondents further magnified by lower response rates. As described earlier in this section, necessary steps will be taken to maximize response rates and thereby minimize the effect, if any, of lower non-response rates on non-response bias. Also, non-response weighting adjustments will be carried out to minimize potential non-response bias. However, despite all these attempts, non-response bias can still persist in estimates.

The non-response bias analysis will primarily involve comparison of available data from both respondents (those who participate in the Survey) and non-respondents (those who did not respond to the Survey). As noted before, the sampling frame for this study will be the Recontact List. As a result, data on significant number of variables (from the underlying G1K survey) relevant for this comparison will be available for both respondents and non-respondents. Statistical tests of hypotheses will be carried out to detect differences, if any,

between these two groups and thereby examine the potential of non-response bias associated with the survey based estimates.

Another approach to examine non-response pattern and bias will be based on 'early' and 'late' respondents to the Survey. As part of the non-response analysis for the Survey, the respondents will be split into two groups: (i) early or 'easy to reach' and (ii) late or 'difficult to reach' respondents. The call design for the Survey will be 5 + 5 and so a maximum of up to 10 calls may be made to each sampled phone number. The total number of calls required to complete an interview with a respondent will be used to identify these two groups - "early" and "late" respondents. This comparison will be based on the assumption that the latter group may in some ways resemble the population of non-respondents. The goal of the analysis plan will be to assess the nature of non-response pattern in the Survey. Non-response bias analysis may also involve comparison of survey-based estimates of important characteristics of the working women population to external estimates. This process will help identify estimates that may be subject to non-response bias. If non-response is found to be associated with certain variables, then weighting based on those variables will be attempted to minimize non-response bias.

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.

The CATI survey was tested with 9 respondents to ensure correct skip patterns and procedures.

5. 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.

The information collection will be conducted for the WB by Gallup:

The representatives of Gallup who consulted on statistical aspects of design are:

Camille Lloyd
Research Director
Gallup Inc.
901 F Street NW
Washington, DC 20004
202-715-3188
Camille_lloyd@gallup.com

Manas Chattopadhyay
Chief Methodologist
Gallup Inc.
901 F Street NW
Washington, DC 20004
202-715-3030
Manas_chattopadhyay@gallup.com

[Reference:](#)

Kennedy, Courtney (2007): Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys, Public Opinion Quarterly, Special Issue 2007, Volume 71 / Number 5: 750-771.