



Family and Medical Leave in 2012: Methodology Report

Contract #GS10F0086K
TO DOLF109630906

September 6, 2012

Prepared for:

Jonathan Simonetta

U.S. Department of Labor
200 Constitution Ave, NW
Washington, DC 20210

Submitted by:

Abt Associates Inc.

55 Wheeler Street
Cambridge, MA 02138

Kelly Daley
Courtney Kennedy
Marci Schalk
Julie Pacer
Allison Ackermann
Alyssa Pozniak
Jacob Klerman

Table of Contents

Acknowledgements	1
Introduction	2
1. The 2012 Employee Survey	3
1.1 Target Population and Sample Design	3
1.1.1 Screening for the Target Population	3
1.1.2 Within-household Selection Procedure	4
1.1.3 Comparison of the 2000 and 2012 Employee Sample Designs	5
1.2 Questionnaire Development	7
1.2.1 2012 Questionnaire Overview	8
1.2.2 Revision to the Screener	9
1.2.3 Revisions to the Main Survey	11
1.2.4 Cognitive Testing	14
1.2.5 Pilot Survey	16
1.3 Data Collection Procedures	17
1.3.1 Interviewer Training	17
1.3.2 Call-Design, Main Study	17
1.4 Response Rates	17
1.4.1 Response Rate for Screening Interview	18
1.4.2 Response Rate for Extended Interview	22
1.4.3 Overall Survey Response Rate	22
1.5 Analysis of Nonresponse of the 2012 Employee Survey	22
1.5.1 Hypothesized Reasons for the Decline in the Employee Survey Response Rate	23
1.5.2 The Nature of Nonresponse Bias	25
1.5.3 An Empirical Assessment of Nonresponse in the 2012 Employee Survey	25
1.5.4 Summary of Nonresponse Analysis for the Employee Survey	37
1.6 Weighting	38
1.6.1 Probability of Selection of the Telephone Number	38
1.6.2 Screening for the Target Population	38
1.6.3 Multiplicity in the Sampling Frames	38
1.6.4 Within-household Selection Procedure	39
1.6.5 Sampling Frame Integration	40
1.6.6 Extended Interview Nonresponse	40
1.6.7 Post-Stratification to the Current Population Survey (CPS)	40
1.6.8 Handling of Extreme Weight Values	41
1.7 Variance Estimation	41
2. The 2012 Worksite Survey	42
2.1 Target Population and Sample Design	42
2.1.1 Stratification	42
2.1.2 Consistency in the 2000 and 2012 Worksite Surveys	43
2.1.3 Notable Differences Between the 2000 and 2012 Worksite Surveys	43

2.2	Questionnaire Development	45
2.2.1	Updates to Screener	45
2.2.2	Updates to Main Questionnaire	46
2.2.3	Pilot Surveys	48
2.3	Data Collection Procedures	48
2.3.1	Interviewer Training	48
2.3.2	Screening Interview	49
2.3.3	Main Survey Administration	49
2.4	Response Rates	50
2.4.1	Response Rate for the Screener Interview	50
2.4.2	Response Rate for the Extended Interview	52
2.4.3	Overall Response Rate	52
2.5	Analysis of Nonresponse of the 2012 Worksite Survey	52
2.5.1	Comparison of Easier to Reach vs. Harder to Reach Establishments	53
2.5.2	Response Propensity Models for Contact and Cooperation	55
2.5.3	Summary of Nonresponse Analysis for the Worksite Survey	58
2.6	Weighting	58
2.6.1	Base Weight for Probability of Selection	58
2.6.2	Weighting Class Nonresponse Adjustment	59
2.6.3	Post-stratification Ratio Adjustment	59
2.7	Variance Estimation	60
2.8	Producing Employee Level Estimates from the Worksite Survey	60
	References	62

Acknowledgements

The final report is a collaboration of the United States Department of Labor’s Chief Evaluation Office (CEO), Wage and Hour Division (WHD), and Abt Associates. This effort has received the strong support of the CEO—including Jean Grossman, Demetra Nightingale, and Jonathan Simonetta. The survey and this report would not have happened without their vision and day-to-day input. At WHD, Libby Hendrix’s understanding of FMLA and her experiences with the fielding and use of the previous survey were invaluable. Other WHD participating in the effort included Mary Ziegler, Helen Applewhaite, and Karen Livingston.

At Abt Associates, Glen Schneider served as Project Quality Advisor, providing key insights throughout the project and careful review of this final report. Katie Heintz and Cindy Taylor provided valuable management guidance. Katherine Wen, Krista Olsen, Johanna Hudgens, Bethany Bradshaw, and Rachel Banay provided solid research assistance. Nancy McGarry provided solid programming support. Jan Nicholson formatted the document.

The survey was fielded by Abt-SRBI under the leadership of Kelly Daley. Allison Ackermann served as Acting Project Director during the early phase of the project and led the Employee Survey. Julie Pacer led the Worksite Survey. Courtney Kennedy and Ben Phillips led the technical design and weighting effort. Marci Schalk also contributed.

Introduction

This report summarizes the methods used by Abt Associates in conducting the 2012 Family and Medical Leave Act (FMLA) Employee and Worksite Surveys for the Department of Labor (DOL).

The **2012 Employee Survey** is as an overlapping, dual frame landline and cell phone random digit dial (RDD) telephone survey. The target population is U.S. adults age 18 or older who were employed for pay in the past 12 months. The survey features both a screener and an extended interview. Adults who needed or took family/medical leave in the 18 months prior to the interview are oversampled and administered an extended interview roughly twice the length of the extended interview for respondents who did not need or take such leave. In order to identify the extended interview respondent, the screener includes a roster of all the adults in the household, including their relevant employment history and leave-taking behavior. Within-household selection is conducted for both landline and cell phone cases.

The **2012 Worksite Survey** is a mixed-mode telephone and internet survey of U.S. businesses. The Study was conducted to obtain estimates of the use of leave under the FMLA and examine the impact on U.S. private business establishments. The sampling frame was drawn from the Dun and Bradstreet (D&B) Market Identifiers (DMI) file. The final sample excluded self-employed without employees, government and quasi-government units (federal, state, and local governments, public educational institutions, and post offices).

Chapter 1 describes the methods for the 2012 Employee Survey and Chapter 2 describes the methods for the 2012 Worksite Survey.

1. The 2012 Employee Survey

This chapter present the methods employed to design and administer the 2012 Employee Survey, including the following:

- Target Population and Sample Design
- Survey Instrument Development
- Data Collection Procedures
- Response Rate Calculations
- Analysis of Nonresponse
- Weighting
- Variance Estimation

1.1 Target Population and Sample Design

The Employee Survey was designed to sample U.S. adults who had been employed for pay in the private or public sector at any time during the 12 months prior to the interview. This target population did not include those who were self-employed since they are not subject to the FMLA. The survey featured an overlapping dual frame random digit dial (RDD) design with national samples from the landline and cell phone RDD frames. The coverage rate provided by this design is estimated to be approximately 98.1 percent based on the most recent estimates from the National Health Interview Survey (Blumberg and Luke 2012). Interviews were conducted in English and Spanish.

The survey featured 2,852 completed extended interviews, including 2,060 from the landline sample and 792 from the cell phone sample. The data collection was conducted by Abt SRBI from February 1 through June 24, 2012. Both samples were provided by Survey Sampling International, LLC according to Abt SRBI specifications. Numbers for the landline sample were drawn with equal probabilities from active blocks (area code + exchange + two-digit block number) that contained one or more residential directory listings. The cellular sample was drawn through a systematic sampling from 1000-blocks dedicated to cellular service according to the Telcordia database.

1.1.1 Screening for the Target Population

The sample design necessitated screening for members of the target population. The same screening and within-household selection procedures were implemented for the landline and cell phone samples. In the survey screener, interviewers determined whether the household contained at least one person 18 years of age or older who had been employed (excluding self-employed) during the last 12 months. For all persons in the household meeting these criteria, the interviewer attempted to determine if they have taken, were taking (at the time of the interview) or needed without taking family or medical leave during the reference period. The screener involved asking the household informant (age 18 or older) to report the following information for each adult in the household:

- Age
- Gender

- Worked for pay or profit in the last 12 months (yes/no)
- Sector (government, private company, non-profit, self-employed)
- Took leave from work in the last 18 months (yes/no)
- Needed but did not take leave from work in the last 18 months (yes/no)

If the roster identified no adults who had worked for pay or profit in the last 12 months (excluding self-employed), then the household was screened out as ineligible for the survey.

Each eligible adult was classified (in real time) into one of three family or medical leave groups: leave-needer (defined as an employed individual who needed to take time off of work for family or medical leave but did not to take leave; i.e., “unmet need for leave”), leave-taker (an employed individual who took family or medical leave), or an employed-only (an individual who did not need or take family or medical leave).¹ This classification informed the within-household respondent selection.

1.1.2 Within-household Selection Procedure

The purpose of the within-households selection procedure was two-fold. The procedure was used to identify one randomly-selected eligible adult for the extended interview, and it was also used to increase the sample size for the key oversampled subgroups (i.e., the leave-needers and the leave-takers). Each eligible adult in the household was assigned a non-zero probability of selection for the extended interview. The leave-needers and leave-takers were assigned higher probabilities of selection relative to the employed-only in order to increase their sample size. It is important to note that this procedure provides full coverage for the target population, and that the survey weights (described below) adjust for the differential probabilities of selection. In the weighted survey estimates, each of these groups is represented in proportion to its actual size. To accomplish these objectives, the within-household respondent selection was conducted in three stages.

Stage 1 took inventory of which family or medical leave *groups* were present in that household (i.e., Is there at least one leave-needer? at least one leave-taker? at least one employed-only adult?). This information was used to determine from which family and medical leave *group* the extended interview respondent should be selected. For households in which all eligible adults were classified as belonging to the same group (e.g., employed-only), that group was automatically selected. For households where multiple family or medical leave groups were represented, the leave-needer and leave-taker groups were selected at a higher rate than the employed-only group because their population incidence rates are significantly lower. The selection rules and rates applied in Stage 1 were as follows:

- If all eligible adults are of the same family or medical leave group (i.e. a leave-taker, a leave-needer or employed-only), that group is selected. Skip to Stage 2.

¹ Adults who were reported as both needing and taking leave were classified as leave-needers *only for the purpose of the within-households selection*. This temporary classification for logistical purposes has no bearing on the analysis of the survey data.

- If the household contains at least one leave-needer and at least one leave-taker, select the leave-needer group with 90 percent probability and the leave-taker group with 10 percent probability. Skip to Stage 3.
- If the household has at least one leave-needer and at least one employed-only adult, select the leave needer-group with 90 percent probability and the employed-only group with 10 percent probability. Skip to Stage 3.
- If the household has at least one leave-taker and at least one employed-only adult, select the leave taker-group with 90 percent probability and the employed-only group with 10 percent probability. Skip to Stage 3.
- If the household has at least one leave-needer, at least one leave-taker, and at least one employed-only adult, select the leave-needer group with 80 percent probability, the leave-taker group with 10 percent probability, and the employed-only group with 10 percent probability. Skip to Stage 3.

Stage 2 applied only to households in which the employed-only group was selected. This stage involved subsampling these households in order to focus limited survey resources on the employees of most interest; i.e., on leave-takers or leave-needers. Past studies suggest that about 80 percent of U.S. workers belong to the “employed-only” group. Conducting full interviews with all of these cases would have been expected to yield over 6,000 interviews but, consistent with the 1995 and 2000 surveys, only about 1,300 completes with employed-only employees were needed for the analysis. Consequently, households in which the employed-only group was selected were randomly subsampled for the extended interview. If the household was not subsampled in Stage 2, then the interviewer thanked the screener respondent for their cooperation and ended the call. It is important to note that the survey weights (described below) adjust for this subsampling so that in the weighted survey estimates, the employed-only group is represented in proportion to its actual size.

The subsampling rate was determined on a replicate by replicate basis. At the start of the field period this subsampling rate was set at 20 percent, and this rate was used for most of the survey replicates released for the study. We evaluated the results of the early sample replicates to determine whether the group selection rates discussed above were on pace to achieve the target sample sizes. Toward the latter part of the field period, the Stage 2 subsampling rate was increased slightly with the specific value varying across (but not within) the replicates.

Stage 3 involved randomly selecting an eligible adult from the family or medical leave group identified in Stage 1 as the extended interview respondent. For households in which there was exactly one adult in the selected family and medical leave group, that adult was automatically selected. In households where there was more than one adult in the selected family and medical leave group, one adult was randomly selected among those in the group. If the selected adult was not present (e.g., not at home), then the interviewer arranged a time to call back and inquired about the best phone number to reach the selected adult.

1.1.3 Comparison of the 2000 and 2012 Employee Sample Designs

The 2012 Employee Survey was designed with two main methodological objectives: 1) rigorously measure the family and medical leave experiences of a representative national sample of U.S. employees, and 2) maintain as much consistency as possible with the 2000 Employee Survey, without

threatening objective 1. The second objective is important because longitudinal comparisons are a critical component of the 2012 FMLA study report. In this section we highlight the key consistencies and differences between the 2000 and 2012 sample designs. Revisions to the survey questionnaire are described in section 1.2.

Consistencies

Many key design elements are consistent in the 2000 and 2012 Employee Surveys. Critically, the target population has essentially remained the same, as has the general mode of survey administration. Both the 2000 and 2012 Employee Surveys were designed to make inference to employed adults in the U.S. Also, interviewing for both the 2000 and 2012 surveys was conducted using computer assisted telephone interviewing (CATI), and the samples were drawn using probability-based, random digit dial methodology. The addition of a cell phone sample in 2012 is discussed below. This consistency in target population and general approach to data collection serves to reduce the likelihood that nonresponse or measurement related factors confound longitudinal comparisons.

Another key consistency is the emphasis on the key subgroups of leave-needers and leave-takers. In 2000 and 2012, the screening instrument was designed to identify leave-needers and leave-takers in the household and select them for the extended interview at a higher rate than employed-only adults. In summary, both the 2000 and 2012 Employee Surveys were national, probability-based, high coverage, random digit dial surveys that screened for the target population and oversampled the key subgroups of leave-needers and leave-takers.

Notable Differences

While most broad design elements of the Employee sample were held consistent, four aspects of the survey were modified in 2012 in order to ensure high data quality. First, and perhaps the most notable change, is the fact that the 2012 Employee Survey featured a cell phone RDD sample in addition to a landline RDD sample. Currently, about one third of adults in the U.S. have a cell phone in their household but not a landline (Blumberg and Luke 2012). While a landline-only RDD design provided high population coverage in 1995 and 2000, the coverage rate in 2012 would have been less than 70 percent. In order to provide near full population coverage (over 98 percent), the 2012 survey featured an overlapping dual frame landline and cell phone design.

Second, the sample selection procedure was also bolstered in 2012. In the 2000 Employee Survey, some of the extended interview respondents lived in the same household. This feature yielded a cluster design, which generally reduces statistical power and effective sample size due to intra-class correlation (the fact that people living together often have correlated values on survey measures). In the 2012 Employee Survey, a cluster design was not used. Only one eligible adult was selected from each sampled household. This design change is expected to yield greater statistical power relative to the 2000 survey.

Third, another difference between the 2000 and 2012 Employee Surveys pertains to the weighting adjustments. The weighting protocol developed for the 2012 Employee Survey is more comprehensive than that used for the 2000 Employee Survey, and this could have implications for the survey estimates. As discussed above, the 2012 weighting includes an adjustment for multiplicity within the sampling frame (e.g., multiple landlines or multiple cell phones in the household). The 2000 survey did not make such an adjustment in the interest of consistency with the 1995 survey,

though it was recommended in the methodology report. It is also important to note that the 2000 Employee Survey weights were post-stratified only to population control totals for gender. The 2012 Employee Survey weights, by contrast, are post-stratified to population control totals for age, gender, education, race/ethnicity, region, and phone service. It is not clear why post-stratification to age, education, and race/ethnicity in particular were not included in the 2000 Employee Survey weights as these variables are correlates of response propensity and some employment and health-related characteristics.

Fourth, there were also changes in the reference period for the screener. We discuss these changes in the next section.

On each of the issues discussed above, the 2012 Employee Survey is expected to be at least as rigorous as the past surveys. One metric on which the 2012 survey does not perform as well is response rate. An extensive nonresponse analysis for the 2012 Employee Survey is presented below in section 1.5.

1.2 Questionnaire Development

As much as possible, the 2012 survey followed the model implemented in the 2000 surveys. That strategy preserves comparability, allowing analyses of changes over time. Despite efforts to maximize comparability, the final 2012 surveys differ substantially from the 2000 instrument for several reasons. Twelve years have elapsed since the last FMLA survey was conducted. In that time, labor markets have continued to evolve, including an increase in women in the workforce, an aging workforce and population, declines in unionization, and changes in the distribution of employment across sectors. Each of these surveys was also conducted in a very different economic environment.² When the first surveys were fielded in calendar year 1995, the unemployment rate was 5.6 percent; when the second surveys were fielded in calendar year 2000, the unemployment rate was 4.0 percent. The 2012 surveys were fielded when the unemployment rate was over 8 percent.

The questionnaire development process for the 2012 Employee proceeded in four main phases, described below:

- 1) Revisions to the Screener
- 2) Revision to the Main Survey
- 3) Cognitive testing
- 4) Pilot Surveys

To facilitate comparisons and identification of trends, we began with the 2000 survey as a base. To ensure that new questions adequately capture the range of issues and experiences with regulatory changes since 2000 and the possibilities for future efforts, we gathered information from various sources, including a list of recommended sources provided by the Department of Labor Wage and Hour Division, the public comments in response to the Department of Labor's 2006 Request for Information, peer-reviewed published literature as well as "gray" literature (e.g., relevant newspaper

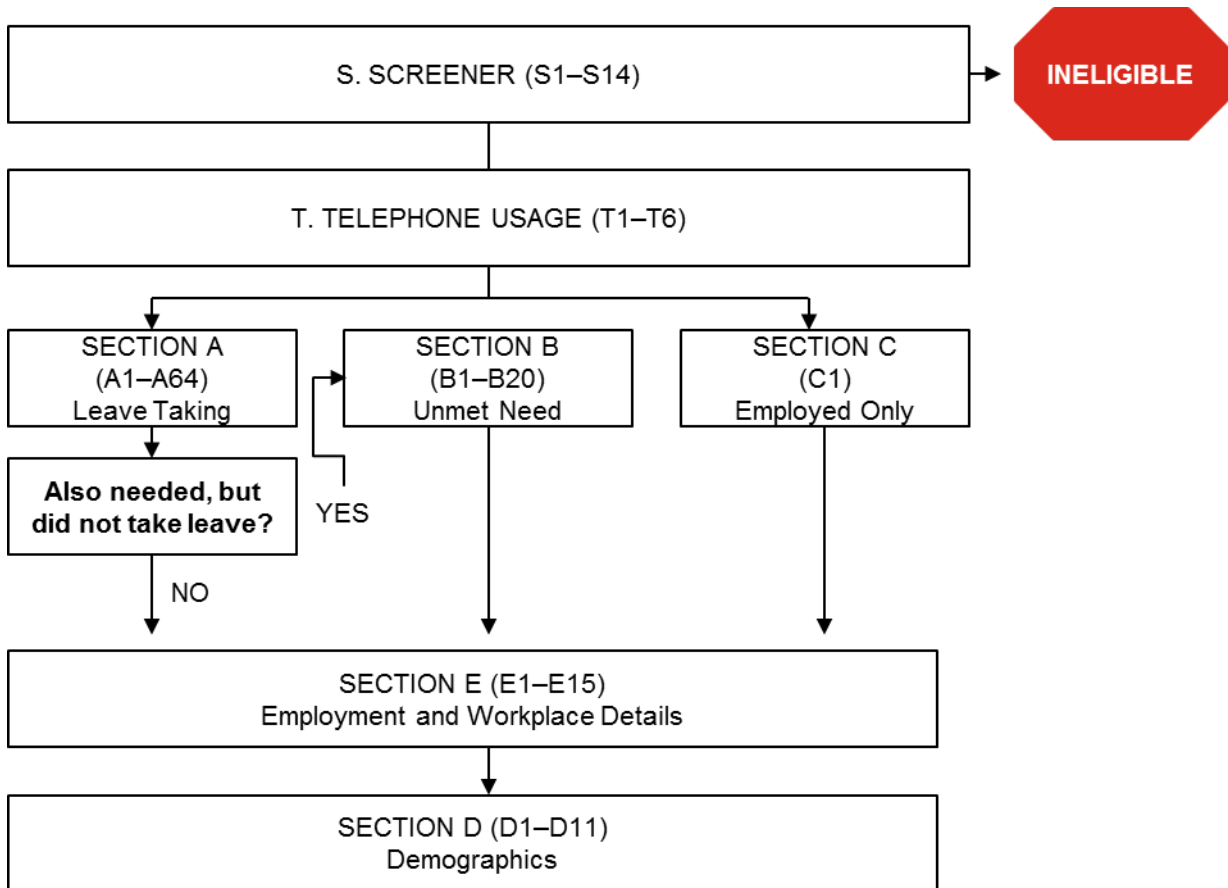
² These raw monthly unemployment rates are drawn from data.bls.gov/timeseries/LNS14000000. The 1995 and 2000 rates are simple averages of the twelve monthly values. The 2012 value is for January through May (the most recent available data as of this writing).

articles, policy papers). We conducted in-person interviews with staff at the Department of Labor. Additionally we held two listening group sessions with representatives from nine employee and four employer stakeholder organizations to elicit their feedback. The questionnaire drafts were reviewed by a Technical Working Group.

1.2.1 2012 Questionnaire Overview

The 2012 Questionnaire included five major sections. The first section is the screener, shown as Section S in the figure below. Corresponding questions in the survey instrument begin with the letter S. During the screener, as described above, the target respondent was selected according to their leave status as a “leave taker”, a “leave needer” or “employed only”. Following the screener, eligible respondents were asked a series of telephone usage questions for the purposes of integrating and weighting the cell and landline sample frames. These are shown as Section T in the figure below. Next, respondents were asked questions related to leave and were taken to either Section A (leave-takers) Section B (leave needers) or Section C (employed only). At the beginning of each section, the respondent’s leave designation was confirmed. Leave takers were also asked whether they *also* had unmet need during the reference period. If so, they were then asked questions in Section B. Employed only respondents proceeded directly to section C after the screener and telephone usage questions. After completing their respective leave sections, all respondents were asked questions related to their employment situation and benefits (Section E), followed by standard demographic questions in Section D.

Figure 1.1.1 Structure of the 2012 Employee Survey.



A discussion of key questionnaire revisions and additions follows below. Appendix C includes a question-by-question crosswalk between the 2000 and 2012 Employee Surveys. Appendix A includes a copy of the 2012 Employee Survey.

1.2.2 Revision to the Screener

This subsection discusses revisions to the screener.

Period of Employment

The 2000 survey screener question about employment asked about employment in an 18-21 month period (between January 1, 1999 and the interview date). In the 2012 Survey, we changed the employment time period to include the last **12** months in order to better align the sample and control totals used for weighting. Survey response rates have fallen dramatically in the twelve years since the last survey was conducted. One way to adapt to this reality is to construct sampling weights from more detailed population information. An employment question based on any employment in the last 12 months allowed us to construct control totals from the March 2011 Current Population Survey (CPS):

In the last 12 months, [have you / has A2-X] worked for the government, a private company, a non-profit organization, or [have you / has A2-X] been self-employed? [IF NECESSARY: Please think about your most recent/main job.]

This change allows us to use control totals from the Current Population Survey on those characteristics of those employed in the last year. Such control totals allow considerably better weighting for differential nonresponse. Comparable control totals for those employed in the last 18 months do not appear to be available. This means, that people who were employed in the last 18 months, but not in the last 12 months (i.e., they only worked some time in months 13 to 18) would have been included in the 2000 survey, but would not have been included in the 2012 survey. This change induces a small difference in the targets populations. We do not believe that this change has major impacts on any outcome.

Definition of Qualifying Leave

Among those satisfying the screener for any employment in the last 12 months, we assigned leave groups based on leave in the last 18 months. This reference period followed the 2000 survey, but we made other changes to the definition of qualifying leave. The wording from the 2000 Survey Qualifying Leave question was:

*Since **January 1, 1999**, {have you/has this person} taken leave from work*

- *to care for a newborn, newly adopted, or new foster child;*
- *for reasons related to your or a family member's pregnancy; or*
- *for {your/their} own serious health condition or the serious health condition of {your/their} child, spouse, or parent? A serious health condition is one that lasted more than 3 days or required an overnight hospital stay.*

The 2012 Survey modified this question to provide more precision in the definition of leave, to improve recall and update the definition of qualifying leave conditions under the FMLA since the

2000 survey. The 2012 question was modified, as shown here. Specific modifications are bolded and/or described below.

*In the **LAST 18 MONTHS**, that is, since [INSERT 18 MONTH PERIOD] [have you/has [FILL A1-X FROM QS6]] taken leave from work for ANY of the following reasons:*

- *to care for a newborn, newly adopted or new foster child; (IF NECESSARY: This includes both maternity AND paternity leave)*
- *for [your own/[FILL A1-X FROM QS6]'s] serious health condition or to care for someone else's serious health condition;*
- *for pregnancy-related reasons (IF NECESSARY: [IF QS8 >1: your own/[FILL A1-X FROM QS6]'s or] a family member's); or*
- *to care for a military service member, or for reasons related to the deployment of a military service member?*

A serious health condition, for purposes of this survey, means a condition that lasted more than 3 days and required treatment by a health care provider, a condition that required an overnight hospital stay, or a long-lasting condition for which one must see a health care provider at least twice a year for treatment. It may also include a condition that makes one permanently unable to work or perform other daily functions, or that requires treatments to keep from becoming incapacitated.

Reference Period for Leave in Screener Interview

Both the 2000 and 2012 surveys asked about all leave taken and leave needed over a reference period to define leave groups for stratification. In the 2000 survey, the start of the reference date was fixed at January 1, 1999. The field period began July 15, 2000 and continued for approximately 10 weeks. This would require the respondent to answer the question in reference to possible range of 18 to 21 months. The reference period for leave taking questions in the screener was changed to a static 18-month period. This change was made so that all respondents were asked to report over the same total period of time, to ease burden and to improve respondent recall.

Qualifying Care Recipient

The 2000 version asked about leave needed or taken to care for the serious health condition of a child, spouse or parent only. In 2012, this was expanded to cover “someone else’s serious health condition”. The language was changed to reflect the 2010 DOL administrative clarification of the definition of “son or daughter” under Section 101(12) of the Family and Medical Leave Act (FMLA) as it applies to standing “in loco parentis” to a child. In loco parentis applies to both an employee who is standing in loco parentis to a child (and therefore needing to take leave to care for this child) *and* an employee who takes leave for a person who stood/stands in loco parentis to the employee: “An eligible employee is entitled to take FMLA leave to care for a person who stood in loco parentis to the employee when the employee was a child. The fact that the employee also has a biological, adoptive, step, or foster parent, does not preclude a determination that another individual stood in loco parentis to the employee when the employee was a child.” (29 U.S.C. § 2611(12)). See also 29 C.F.R. §§ 825.122(c), 825.800. The 2012 screener description asks the respondent’s own or someone else’s serious medical condition, without specifying the relationship. This change was made to ensure that respondents were correctly screened and selected under the updated qualifying conditions under the FMLA.

Military Family Leave Entitlements

The National Defense Authorization Act for Fiscal Year 2008 (2008 NDAA), Public Law 110-181, amended the FMLA to allow eligible employees to take up to 12 workweeks of job-protected leave in the applicable 12-month period for any “qualifying exigency” arising out of the active duty or call to active duty status of a spouse, son, daughter, or parent. The 2008 NDAA also amended the FMLA to allow eligible employees to take up to 26 workweeks of job-protected leave in a “single 12-month period” to care for a covered service member with a serious injury or illness. These two types of FMLA leave are known as the military family leave entitlements. This change was made to ensure that respondents were correctly screened and selected under the updated qualifying conditions under the FMLA.

Definition of Serious Health Condition

The 2000 survey definition for serious health conditions did not specify chronic conditions. Given the growing interest in intermittent leave, the definition was revised to include: “a long lasting condition for which one must see a health care provider at least twice a year for treatment.”

1.2.3 Revisions to the Main Survey

In this section we describe the key substance changes to the main survey questionnaire from the 2000 and 2012 FMLA Employee Surveys.

Addition of Dual Taker/Needer FMLA Designation

In accordance with the 2000 survey, leave designation for the target respondent was confirmed during the extended interview. If the target respondent reported a different leave designation, this was the designation used for the extended interview. However, in contrast to the 2000 survey, in the 2012 survey, each confirmed leave takers is asked if they *also* needed to take leave but did not during the reference period. If they answered “yes”, they were also asked all of the questions in the Section B, the leave needing section of the questionnaire. For purposes of analysis, these respondents were given a second FMLA designation “dual taker/needer” in the data set.

Change in Qualifying Care Recipient

The Family and Medical Leave Act (FMLA) entitles an eligible employee to take up to 12 workweeks of job-protected unpaid leave for the birth or placement of a son or daughter, **to bond with a newborn or newly placed son or daughter**, or to care for a son or daughter with a serious health condition. See 29 USC 2612(a)(1). The response categories to the 2000 Survey questions A3 and B2 have been updated to distinguish between “to care for a newly adopted child” and bonding with a new child ((2012 Survey questions A5 and B6A). These categories include:

- *TO BOND WITH NEWBORN*
- *TO BOND WITH NEWLY ADOPTED CHILD*
- *TO BOND WITH NEWLY PLACED FOSTER CHILD*

Additionally, “DOMESTIC PARTNER’S HEALTH CONDITION” was added to the response options.

Leave for Military-Related Reasons

Changes for military leave entitlements (see discussion above) are reflected in the main questionnaire by way of an additional response category in questions A5, A5a, B6, B6a. These questions collect

information about the reasons for taking leave or needing to take leave. Additionally, for those who report taking/needing leave to care for a military member, follow-up questions were added to ask about the military member's relationship to the respondent and the total amount of time needed for that particular leave reason. (A9a, A19a, B10a, B10b)

Focal Leave

In the early years of the FMLA, concern focused on the impact of long leaves. The 1995 and 2000 surveys therefore asked detailed questions about the longest (and second longest leaves). This approach has the unfortunate side effect of collecting no information about shorter leaves. Given increased interest in intermittent (i.e., possibly short) leaves, the 2012 survey changes the questions to gather information about both the **longest** and the **most recent leave**. Asking about the "most recent leave" yields an (approximately random, but length biased) sample of all leave occasions.

When the focus is on long leaves, it seems reasonable to ask about a "focal leave". Given interest in intermittent leave, it seems more natural to ask about "all leave taken for this medical condition". The 2012 Survey revised questions to focus on leave taken by medical condition. This change affects several questions, discussed below.

Nature of the Serious Health Condition

In the 2000 survey, respondents were asked an open-ended question about the health condition for which they needed or took leave. "What health condition did you [your spouse, or other care recipient] have?" Open-ended questions allow for incommensurate responses and would involve extensive coding in order to have analytic value. We revised the question to gather information about the nature of the health condition into the three categories that were most appropriate and useful for analysis related to leave-taking and leave-needing. These include:

- A one-time health matter, such as appendicitis or injury;
- The treatment of an injury or illness that now requires routine scheduled care, such as chemotherapy or physical therapy; or
- An ongoing health condition that affects one's ability to work from time to time, such as diabetes, migraines, depression, or Multiple Sclerosis?

Intermittent Leave

The 2000 survey asked about intermittent leave with Question A5b and A5c, shown below:

2000 Survey Version:

A5b/A5c Sometimes people alternate between work and leave. That is, they repeatedly take leave for a few hours or days at a time because of ongoing family or medical reasons. Have you taken this kind of leave since January 1, 1999?

Was this kind of leave less than half, about half, or more than half of all the time you spent on family or medical leave since January 1, 1999?

In the 2012 Employee Survey, these questions are modified to reflect the changes described above about focal leave:

2012 Survey Version:

A14. *Did you take this time off continuously -- that is, all in a row without returning to work -- or did you take leave on separate occasions?*

If the answer to A4 is SEPARATE occasions, then:

A15. *How many separate blocks of time did you take off from work during this leave? [IF NECESSARY: Please think about special events, holidays, or seasons to help you remember.]*

A16. *In what month and year did the **last** block of time for this leave begin? [IF NECESSARY: Please think about special events, holidays, or seasons to help you remember.]*

A17. *And in what month and year did this leave end? [IF NECESSARY: Please think about special events, holidays, or seasons to help you remember.]*

Ascertaining Respondents' Knowledge of FMLA Eligibility

To ascertain the level of employee awareness about the FMLA, all respondents were asked a new question (E4a). The question lists possible reasons for needing leave and asks respondents whether the reasons are covered by the FMLA. To remain within the time parameters of the survey, rather than asking each respondent all 11 items, 4 items were subsampled per respondent and asked in a random order (one of which came from the “no” items and one of which came from the military-related items).

E4a. *To the best of your knowledge, are employees who are covered by the federal FMLA law entitled to take leave for the following reasons?*

- a. *For the care of a newborn?*
- b. *For an employee's own serious health condition?*
- c. *For the care of a child with a serious health condition?*
- d. *For the care of a spouse with a serious health condition?*
- e. *For the care of a parent with a serious health condition?*
- f. *For the care of a grandparent with a serious health condition?*
- g. *For the care of a grandchild with a serious health condition?*
- h. *For the care of a sibling with a serious health condition?*
- i. *For the care of an adopted child or foster child?*
- j. *For the care of a military service member?*
- k. *For reasons related to the deployment of a military service member?*

Medical Certification

An employer may require that the need for leave for a serious health condition of the employee or the employee's immediate family member be supported by a certification issued by a health care provider. Questions were added to the 2012 Employee Survey to cover issues related to certification and re-certification of medical conditions for leave takers.

Other Changes Relative to the 2000 Survey

In addition to the revisions listed above, some questions were deleted from the 2000 Employee Survey. These were deemed as no longer relevant:

- A14a. *Which method (to cover work while you were away on your leave) was used most often?*
- C2. *“Over the next 5 years, how likely do you think it is that you will need to take a leave from work for your own serious health condition, the serious health condition of your child, spouse, or parent, or for the arrival of a newborn, newly adopted, or new foster child.”*
- C11. *Please tell me whether you agree or disagree with the following statements:*
- a) *Every employee should be able to have up to 12 weeks of unpaid leave in a year from work for family and medical problems*
 - b) *Having to provide employees with up to 12 weeks of unpaid leave in a year for family and medical problems is an unfair burden to employees’ co-workers*
- C11e. *“Would you say that your co-workers taking leave had a positive impact on you, a negative impact on you, or neither?”*
- C12a. *“Of (the employment benefits) offered, which two are the most important to you?”*

See Appendix C for the complete list of changes between the 2000 and 2012 Employee Survey.

1.2.4 Cognitive Testing

The objective of cognitive testing is to identify problems respondents are likely to have with any part of the response process and to help eliminate sources of response error. We used cognitive interviewing to aid in the development of new survey items and test the appropriateness of published survey questions for use in this context. Specifically, we sought to ascertain if the individual question wording and response categories adequately captured the range of respondent experiences with taking leave and needing to take leave, particularly on an intermittent basis. Secondly, we attempted to identify recall issues for questions pertaining to multiple leave occasions or multiple conditions over 12 and 18-month periods. Third, we tested respondents’ understanding of technical terminology related to their employers’ leave taking and benefits policies. Finally, we tested the overall flow of the newly designed questionnaire under a variety of distinct respondent conditions (leave takers, leave needers, intermittent and long-term leave, for their own serious health condition and to care for others.)

We conducted cognitive tests on the Employee Survey with nine volunteer respondents in Chicago (the location of the core survey design staff for this project). These purposively selected respondents included employees who took or needed to take leave from work for a variety of family and medical reasons in order to test the applicability of questions on different types of leave takers or leave needers. These respondents came from a diversity of backgrounds and education levels in order to test applicability of the questions for different types of employees (salaried versus hourly, for example) and to capture the range of possible comprehension issues.

Generally speaking, administration of the questionnaire went well. Respondents understood the basic requirements of the questions being asked and were generally cooperative and helpful. Respondents were forthcoming with their answers and no one refused outright to answer any questions. Specific question testing that resulted in wording changes are described below.

Focal Leave

As described above, a significant change from the 2000 survey involved around the focal leave for the questionnaire. The question designed to capture this information required some development. Although intended to capture information about time off from work related to *one* (medical or family leave) reason, testing revealed that respondents did not understand the question as intended.

Testing revealed the following challenges:

1. The word “condition” is inappropriate for someone who took leave to care for another person, or to care for a newborn/newly adopted child.
2. Despite using the phrasing “separate occasions”, respondents did not distinguish between time taken off on separate occasions for the same condition (intermittent leave) and time taken off for separate reasons or health conditions.

The question was revised to emphasize the *reasons* for taking time off of work so as to avoid confusion with the word “leave.” Cognitive testing respondents reported that they understood each time they took off of work as a separate leave, even if it was for the same underlying serious medical condition. Respondents better understood the concept of leave when we specified that the leave refers ALL time off taken for the same conditions. The final question was revised to then ask separately about the total number of reasons and conditions, and subsequent questions ask for information about the amount of time taken for each single reason.

We are interested in the number of times you took leave from work for A SINGLE reason or condition (yours, or that of the person you cared for), and this is regardless of whether you took time off all at once or in separate blocks of time. So, for how many TOTAL reasons or conditions did you take leave from work since [INSERT 18 MONTH PERIOD]?

Employer Designation of Leave

The goal of question A15 was to measure whether or not respondents knew if their leave was designated as FMLA leave by their employer. However, prior to testing, survey reviewers at DOL were concerned that using the terms “family and medical” leave in the question stem would prime the respondent to answer “yes” when it had not actually been filed by the employer under FMLA.

A15. Was the leave you just told me about designated by your employer as family and medical leave?

- 1 YES
- 2 NO
- 8 DK (VOL)
- 9 REF (VOL)

We tested this question as an open-end: “how did your employer designate your leave?” with follow-up probes. Upon the first posing of this question, not a single respondent used words that designated the leave as FMLA eligible leave. Several respondents asked for clarification. Interviewers then asked respondents to please re-phrase the question using his/her own words. Responses included: “You mean, did they give it to me?” “Did they let me take it?”

After uncovering these issues, we included a follow up probe, replacing the word “designate” with “classified”: “How did your employer *classify your leave?*” and this probe resulted in a more appropriate response: “It was a leave of absence.” The question was revised to clarify this language, and it was asked as an open ended question coded by interviewers:

How did your employer designate or categorize the leave you just told me about? That is, what type of leave did your employer assign to your time off? [DO NOT READ LIST]

- 1 VACATION LEAVE
- 2 SICK LEAVE
- 3 FAMILY AND MEDICAL LEAVE
- 4 OTHER (SPECIFY): _____
- 8 DK (VOL)
- 9 REF (VOL)

1.2.5 Pilot Survey

Given the complexity of the questionnaire and the number of changes since 2000, we divided the pilot into two rounds. Round 1 was designed to identify interviewer problems with the screening and recruitment processes, uncover respondent problems in understanding or answering new and revised questions, and identify unanticipated responses and inadequacies of the interviewer instructions (for handling respondent questions, probes, or out of range responses). This pilot took place in February, 2011, prior to OMB submission. As a result of the Round 1 pilot, several interviewer aid materials were revised and expanded. An event history calendar was developed to assist interviewers with respondents who took or needed leave for multiple conditions during the reference period.

We conducted an additional 15 interviews during the second pilot, which took place in January, 2012. This pilot focused on reacquainting the interviewers with the instrument and ensuring proper CATI administration. During the course of this pretest, survey administration length for leave takers was running far beyond the originally estimated time. As a result, in consultation with DOL, the following 2000 Employee Survey questions were deleted from the 2012 Employee Survey after the pretest:

- A7. *I'm going to read you some reasons why some people might be worried about taking family or medical leave. For each of these, please tell me if you were worried. Were you worried about taking family or medical leave.*
- A12. *Would you say using family and medical leave had a positive effect or no effect at all on... [RANDOMIZE]*
 - a. *Your ability to care for family members?*
 - b. *Your ability to select a satisfactory childcare provider?*
 - c. *Your ability to select a satisfactory caretaker for a sick family member?*
 - d. *Your or your family member's physical health?*
 - e. *Your or your family member's emotional well-being?*
- A13. *Which effects did your family and medical leave have on your or your family member's physical health? Would you say...*
 - a. *A quicker recovery time*
 - b. *It was easier to comply with doctor's instructions*
 - c. *It delayed or avoided need to enter nursing home or other long-term care facility, or*

d. Was there another effect (SPECIFY)?

A19. *Now I'm going to ask you some questions about your feelings regarding your leave. How easy or difficult was it to get your employer to let you take time off? Would you say it was...*

1.3 Data Collection Procedures

Interviewing for the Employee Survey was conducted using computer assisted telephone interviewing (CATI). Interviewing was conducted in English and Spanish.

1.3.1 Interviewer Training

Intensive trainings were conducted with interviewers to prepare them for administration of the survey. The first training reviewed general interviewing principles and unique study procedures and requirements. It also allowed interviewers access to the CATI equipment, to gain familiarity with the questionnaire and to perform practice interviews. At the start of the training, the project directors explained the purpose and goals of the study. In telephone surveys, the most critical issue is usually to ensure that the interviewer understands the questionnaire fully, and knows how to ask the questions properly and record the responses accurately. Abt SRBI project staff reviewed important considerations in the questionnaire, including probing, expected respondent questions, and ambiguity. We reviewed the questionnaire, the question-by-question specifications, and questions and problems that interviewers had concerning the questionnaire. Mock interviews were conducted across all of the interview types (leave-taker, leave needer, dual, and employed only).

1.3.2 Call-Design, Main Study

Interviewers placed phone calls from 5:00 pm to 9:00 pm on weekdays, from 10:00 am to 6:00pm on Saturdays, and from noon to 9:00 pm on Sundays. Daytime calling during the week was used periodically to reach non-contacts. In addition, special arrangements were made to accommodate other times of the day based on a respondent's request (i.e., outside the regular calling hours listed above). To increase the probability of completing an interview, we established a differential call rule requiring that call attempts be initiated at different times of the day and days of the week.

The maximum number of call attempts was originally set at 15. However, this number was eventually increased to 20 for noncontacts and callbacks. All calls to cell phones were manually dialed. Landline telephone numbers were dialed using an autodialer. Telephone numbers were dialed until contact was established with a respondent associated with the number, or until the telephone number was determined to be incorrect or out of service. For participants completing the questionnaire on a cell phone, \$10 incentives were issued to compensate for minutes. See Appendix D for more information on incentives for landline cases. The average length of completed interviews was 18.2 minutes.

1.4 Response Rates

We computed the response rate for the Employee Survey in three steps. The first step calculated a response rate for the screening interview. The second step calculated a response rate for the extended interview. The third step combined the two response rates to produce the overall survey response rate. Each of these steps is described in detail below.

Unless otherwise noted, the response rates in this report are computed according to current AAPOR Standard Definitions of case codes and outcome rates (AAPOR 2011). As discussed below, the formulas used to produce the response rates for the 1995 and 2000 surveys are out-dated, especially given the fact that the 2012 Employee Survey included a cell phone sample in addition to a landline sample.

1.4.1 Response Rate for Screening Interview

The final outcomes of call attempts for the screening interview and the extended interview are presented in Exhibit 1.4.1. These outcomes are presented separately for the landline and cell phone samples, as well as combined (unweighted).

Exhibit 1.4.1 Dispositions for the 2012 Employee Survey by Sample

Disposition	AAPOR Code	Total Sample		Landline Sample		Cell Sample	
		Screener	Extended	Screener	Extended	Screener	Extended
Interview (Category 1)							
Complete	1.100	19,567	2,852	15,113	2,060	4,454	792
Eligible, non-interview (Category 2)							
Refusal	2.110	31,349	1,614	19,578	1,109	11,771	505
Respondent never available	2.210	320		219		101	
Telephone answering device	2.220	21,497		11,600		9,897	
Physically or mentally unable	2.320	2,203		1,781		422	
Language problem	2.330	1,003		654		349	
Unknown eligibility, non-interview (Category 3)							
Always busy	3.120	1,794		1,364		430	
No answer	3.130	20,342		17,420		2,922	
Call blocking	3.150	98		20		78	
Not eligible (Category 4)							
Fax/data line	4.200	6,505		6,433		72	
Disconnected number	4.320	138,156		123,916		14,240	
Temporarily out of service	4.330	3,813		2,649		1,164	
Business, gov't office, other org.	4.510	11,976		10,393		1,583	
No eligible respondent	4.700	1,840		123		1,717	
Total		260,463	4,466	211,263	3,169	49,200	1,297

Based on the AAPOR RR(3) formula, the overall response rate for the screener is 23.6 percent. This rate is based on the combined sample. Looking at the samples separately, the screener AAPOR RR(3) is 28.1 percent for the landline sample and 15.3 percent for the cell sample. These rates are presented in the first row of Exhibit 1.4.2 below.

The 2000 Employee Survey methodology report provides two different screener response rate formulas: a “higher” response rate formula that originated in the 1995 survey report and a “lower” response rate formula. The two formulas reflect different assumptions about the eligibility of telephone numbers in which no person ever answered the phone. The “lower” response rate formula assumes that 27 percent of the “Ring No Answer” numbers represent eligible (working and residential) numbers, and that 60 percent of the “Telephone Answering Device” numbers are eligible. According to the 2000 report, the rate applied to the “Ring No Answer” cases is based on research that tracks telephone numbers through the telephone company. The report does not provide an explanation for why only 60 percent of the “Telephone Answering Device” numbers are assumed to be working and residential. The “higher” response rate formula assumes that none of the “Ring No Answer” numbers are eligible and all of the “Telephone Answering Device” numbers are eligible.

Applying the assumptions incorporated into the response rate formulas used in the 1995 and 2000 survey reports lead to somewhat higher response rates relative to the standardized AAPOR RR(3) formula. Under AAPOR RR(3), 100 percent of the Telephone Answering Device numbers are assumed to be eligible (not 60%), and 32 percent of the Ring No Answer cases are assumed to be residential (not 27% or 0%). Consequently, more of these numbers are counted in the denominator of the response rate, leading to a lower outcome. When the AAPOR RR(3) formula is applied to the 2000 Employee Survey dispositions, the screener response rate is three percentage points lower than the rate that is presented in the report. To be sure, this difference in the formulas does not account for the majority of the difference between the 2000 and 2012 Employee Survey response rates. That said, it is important to note that standards for computing response rates have changed over time and there are consequences for the response rates realized.

Exhibit 1.4.2 Screener, Extended Interview, and Overall Response Rates, by Formula

Response Rate Formula	Total Sample			Landline Sample			Cell Sample		
	Screener	Extended	Overall	Screener	Extended	Overall	Screener	Extended	Overall
AAPOR Response Rate 3* Formula: $I / ((I+P) + (R+NC+O) + e(UH+UO))$	23.6%	63.9%	15.1%	28.1%	65.0%	18.3%	15.3%	61.1%	9.4%
2000 FMLA "Lower" RR Formula Formula: $C / (C+R+.27NA+.6M+LP+MC+ONR)$	26.7%	63.9%	17.0%	30.6%	65.0%	19.9%	18.6%	61.1%	11.4%
2000 FMLA "Higher" RR Formula Formula: $C / (C+R+M+LP+MC+ONR)$	25.8%	63.9%	16.5%	30.9%	65.0%	20.1%	16.5%	61.1%	10.1%

* The "e" coefficient in AAPOR RR(3) was computed as $(I+R+NC+O) / ((I+R+NC+O)+NE)$

- I = completed interview
- P = partial interview
- R = refusal
- NC = non-contact
- O = other nonresponse
- NE = not eligible
- UH = unknown if household
- MC = maximum calls
- NA = no answer
- M = answering machine
- LP = language problem
- ONR = other nonresponse

1.4.2 Response Rate for Extended Interview

The extended interview response rates (Exhibit 1.4.2) represent the proportion of interviews that were completed among those eligible and selected for the extended interview. The AAPOR RR(3) for the extended interview was 63.9 percent.

1.4.3 Overall Survey Response Rate

The overall survey response rate is computed as the product of the screener and extended interview rates. The AAPOR RR(3) overall survey response rate is 15.1 percent. This figure is as high as 17.0 percent when using the alternative response rates presented in the 2000 Employee Survey report. All of the Employee Survey response rates reported here are based on unweighted data due to the fact that within each frame, all telephone numbers had the same probability of selection.

The total sample response rates shown in Exhibit 1.4.2 are based on combining the landline and cell phone samples in an unweighted fashion. A more complex and arguably superior approach to computing overall survey response rate for dual frame RDD surveys was discussed in the 2010 AAPOR Cell Phone Task Force Report (pp. 52-53). That report recommends combining the response rates for each frame using weights that are proportional to each segment of the population sampled from the respective frame.

To implement this approach, frame figures were derived from the most recent release of telephone service estimates in the National Health Interview Survey (Blumberg and Luke 2012) and re-based on the population of adults living in telephone households. The landline frame is estimated to cover 67.1 percent of the population, and the cell phone frame is estimated to cover 91.5 percent of the population. Nearly one-third of the population (32.9%) lives in a cell-only household, 8.5 percent live in a landline-only household, and 58.6 percent live households with both a landline and cell phone. These frame figures are then used to determine the proportion of the target population represented by the landline sample versus the cell phone sample. The proportion of the population represented by the landline sample is estimated to be 37.8 percent, which is the sum of the percentage landline-only and one half of the percentage in the overlap domain. Similarly, the proportion of the population represented by the cell phone sample is estimated to be 62.2 percent, which is the sum of the percentage cell phone-only and one half of the percentage in the overlap domain. The overall survey response rate is then computed as $(37.8\% \times 18.3\%) + (62.2\% \times 9.4\%) = 12.8\%$. This response rate is perhaps the most accurate rate available for the 2012 Employee Survey, though it may not be the best for comparative purposes due to the fact that this method of integrating frame response rates proportional to their share of the population does not appear to have caught on in the literature (we have not been able to find any published examples).

1.5 Analysis of Nonresponse of the 2012 Employee Survey

The response rate achieved in the 2012 Employee Survey (15.1%) was noticeably lower than that achieved in the 2000 Employee Survey (58.3%). This section has three purposes. First, it examines the potential reasons for this drop off in the response rate. Next, it discusses the implications of nonresponse from a theoretical perspective. Then it examines the potential threat posed by nonresponse to Employee Survey estimates from an empirical perspective based on four nonresponse analyses.

1.5.1 Hypothesized Reasons for the Decline in the Employee Survey Response Rate

We have identified three factors that we assert are likely to account for much of the drop in the response rate: societal changes, challenges associated with the cell phone sample, and differences between the within-household selection procedures used in 2000 and 2012. As noted by Tourangeau (2004), most survey researchers attribute the decline in survey response rates to societal factors. These factors include the general decline in civic engagement (Putnam 1995; see also Groves et al. 2000), increased concern about privacy and confidentiality (Singer et al. 1993), rising hostility toward telemarketers, and the possibility of identity theft. In addition, shifts in the demographic composition of the U.S. population are likely compounding nonresponse. Some of the fastest growing segments of the population (e.g., Hispanics) are known to have generally lower response rates to surveys relative to other Americans.

This constellation of factors has led to a continued, general decline in survey response rates over the past 12 years, particularly in random digit dial (RDD) surveys. To put the 2012 Employee Survey response rate in context, we compiled response rates from RDD surveys with similar sample designs and sponsorship conducted since 2005, which is roughly when survey designers first began supplementing landline RDD samples with cell RDD samples. The surveys and associated response rates (AAPOR (3)) are presented in Exhibit 1.5.1.

Exhibit 1.5.1 Published Response Rates for RDD Surveys Conducted 2005–2011

Survey	Response Rate (AAPOR 3)
Landline RDD	
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%
Cell RDD	
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%

This exhibit has several implications. First, combined (landline plus cell sample) response rates are rare for dual frame RDD surveys. We found no response rates of this nature. Also, 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, though that that is a fairly unusual survey given the extremely low incidence of the target population and the exploratory nature of the project. The national dual frame surveys conducted for the Pew Research Center provide the only national production (not pilot) cell sample response rates that we 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.³ 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 considering only surveys fielded since 2008, the landline response rate is 19%.

We also note that averages are of somewhat limited value because the surveys in Exhibit 1.5.1 differ from each other in content, burden, calling protocols, and other important aspects. That said, they do serve as a guidepost for what can be achieved with RDD designs in current conditions.

A second, closely-related factor contributing to the lower response rate in 2012 is the inclusion of the cell phone sample. As noted in the AAPOR Cell Phone Task Force Report (2010), the response rates to cell RDD samples tend to be lower than those for landline samples conducted as part of the same study. Thus, the mere inclusion of a cell sample in 2012 was expected to lower the Employee Survey response rate somewhat. The exact explanation for lower response rates in cell samples relative to landline samples is not known. Some research suggests, however, that people who only give their cell number to friends and family do not expect to receive survey-type requests on that phone (Kennedy et al. 2009). In other words, for many people the cell phone still appears to be a “private” device on which they are not receptive to contact from unfamiliar callers.

A related issue is that the within-household selection procedure was especially difficult in the cell phone sample. One eligible adult from the screener respondent’s household was randomly selected for the extended interview. When the screener respondent was not selected, this led to a “hand-off” between the screener respondent and the adult selected for the extended interview. Such hand-offs tend to be more difficult in cell phone samples than landline samples because landlines are often a household-level device while cell phones are typically a personal device. Furthermore, approximately one-third of cell phone respondents are reached while they are away from their home (Lavrakas et al. 2009) making it less likely that the recipient of the hand-off is present. Given these circumstances, it is not surprising that studies have shown that cell phone sample response rates are lower when trying to execute a handoff to another person in the household (Brick 2009; Lavrakas et al. 2009). In the 2012 Employee Survey, the extended interview response rate was 33 percent among the cases requiring a hand-off versus 84 percent among the cases in which the screener and extended interview respondent were the same person. This aspect of the selection procedure clearly had a significant negative effect on the response rate. From a nonresponse *bias* standpoint, however, there is no reason to expect that this would systematically affect the survey estimates. Indeed, persons who took leave or had unmet need or were no more or less likely to require a hand-off than employed-only persons.

A third potential factor contributing to the lower response rate in Employee Survey 2012 is the fact that – unlike in 2000 – only one extended interview was completed with each sample household. In the 2000 Employee Survey Methodology Report, there is a reference to the fact that “more than one person per household could be sampled.” Unfortunately there is no indication as to the scale of this clustering, and the publically released dataset does not contain a household identification variable to investigate this. As discussed above, interviewing multiple respondents per household was not done

³ The two most senior survey researchers at the Pew Research Center were both elected president of the American Association for Public Opinion Research.

in 2012 out of concern that it would reduce the statistical power and effective sample size due to intra-class correlation. From a response rate standpoint, however, not accepting multiple respondents per household probably hurt the 2012 response rate when compared to the 2000 rate. Indeed, when the target sample size is fixed (e.g., $n=3,000$) it is generally more difficult to interview respondents in 3,000 unique households than to achieve 3,000 interviews by conducting multiple interviews in some of the same sample households. Details on how the clustering was accounted for in the 2000 response rate calculations—and also in the analysis—do not appear to be available.

1.5.2 The Nature of Nonresponse Bias

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

One reason why estimates from low response rate surveys can still be accurate is the ability to apply statistical weighting to correct for differential nonresponse across demographic subgroups. In particular, it is best practice for survey samples to be statistically adjusted so that the weighted survey data align with benchmark data for the target population, and the 2012 Employee Survey includes such an adjustment. Specifically, the responding sample is aligned to benchmark data for the target population derived from the Current Population Survey using raking ratio adjustment. Raking ratio adjustment can be particularly effective in minimizing the risk of nonresponse bias. Raking ratio adjustment is based on a model that assumes nonrespondents are missing at random from within their respective adjustment cells, so the technique has its limitations. The missing at random assumption is rarely testable in practice. This adjustment does, however, account for many of the well-established correlates of nonresponse that are likely related to survey estimates; specifically age, gender, race and education.

1.5.3 An Empirical Assessment of Nonresponse in the 2012 Employee Survey

Four methods were used to evaluate nonresponse to the 2012 Employee Survey: 1) a nonresponse follow-up survey, 2) a comparison of easier-to-reach versus harder-to-reach respondents, 3) response propensity modeling, and 4) a comparison of survey estimates with external benchmarks. Each of these methods provides a different perspective on the potential risk of nonresponse in the 2012 Employee Survey.

Nonresponse Follow-up Survey (NRFU)

A nonresponse follow-up survey (NRFU) was conducted shortly after the Employee Survey was completed. The NRFU attempted to interview a subsample of nonrespondents to the Employee Survey in order to assess whether nonrespondents had different characteristics than respondents.

Based on lessons learned from the 2000 Employee Survey NRFU, the 2012 NRFU focused exclusively on cases that had completed a screener but failed to respond to the extended interview. In 2000, attempts to reach nonrespondents to the screener yielded too few cases to support any analysis beyond the screener data: Only 2.2% of the households that had not responded to the screener in the main survey yielded a completed interview. This was “too few (interviews) to draw definitive conclusions” according to the methodology report.

In order to avoid this same result in 2012, the NRFU sample was limited to a subset (n=600) of the total n=1,077 households that completed the screener, contained at least one eligible adult, and had not responded to the extended interview. Both landline and cell phone cases were included in the NRFU sample, approximately in proportion to their relative shares of the entire pool of screened nonrespondents from the main survey. The landline cases sub-sampled for the NRFU were matched against a directory of postal addresses with known landlines, and a postal address was appended where available. Both address-matched and unmatched cases were included in the 2012 NRFU because adults living in households that are not in commercial address databases have different demographic and socioeconomic characteristics than individuals living in address-matched households. If the NRFU had been limited to address-matched cases, this could have led to confounds between the NRFU respondents and the main survey respondents. All landline sample cases that were matched to an address (though reverse lookup) received a letter encouraging them to cooperate with the NRFU interview.

NRFU sample cases were offered \$20 post-paid remuneration for completing the interview. The NRFU was conducted via CATI and featured a shortened version of the Employee Survey instrument (see Appendix E). The NRFU was conducted from July 9 to July 31, 2012 and yielded 137 completed extended interviews (98 with landline sample cases and 39 with cell phone sample cases) for a NRFU response rate of 22.8 percent.

About two-thirds of the completed NRFU interviews (64 percent) were with employed-only adults, 9 percent were with respondents with unmet need for leave, and 27 percent were with leave-takers. It is important to note that leave status of respondents (FMLA group) for these cases was known (from the screener data) prior to the fielding of the NRFU. Consequently, FMLA group is not considered an outcome variable in the NRFU analysis. The relationship between FMLA group and nonresponse to the extended interview is discussed in detail in the section on response propensity modeling below. The focus of the NRFU analysis is on employment and leave-related characteristics that were not captured in the screener.

Exhibit 1.5.2 compares unweighted characteristics of all 2,852 Employee Survey respondents to the characteristics of the 137 eligible adults reached in the NRFU. The results suggest no major differences between the NRFU respondents and the Employee Survey respondents. The likelihood of being familiar with FMLA and of having an employer who is covered by FMLA are highly similar for these two groups. Some 77 percent of main survey respondents gave answers that indicated their employer was covered by FMLA, and this compared with 78 percent of the NRFU respondents.

Exhibit 1.5.2 Characteristics of Main Survey Respondents versus NRFU Respondents

	Main Survey Respondents	NRFU Respondents
Employer is covered by FMLA	77%	78%
Respondent is eligible for FMLA	67%	60%
Ever hear of the Family and Medical Leave Act?		
Yes	74%	73%
No	25%	27%
Don't know/Refused	1%	0%
Total	100%	100%
Are you eligible to receive...		
Flexplace or telecommuting (% Yes)	23%	22%
Paid family leave (% Yes) *	49%	42%
Paid vacation (% Yes)	78%	63%
Paid sick time (% Yes) *	70%	59%
Education†		
High school or less	27%	27%
Some college/Associate's	30%	34%
College graduate	42%	39%
Don't know/Refused	1%	0%
Total	100%	100%
Ethnicity†		
Hispanic (% Yes)	9%	12%
Race†		
Black/African-American (% Yes) *	12%	7%
Number of children under age 18 in your care		
None	59%	62%
One or more	40%	38%
Don't know/Refused	1%	0%
Total	100%	100%
Marital Status		
Married/Living with partner	66%	73%
Divorced/Separated	14%	9%
Never married	15%	16%
Widowed	3%	1%
Don't know/Refused	1%	0%
Total	100%	100%
Sample size	2,852	137

† Indicates that this variable was adjusted for in the Employee Survey weighting.

* Difference of proportions test $p < .05$

Some modest differences appear with respect to being eligible for FMLA and certain employer-provided benefits. NRFU respondents were somewhat less likely to give responses indicating that they are eligible for FMLA (60 percent versus 67 percent). The NRFU respondents were also somewhat less likely to self-report being eligible for paid family leave, paid vacation, and paid sick time. These results suggest that the main survey could possibly have somewhat over-estimated employee eligibility for FMLA and these benefits. However, it is also important to note that NRFU respondents were also less likely to have taken leave during the reference period relative to the main survey respondents. It seems plausible that people who take FMLA-related need are perhaps more likely to work for an employer that offers these kinds of benefits.

Exhibit 1.5.3 presents estimates that are based on the leave-takers in the main survey and in the NRFU.⁴ There were only 37 such respondents in the NRFU and so the estimates can only be used to check for very large differences from the leave-takers interviewed in the main survey. No such large differences are evident. The leave-takers from the main survey appear to be quite similar to those from the NRFU with respect to the number of reasons they took leave, the nature of the condition for which the leave was taken, and the circumstances in which they returned to work.

⁴ We also sought to compare NRFU respondents with unmet need for leave to those interviewed in the main survey. Unfortunately only 13 such respondents completed the NRFU extended interview. This case base was too small to support any meaningful analysis.

Exhibit 1.5.3 Characteristics of Leave Takers Interviewed in the Main Survey versus the NRFU

	Main Survey Respondents	NRFU Respondents
Number of total REASONS leave-takers took leave from work in past 18 months		
One reason	69%	72%
Two or more reasons	29%	25%
Don't know/Refused	1%	3%
Total	100%	100%
Number of total REASONS leave-takers took leave from work in past 12 months		
One reason	77%	72%
Two or more reasons	22%	28%
Don't know/Refused	2%	0%
Total	100%	100%
Main reason for most recent leave		
Own illness/disability/other health condition	57%	62%
Maternity-related	1%	5%
Maternity-related and newborn care	2%	0%
Miscarriage	0%	0%
Newborn care	12%	11%
To bond with newborn	2%	0%
To bond with newly placed foster child	0%	3%
Child's health condition	5%	0%
Spouse's health condition	7%	8%
Parent's health condition	9%	8%
Other relative's health condition	3%	0%
Deployment of military member	1%	0%
Don't know/Refused	1%	3%
Total	99%	100%
Nature of condition for the most recent leave		
One-time health matter	44%	34%
Condition requiring routine scheduled care	16%	19%
Condition affecting work from time to time	25%	19%
Other	14%	28%
Don't know/Refused	1%	0%
Time off was taken...		
In one continuous block of time	75%	78%
On separate occasions	24%	22%
Total	100%	100%
After your leave ended, did you...		
Went back to work for same employer	91%	92%
Went back to work for new employer	1%	0%
Did not return to work	8%	8%
Don't know/Refused	1%	0%
Total	100%	100%
Sample size	1,332	37

Note - None of the difference in proportions tests is statistically significant at the .05 level.

Roughly 70 percent of both groups took leave for exactly one reason during the past 18 months, and that tended to be for the respondent's own disability or health condition. The clear majority of leave takers from both surveys took their leave in one continuous block and returned to work for the same employer when the leave ended. The proportion of main survey leave takers reporting that their leave was a one-time health matter (44 percent) is somewhat higher than the proportion of NRFU leave takers reporting this (34 percent), but that difference is not statistically significant and appears to be attributable to the high proportion of "Other" responses among the NRFU leave takers. On balance, the leave-takers from the main survey and the NRFU appear to have quite similar leave experiences. There is little evidence, based on the NRFU for potential nonresponse bias when looking at leave-taker estimates.

Comparison of Easier to Reach versus Harder to Reach Respondents

The second technique used to assess the risk of nonresponse bias is an analysis of the level of recruitment effort. Here we compare the leave-related characteristics of respondents who were easy to reach with respondents who were harder to reach. The harder-to-reach cases serve as proxies for the nonrespondents who never completed the extended 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 nonresponse bias.

In this analysis the level of effort in reaching the respondent is considered with respect to three dimensions: (1) *ease of "contactability"* as defined by the number of calls required to complete the interview; (2) *amenability* as defined by whether or not the case was a converted refusal; and (3) in terms of *both contactability and amenability* as defined by a hybrid metric combining number of call attempts and converted refusal status. Just over half (54.9 percent) of the 2,852 extended interview respondents completed the interview on the first, second, or third call. The remainder (45.1 percent) required at least four calls, with a maximum of 14 calls. About 1 in 25 respondents (4.5 percent) was a converted refusal.⁵ Some 46.4 percent of the respondents either required four or more calls or was a converted refusal. These cases are referred to as "hard to reach" in this analysis, and respondents who never refused and completed the interview in three or fewer calls are referred to as the "easy to reach."

Exhibit 1.5.4 presents several leave-related characteristics for these various groups. In this table each respondent is represented three times according to number of attempts they required (contactability), whether or not they ever refused (refusal behavior), and whether they were easy or hard to reach (hybrid metric).

⁵ In some of these cases the refusal may have come from the screener respondent rather than the extended interview respondent, if these happened to be different people.

Exhibit 1.5.4 Leave-related Characteristics by Level of Effort Groups

	Contactability		Refusal Behavior		Hybrid ^a	
	3 or fewer attempts %	4 or more attempts %	Never Refused %	Converted Refusals %	Easy to Reach %	Hard to Reach %
FMLA group ^{b,c}						
Leave-taker	41.3	35.6	38.8	36.2	41.3	35.7
Unmet need for leave	16.7	14.5	15.9	11.8	16.6	14.7
Employed-only	42.0	49.9	45.3	52.0	42.1	49.6
Total	100.0	100.0	100.0	100.0	100.0	100.0
Employer is						
Not covered by FMLA	23.9	21.5	22.6	28.3	23.6	21.9
Covered by FMLA	76.1	78.5	77.4	71.7	76.4	78.1
Total	100.0	100.0	100.0	100.0	100.0	100.0
Employee is						
Not eligible for FMLA	33.6	33.1	33.3	36.3	33.5	33.3
Eligible for FMLA	66.4	66.9	66.7	63.7	66.5	66.8
Total	100.0	100.0	100.0	100.0	100.0	100.0
Heard of FMLA						
Yes	73.0	74.2	73.4	76.4	72.8	74.5
No	25.8	25.1	25.6	22.1	26.0	24.8
Don't know/Refused	1.2	0.7	1.0	1.6	1.2	0.8
Total	100.0	100.0	100.0	100.0	100.0	100.0
Minimum sample size	1,404	1,168	2,459	113	1,372	1,200

Source: Employee Survey, figures are unweighted

^a The easy to reach group consists of respondents who completed on three or fewer attempts and never refused. The hard to reach group consists of respondents who required four or more attempts or were a converted refusal.

^b Indicates that the chi-square test the difference between the Contactability classes is statistically significant at the .05 level.

^c Indicates that the chi-square test the difference between the Hybrid classes is statistically significant at the .05 level.

The only significant results from the table is that respondents who completed the interview on the first few attempts were more likely to have taken leave (41.3%) or had unmet need for leave (16.7%) than those who took (35.6%) or had unmet need for leave (14.5%) but completed the interview after four or more attempts (combined 58.0% versus 50.1%, chi-square $p < .001$). One clear post-hoc explanation is that the employees with unmet need for leave and leave-takers may have felt that the survey was more relevant to them, and they may have therefore been more eager to participate than those who with no relevant leave experiences. This result does not hold when looking at converted refusers versus those who never refused. The pattern is significant for the hybrid measure, but this simply reflects the fact that the hybrid measure is largely a function of the number of call attempts.

The significant relationship between contactability and leave taking/unmet need for leave may be related to the fact that the survey introduction announces the fact that this was a survey about medical leave. Respondents were told, “We are conducting a national study to find out about employees’ use of, and attitudes about, family and medical leave policies in their workplace.” For many people who had not taken leave, this introduction may have led them to conclude that the survey was not important to them. This dynamic is predicted by Leverage Salience Theory (Groves et al. 2000), which posits that sample members base their cooperation decisions on the aspects of the survey that are made salient to them during recruitment. In this case, the topic was made very salient, possibly to the detriment of the composition of the responding sample. If future Employee Surveys are conducted, consideration should be given to not including such an explicit statements about the survey content. The balance between informed consent about the survey and the threat posed by differential nonresponse is one that should be discussed with relevant institutional review boards and oversight agencies (e.g., OMB).

The exhibit also shows how easy and harder to reach respondent groups compare with respect to employer coverage, employee eligibility for FMLA, and whether or not the respondent had heard of FMLA. On all three of these measures, there were no differences between the easy to reach respondents and the harder to reach respondents. The negligible differences observed for these other measures suggests that other survey variables are likely to be unrelated to this easy/hard to reach dimension, especially when the analysis is conditioned upon individuals belonging to a given FMLA group (leave-taker, unmet need for leave, employed-only).

Response Propensity Modeling

The third technique used 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. The response propensity model allows the researcher to identify the most powerful predictors of response when all available predictors are tested simultaneously. In this analysis, the primary research question is whether or not employment-related or leave-related characteristics are associated with response propensity, especially when controlling for factors included in the weighting protocol. If employment-related or leave-related variables show a significant association with response to the extended interview (after controlling for other factors), this would be evidence of possible nonresponse bias. If, however, the employment and leave-related predictors do not have a significant effect, this suggests that the weighting adjustments are likely to have been effective in reducing nonresponse bias.

In order for a response propensity model to be informative, the researcher must know the values for respondents and nonrespondents on one or more predictors of survey response. In RDD surveys, propensity models are often quite limited because little information is generally known for the nonrespondents. This is the case for the screener component of the Employee Survey. The only types of variables known for the nonrespondents to the screener are sampling frame, region, and level of effort data. Nothing is known about the employment or leave-related characteristics of the screener nonrespondents.

A much richer model is possible, however, if we condition on cases completing the screener and model propensity to respond to the extended interview. Based on the screener, we know the age, gender, telephone service, FMLA group, employment sector, and other variables for both the respondents and nonrespondents to the extended interview. This response propensity analysis, thus, conditions on households completing the screener and examines which variables were associated with response to the extended interview.

A logistic regression was used to model response to the extended interview conditional upon completion of the screener. The results are presented in Exhibit 1.5.5.⁶ The strongest predictor of response to the extended interview is the hand-off flag. The hand-off flag has value 1 for cases in which the screener respondent happened to be the person selected for the extended interview and value 0 for cases in which these screener respondent needed to hand-off the phone because someone else in the household was selected to complete the extended interview. This result reflects the fact that in each sample household, only one eligible adult was selected to complete the extended interview. *The extended interview response rate was 33 percent among the cases requiring a hand-off versus 84 percent among the cases in which the screener and extended interview respondent were the same person.*

⁶ Interviewing effort variables, such as the number of call attempts and an indicator for converted refusal cases, were intentionally excluded from this model because they are endogenous and also because a significant association with the outcome being modeled would not communicate any information about the potential risk to survey estimates from nonresponse bias. Interviewing effort variables are considered separately in the analysis of easier to reach versus harder to reach cases.

Exhibit 1.5.5 Logistic Regression Estimating the Probability of Response to the Extended Interview Conditional on Completion of the Screener

Parameter	Estimate	s.e.	Wald χ^2	p value
Intercept	0.528*	0.215	6.02	0.01
Sampling frame = Cell Phone RDD	-0.168**	0.057	8.65	<.01
Hand-off = Yes	-1.232***	0.043	835.07	<.0001
Sampling frame x Hand-off (interaction)	-0.133**	0.042	9.97	<.01
Region = Northeast	-0.125	0.067	3.52	0.06
Region = Midwest	0.105	0.066	2.52	0.11
Region = South	-0.027	0.058	0.21	0.64
Number of eligible adults in HH (log)	-0.262	0.144	3.33	0.07
FMLA group = leave-taker	0.030	0.053	0.31	0.57
FMLA group = unmet need for leave	-0.135*	0.066	4.18	0.04
R gender = male	-0.107**	0.037	8.15	<.01
Telephone service = landline and cell	-0.043	0.080	0.29	0.59
Telephone service = cell-only	0.116	0.116	1.01	0.32
R employment sector = government	0.038	0.070	0.29	0.59
R employment sector = non-profit	0.046	0.090	0.27	0.61
R age	0.003	0.003	1.37	0.24

Model Diagnostics:

Area under ROC curve (c) 0.792

-2 Log Likelihood 4,590.6

Sample size 4,498

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

Reference groups for categorical variables: hand-off (no, the screener respondent was selected as the extended interview respondent), region (West), FMLA group (employed-only), telephone service (landline-only), R employment sector (private sector), R gender (female)

The sampling frame was also associated with extended interview response. Landline sample cases were somewhat more likely to complete the extended interview than cell phone sample cases. Based on the literature we expected that the hand-off issue would be more problematic in the cell phone sample than in the landline sample, so an interaction term was included in the model. The interaction of sampling frame and hand-off flag was statistically significant ($p < .01$). Indeed, while hand-offs decreased response propensity in both samples, the effect was stronger in the cell sample than the landline sample.

In terms of potential nonresponse bias, these findings do not necessarily represent cause for concern because the weighting protocol addresses the integration of the sampling frames, and it also includes a post-stratification to telephone service groups (landline-only, cell-only, dual service). Furthermore, there is no evidence that the incidence of hand-offs varied across the FMLA groups (based on a separate analysis, not shown in the Exhibit).

The only other highly significant effect observed in the model was for the gender of the adult selected for the extended interview. Over two-thirds (68 percent) of the women selected completed the extended interview, compared with 59 percent of the men selected (not shown). It is important to note that the weighting included an extended interview nonresponse adjustment for both gender and age. This weighting adjustment is expected to have minimized the risk on nonresponse bias associated with this effect from gender shown in Exhibit 1.5.5.

The other result of note in Exhibit 1.5.5 is the marginally significant effect associated with having an unmet need for leave. In bivariate analysis, the extended interview response rates (conditional on screener completion) were 61 percent for employees with unmet need for leave, 64 percent for the leave-takers, and 64 percent for the employed-only. This variation is quite modest, and indeed the Chi-Square test in the bivariate analysis is not statistically significant ($p=0.38$). It is somewhat puzzling that this non-significant bivariate result becomes marginally significant in the multivariate model in Exhibit 1.5.5. We were unable to identify any obvious post-hoc explanation for this result, especially in light of the lack of association between these FMLA groups and the main predictor of response, handing-off. The effects associated with region, the number of eligible adults in the household (log transformed), household telephone service, employment sector, and age were all non-significant at the $\alpha=0.05$ level in the model.

In sum, the fact that the bivariate association between response and FMLA group is not significant and that the Wald test for the employees with unmet need for leave coefficient in the model is only marginally significant, amounts to little evidence of potential nonresponse bias. The risk to estimates from nonresponse bias appears to have been more substantial at the screener stage than at the extended interview stage.

Comparisons to External Benchmarks

One limitation of the previous three 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 nonrespondents; the level of effort analysis relies on the “harder-to-reach” respondents as proxies for nonrespondents; the response propensity model captures only variation between the screened extended interview respondents and the screened extended interview nonrespondents. In this section we make comparisons to external benchmarks in order to evaluate the total level of nonresponse bias in the 2012 Employee Survey.

Specifically, we compare the weighted final respondent estimates from the Employee Survey to the Current Population Survey (CPS). The CPS is considered to be a “gold standard” survey due to its more rigorous protocol (e.g., area-probability sampling with in-person interviewing) and a higher response rate than the 2012 Employee Survey. By virtue of its more rigorous design, the estimates from the March 2011 CPS are assumed to contain less nonresponse bias than those from Employee Survey.

The strength of this approach is that the benchmark survey (CPS) is well known to be a high quality federal survey, and so obtaining similar estimates would give some confidence about the 2012 Employee Survey (Groves 2006). One weakness of this approach is that not all of the key survey variables in the Employee Survey are collected in the CPS, and so the analysis is somewhat limited in scope. Another weakness is that the measurements collected in the 2012 Employee Survey are not identical to the measurements collected in the CPS. The CPS features in-person interviewing in addition to the CATI data collection used exclusively in the Employee Survey. Furthermore, the question wording for the comparison questions varies somewhat between the two surveys. Either of these factors may lead to measurement error differences contaminating the comparison. A third weakness of this approach is that the coverage and nonresponse characteristics of the CPS are not completely known. While the CPS provides the best available estimates for the comparison measures, the CPS estimates may themselves contain some level of error (beyond sampling error).

CPS weighted estimates were computed based on the population of adults aged 18 and old with a telephone who were employed for pay within the past 12 months (excluding self-employed) to match the target population of the Employee Survey target as closely as possible. Three variables identified in the CPS were also administered in the Employee Survey but not used in the weighting protocol.⁷ These variables are marital status, union membership, and employment status. The weighted estimates from both surveys are presented in Exhibit 1.5.6.

Exhibit 1.5.6 Weighted Estimates from the Current Population Survey and Employee Survey

Characteristic	Current Population Survey %	Employee Survey %	Difference %
Marital Status			
Married	55.1%	54.2%	-0.9%
Not married	44.9%	45.8%	0.9%
Labor Union Membership			
Yes	11.4%	14.5%	3.1%
No	88.6%	85.5%	-3.1%
Current Employment Status			
Employed	95.1%	88.4%	-6.7%
Unemployed/Not in Labor Force	4.9%	11.6%	6.7%

Sources: 2012 Employee Survey and March 2011 CPS. Estimates from both surveys are weighted. Estimates exclude item nonresponse.

The weighted Employee Survey estimate for percent married is highly similar to the estimate from the CPS (54.2 percent versus 55.1 percent). The full array of response options (e.g., separated, widowed) are not compared here because the list of options differed between the two surveys. With respect to married/unmarried, however, this comparison suggests minimal potential for nonresponse bias.

There was a somewhat larger discrepancy, however, with respect to union membership. The estimate from the Employee Survey is that 14.5 percent of the target population are union members, which compared with 11.4 percent from the CPS. One post hoc explanation for this difference is that union members may be more attuned to issues of benefits and employed leave policy, and they may have therefore been more interested in participating in the Employee Survey relative to non-union workers. Unfortunately, there are no data available to test that hypothesis.

There is also a noticeable difference for the estimated percent currently employed. The estimate from the CPS is nearly seven percentage points higher than the estimate from the Employee Survey. Based on the available information, this difference appears to be attributable to a definitional peculiarity in the CPS difference rather than nonresponse. When the CPS public use micro dataset is filtered on the population of adults employed for pay within the past 12 months (excluding self-employed), the estimate for percent currently employed does not include persons who are unemployed and not

⁷ Several demographic variables such as age, gender, education, and race/ethnicity are measured in both the CPS and the Employee Survey. These variables were intentionally excluded from this analysis, however, because they were included in the raking ratio estimation for the Employee Survey weights. In other words, the Employee Survey was statistically adjusted to match external benchmarks on these measures, and so comparing those weighted characteristics to the CPS would not be informative about the risk of nonresponse bias.

looking for work. Such individuals are considered out of the labor force and so they do not appear in the denominator of this estimate. In the Employee Survey, however, such individuals were asked about their current employment status and they are represented in the denominator of the estimate in Exhibit 1.5.6. Given that this definitional difference may explain at least some of the discrepancy between these two estimates, we do not necessarily view the current employed metric as an especially informative point of evaluation. It is included here for the sake of comprehensiveness as well as to illustrate the kinds of issues that can limit generalizability from a benchmark comparison analysis.

1.5.4 Summary of Nonresponse Analysis for the Employee Survey

These analyses suggest that nonresponse bias may pose a small risk to some of the estimates from the Employee Survey. Perhaps the most informative result is the fact that harder-to-reach respondents (those requiring four or more call attempts) were less likely to have taken leave or have an unmet need for leave, relative to those who were easy to reach. As discussed above, one potential post hoc explanation is that the leave-takers and employees with unmet need for leave may have felt that the survey was more relevant to them, and they may have therefore been more eager to participate than those who with no relevant leave experiences.

Generally speaking, level of effort analysis is not particularly rigorous or definitive (e.g., Montaquila et al. 2008) because it only evaluates variation among the survey respondents and it relies upon a fairly strong assumption that the harder-to-reach respondents are good proxies for the nonrespondents who are never reached. That said, one advantage of the level of effort analysis in this context is that it does not speak to just the screener survey or just the extended interview; the harder-to-reach respondents are potentially signaling a pattern that carries over to the entire set of nonrespondents to the Employee Survey. In this way, the level of effort analysis is different from the NRFU and the response propensity model because in this analysis those two approaches only address nonresponse to the extended interview.

A nonresponse follow-up survey (NRFU) is quite similar in spirit to a level of effort analysis. In this study, however, the NRFU was purposefully limited to nonrespondents who had completed the screener. This decision was based upon our review of the 2000 Employee Survey NRFU that essentially found it impractical to attempt to reach the screener nonrespondents due to issues of low cooperation and low target population incidence. Analysis of the 2012 Employee Survey NRFU generally found small difference between the main survey respondents and the main survey nonrespondents who completed the NRFU. This was especially true for estimates based on leave-takers. That said, NRFU respondents were somewhat less likely to give responses indicating that they are eligible for FMLA and somewhat less likely to self-report being eligible for paid family leave, paid vacation, and paid sick time.

The response propensity analysis and the benchmark comparison analysis provided little evidence that nonresponse bias was a major threat to the Employee Survey estimates. Both analyses were limited, however, in the variables that were available. The response propensity modeled showed that hand-offs between the screener and extended interview respondent greatly decreased the likelihood of response, but there is no indication that this compromised any estimates from the survey. Testing revealed no association between the hand-offs and FMLA group, as would be expected. In the benchmark comparison analysis, the limited set of comparison variables and differences in measurement and definitions limit generalization to other estimates in the Employee Survey.

1.6 Weighting

Several stages of weighting adjustments were needed to account for the complex nature of the Employee Survey design. The weights account for numerous factors including (1) the probability of selection of the phone number from the sampling frame, (2) the presence of unresolved numbers in the sample (unknown eligibility), (3) nonresponse to the screener, (4) the probabilities of selection of the extended interview respondent, (5) the subsampling of households where the “employed-only” group was selected, (6) the overlap of the landline and cell RDD frames, and (7) nonresponse to the extended interview. Missing data on the variables used in the weighting were filled using hot deck imputation.

1.6.1 Probability of Selection of the Telephone Number

The first step in the weighting process was to account for the probability of selection of the telephone number. This adjustment was computed separately for the landline and cell phone samples, and it equaled the total size of the sampling frame divided by the total count of numbers released in the sample. The adjustment values were 1,339.1429 for the landline sample and 8,778.2236 for the cell phone sample.

1.6.2 Screening for the Target Population

This was followed by an adjustment for nonresponse to the screener. The adjustment was computed as the ratio of all residential telephone numbers divided by the count of telephone numbers for which a screening interview was completed. A portion of sample numbers in each frame could not be confirmed as working and residential (e.g., ring no answer all attempts, busy all attempts). The eligibility rate (here eligible refers to “working and residential”) among these undetermined numbers was estimated empirically using the eligibility rate among the number for which eligibility was determined. The count of undetermined numbers estimated to be eligible (working and residential) was computed separately for each Census region and added into the numerator of the screener nonresponse adjustment. The adjustment values are presented in Exhibit 1.6.1.

Exhibit 1.6.1 Screener Nonresponse Adjustment Value, by Census Region

Census Region	Screener Nonresponse Adjustment Value
Northeast	3.9456
Midwest	3.6363
South	3.9358
West	4.0796

1.6.3 Multiplicity in the Sampling Frames

The next weighting step makes an adjustment for multiplicity within the sampling frame. Households with multiple voice-use landlines could have been selected more than once into the landline sample. To address this, landline sample cases were weighted by the inverse of the number of working voice-use landlines in the household. The denominator of this adjustment was capped at two in order to avoid excessive variance in the weights. The capping affected the 1.2% of the landline cases (where three or more landlines were reported).

A similar adjustment for multiplicity was made for the cell sample cases. The probability that a household was reached in the cell phone sample was a function of the number of cell phones used by adults living in the household.⁸ To address this, cell sample cases were weighted by the inverse of the total number of cell phones used by adults in the household. The denominator of this adjustment was capped at four in order to avoid excessive variance in the weights. The capping affected the 5.8% of the landline cases (where five or more cell phones were reported). The Employee Survey provided in Appendix A shows the question wording used to measure household telephone usage.

1.6.4 Within-household Selection Procedure

The product of the probability of selection of the telephone number adjustment, the screener nonresponse adjustment, and the multiplicity in sampling frame adjustment is a household-level weight that accounts for factors influencing the probability that the household was selected for the Employee Survey. The next stage of weighting addresses the fact that we only sampled one eligible respondent from each household to complete the extended interview. The within-household selection procedure is described above in Section 1.1.2 and is not repeated here.

The first adjustment for within-household selection accounts for the fact that within each household, one family and medical leave *group* was selected among those present in household based on the roster information. When reported in the household, the leave-needer and leave-taker groups were selected with a higher probability relative to the employed-only group. When both the leave-needer and leave-taker groups were present, the leave-needer group was selected with a higher probability than the leave-taker group. The adjustment was computed as the inverse of the probability that the family and medical leave group was selected. The values of this adjustment ranged from 1.0 to 10.0, reflecting the imperative to interview as many leave-needers and leave-takers as possible within a probability-based framework.

The next step was to adjust for the fact that households in which the employed-only group was selected were subsampled for the extended interview. The subsampling was motivated by the need to contain study costs while minimizing the loss of information. The adjustment applied to the subsampled cases was computed as the inverse of the subsampling rate applied for the sample replicate. In most replicates the subsampling rate was 20 percent (yielding a weighting adjustment of 5.0) but the subsampling rate did vary across replicates released toward the end of the field period.

The final weighting adjustment related to within-household selection accounts for the fact that only one adult from the family and medical leave group selected in the household was randomly selected to be the extended interview respondent. Specifically, the adjustment was computed as the number of eligible adults in the household who belonged to the family and medical leave group selected. For example, if the leave-taker group had been selected for the household and there were three leave-takers reported in the roster, then the adjustment would have been 3.0.

For sampling and weighting purposes it was necessary to classify all eligible adults into mutually exclusive groups (leave-needer, leave-taker, employed-only). In reality, however, some people both needed leave and took leave during the reference period. In the substantive survey report,

⁸ Cell phones answered by children under age 18 were treated as ineligible sample units. No screening was attempted for these cases.

respondents who were both leave-needers and leave-takers are addressed carefully. When they are combined with one group or the other, this is explicitly mentioned. For purely sampling and weighting purposes (not survey estimation), people who were both leave-needers and leave-takers were treated as leave-needers because that is the rarer of the two characteristics. We wish to emphasize that this treatment for sampling and weighting purpose is completely separate from how this group is handled in the final survey estimates.

1.6.5 Sampling Frame Integration

The next weighting step accounts for the overlap between the landline RDD and cell RDD sampling frames. The dual service (landline and cell-only) households from the two frames were integrated in proportion to their effective sample sizes (Frankel et al. 2007). The first effective sample size was computed by filtering on the dual service landline sample cases and computing the coefficient of variation (cv) of the final base weight. The final base weight is defined as the product of the aforementioned adjustments (as appropriate for each case). The design effect for these cases was approximated as $1 + cv^2$. The effective sample size (n_1) was computed as the unweighted sample size divided by the design effect. The effective sample size for the cell frame dual service cases (n_2) was computed in an analogous way. The compositing factor for the landline frame dual service cases was computed as $n_1/(n_1 + n_2)$. The compositing factor for the cell phone frame dual service cases was computed as $n_2/(n_1 + n_2)$. The compositing values were 0.8348 for landline sample dual service cases and 0.1652 for cell phone sample dual service cases. The sample frame integration was set to 1.0 for landline-only and cell phone-only cases, which do not belong to the overlap domain.

1.6.6 Extended Interview Nonresponse

A subset of the respondents selected for the extended interview did not complete the full interview. The survey nonresponse literature (for example, Groves and Couper 1998) suggests that propensity to nonrespond to the extended interview may be correlated with demographic (among other) factors related to survey variables. An extended interview nonresponse adjustment for the combined sample was included in the weighting in order to reduce the risk of nonresponse bias to the survey estimates. This adjustment was conducted within weighting classes and was based on the roster information for the gender and age of the selected extended interview respondent. The weighting classes were defined as the cross-classification of gender and age (three categories). The adjustment was computed as the weighted sum of responding and nonresponding persons in the weighting class divided by the weighted sum of respondents in the weighting class. The adjustment values for each weighting class are presented in Exhibit 1.6.2.

Exhibit 1.6.2 Extended Interview Nonresponse Adjustment Values, by Weighting Class

Weighting Class	Extended Interview Nonresponse Adjustment Value
Male age 18-29	1.6036
Male age 30-49	1.5927
Male age 50+	1.5810
Female age 18-29	1.6556
Female age 30-49	1.4751
Female age 50+	1.4681

1.6.7 Post-Stratification to the Current Population Survey (CPS)

To help reduce possible residual nonresponse and non-coverage errors, the final estimation weights included raking ratio adjustment to reflect the most recent population information available. The

demographic control totals were computed from the March 2011 Current Population Survey, Annual Social and Economic Supplement (CPS-ASEC) public use microdataset. Telephone service control totals were computed from the 2010 NHIS public use microdata file and updated using the most recent topline release of estimates for the entire adult population (Blumberg and Luke 2012). The survey data were aligned to population control totals for gender, age, education, race/ethnicity, region, and phone service.

1.6.8 Handling of Extreme Weight Values

The distribution of the post-stratified weights was examined for any extreme values. The distribution of the final weights was truncated at the 99th percentile. This trimming was performed in order to reduce extreme variance in the weights and ultimately improve the precision of the weighted survey estimates. The sum of the final weights was set to equal the total number of net U.S. adults in the target population based on the March 2011 Current Population Survey. Exhibit 1.6.3 reports the coefficient of variation, minimum, and maximum values of the final weight before and after the trimming.

Exhibit 1.6.3. Coefficient of Variation, Minimum, and Maximum Values of the Employee Survey Final Weight, Before and After Trimming

	Before Trimming	After Trimming
Coefficient of variation	1.433	1.301
Minimum value	2,260.108	3,211.342
Maximum value	913,742.287	334,493.394

1.7 Variance Estimation

Due to the complex nature of the 2012 Employee Survey, formulas commonly used in RDD surveys to estimate margins of error (standard errors) are inappropriate. Such formulas would understate the true variability in the estimates. To account for the complex design, a repeated sampling technique (specifically jackknife delete two repeated replication (Krewski and Rao 1981)), was used to create replicate weights for this study. The subsamples (replicates) were created using the same sample design, but deleting a portion of the sample, and then weighting each subsample up to the population total. A total of 80 replicates were created by combining telephone numbers to reduce the computational effort. Statistical significance tests presented in the report were computed using appropriate complex survey software and procedures.

2. The 2012 Worksite Survey

This chapter presents the methods employed to design and administer the 2012 Worksite Survey, including the following:

- Target Population and Sample Design
- Questionnaire Development
- Data Collection Procedures
- Response Rate Calculations
- Analysis of Nonresponse
- Weighting
- Variance Estimation

2.1 Target Population and Sample Design

The 2012 Worksite Survey was a sequential multi-mode (Web and CATI) survey of 1,812 U.S. business establishments. The field period was March 12 through June 15, 2012. A total of 634 interviews were completed on the Web and 1,178 interviews were completed by CATI. The target population consisted of all private-sector business establishments excluding self-employed businesses without employees, government entities, and quasi-government entities (e.g., public educational institutions and post offices). This sample universe differs from that Employee Survey, which includes both private and public employees.

As in the 1995 and 2000 surveys, the sample was drawn from the Dun's Market Identifiers (DMI) file, which provides all essential frame information (e.g., employee size, NAICS code, contact information) for 15.2 million private business establishments. This DMI file is considered the most comprehensive commercially available business list. For the purposes of the sample, an establishment was defined as the business located at a particular address or location. Following the FMLA statute and regulations, in the main report we refer to "establishments" as "worksites". Data were collected with respect to this location, even if the employer had other locations. The target respondent for each worksite was the human resources director or the person responsible for the company's benefits plan.

2.1.1 Stratification

The sample design for the 2012 survey mirrored that used for the 1995 and 2000 surveys. Maintaining the same sample specifications served to keep reports in each survey as comparable as possible. The sampling frame was stratified by the cross-classification of size (number of employees) and North American Industry Classification System (NAICS) grouping. Stratification increases sample efficiency by ensuring that the distribution of sample units mirrors the distribution of the population on the stratification dimensions. As a result, stratification can increase the precision of survey estimates associated with the stratification dimensions. Stratification can also help to achieve a sufficient number of interviews in key stratification cells. For example, just 0.2 percent of establishments on the DMI frame have 250 or more employees, but this group of establishments represents 17.0 percent of employees. Under a simple random sample, the sample size for this

subgroup of large establishments would not have been expected to support the robust analysis that was achieved by the stratified design.

The Worksite Survey used the following employment size classes: (1) 1-9; (2) 10-19; (3) 20-49; (4) 50-99; (5) 100-249; (6) 250-999 (7) 1,000+. This classification facilitates comparison to the Quarterly Census of Employment and Wages (QCEW), which was used in the weighting protocol described below. In order to stratify by industry, the Worksite Survey used four groups based on 2-digit NAICS codes. These groups are reported in Exhibit 2.1.1. Cross-classifying seven employer size groups and four industry groupings yields 28 sampling strata (7 size categories times 4 NAICS groups).

Exhibit 2.1.1 NAICS Groupings for 2012 Worksite Survey

Group	NAICS Codes
NAICS Group I	Agriculture, Forestry, Fishing and Hunting (11); Mining, Quarrying, and Oil and Gas Extraction (21); Construction (23); Manufacturing (31-33)
NAICS Group II	Utilities (22); Wholesale Trade (42); Retail Trade (44-45); Transportation and Warehousing (48-49)
NAICS Group III	Information (51); Finance and Insurance (52); Real Estate and Rental and Leasing (53); Professional, Scientific, and Technical Services (54); Management of Companies and Enterprises (55); Administrative Support and Waste Management and Remediation Services (56)
NAICS Group IV	Educational Services (61); Health Care and Social Assistance (62); Arts, Entertainment, and Recreation (71); Accommodation and Food Services (72); Other Services (81)

The process of allocating sample to the 28 strata featured three steps. First, consistent with the 2000 Establishment Survey, the sample was allocated to establishment size classes proportional to the *square root* of the aggregate number of employees working for establishments in the class. This allocation facilitates a robust sample size for large establishments without yielding a sample distribution as extreme as would be realized under a strictly proportional allocation. Next, the sample sizes for the strata for establishments of size 20-49 and 50-99 were further increased over the initial allocation in order to support special analysis of establishments in these size groups, which straddle the FMLA coverage threshold of 50 employees. Finally, within each size class, the sample was allocated to the industry groups proportional to the aggregate number of establishments in the industry group. All estimates in the survey report are adjusted for the over-sampling by weighting establishments by the inverse of their probability of selection.

2.1.2 Consistency in the 2000 and 2012 Worksite Surveys

The sample design for both the 2000 and 2012 surveys used the DMI as the sampling frame, stratified by size and industry, allocated to size groups proportional to the square root of the aggregate number of employees in a given size class, and allocated to industry groups proportional to the number of establishments in the industry group. The definition of the target population was also kept consistent across the surveys.

2.1.3 Notable Differences Between the 2000 and 2012 Worksite Surveys

A major design difference between the earlier Worksite Surveys (1995 and 2000) and the current survey is that the 2012 survey is the first to allow the informant to complete the survey instrument on the Web. Some 35 percent of the interviews in 2012 were completed on the Web, and 65 percent

were completed with CATI. In prior iterations of the Worksite Survey, all interviews were conducted using CATI. Provision of the Web option was expected to bolster the overall response rate by giving key informants additional flexibility in the time, location and pace of completing the survey. This is particularly important given the potential need for the respondent to consult with administrative records. Consistent with this consideration, the CATI system was designed to allow the respondent to leave and re-enter the survey as frequently as they wished and at any time.

Any time multiple modes are used to collect survey data, careful consideration should be given to the possibility of mode effects. The International Handbook of Survey Methodology (de Leeuw et al. 2008) defines “mode effect” as: “The effect that using a specific mode has on the responses that are obtained in that mode.” At the design stage, this issue was addressed by standardizing the questionnaires administered to Web and CATI respondents as much as possible.

While a mode effects analysis is beyond the scope of this report, several observations are relevant. The 2012 Worksite Survey, like most multi-mode establishment surveys, did not feature a “control group” of sample units that were administered the CATI mode but not the Web mode. Absent this control group, one cannot rigorously test for mode effects because mode is confounded with amenability toward completing the survey and perhaps other variable which were not measured.⁹ Web respondents were generally more amenable than CATI respondents.

The decision to not use a control group for this particular study was intentional. Using a control group would have increased survey costs, yet any effect from mode was expected to be minimal. Key informants responding in both modes were made aware of the fact that the study was being conducted for the Department of Labor. If they were concerned about reporting certain facts to the federal government, this concern would likely have influenced the response process in either mode. The anticipated risk from a mode effect did not justify the higher cost associated with a control group. If future Worksite Surveys are conducted, however, it might be informative to empirically test (e.g., via a randomized design) the assumption that mode has no meaningful effect on response distributions.

A small change that was made to the Worksite Survey in 2012 pertains to the definition of the sampling strata. While the stratification variables—employee size and industry—are consistent with the 2000 survey, the definition of the classes was updated to reflect current federal classifications. The employee size classes were changed from 1-10, 11-24, 25-49, 50-99, 100-250, 251-999, 1000+ in 1995 and 2000 to 1-9, 10-19, 20-49, 50-99, 100-249, 250-999, 1,000+ in 2012. This change was made because the older size classifications do not match the size classifications used in the Quarterly Census of Employment and Wages (QCEW). The industry codes were also changed to reflect the conversion from Standard Industrial Classification (SIC) codes to NAICS. Attempts to reproduce the classification of the 1995 and 2000 surveys in NAICS codes yielded unbalanced sample sizes between groupings. Instead a fresh set of NAICS groupings were defined. The industry groups used in the 2012 Worksite Survey are presented above Exhibit 2.1.1.

Finally, the response rate for the 2012 Worksite Survey (20.9%) is noticeably lower than that reported for the 2000 Establishment Survey (65.0%). We conducted an extensive nonresponse analysis and discuss the potential implications in Section 2.5 below.

⁹ Alternative mode effects approaches (e.g., propensity-based) have been proposed, but they tend to rely on strong model assumptions that are generally not testable.

2.2 Questionnaire Development

The questionnaire for the 2012 Worksite Survey consists of two primary components: a screening questionnaire and a main survey. The screener was used to confirm that sampled establishments still existed and to obtain the correct contact information for the most knowledgeable person at the firm to answer the main survey. Screening also confirmed that the establishment was neither a government nor quasi-governmental organization. The main survey consists of questions designed to gather information on leave taking policy, the number of employees who took leave under FMLA and the impact of leave taking on the business.

As much as possible, the 2012 surveys followed the model implemented in the 2000 surveys. That strategy preserves comparability, allowing analyses of changes over time. Despite efforts to maximize comparability, the final 2012 surveys differ substantially from the 2000 instruments to reflect the many economic and labor force changes that have occurred since that time. We anticipated that the Web option would ease burden of reporting administrative data. Indeed, the 2012 survey response rate to Q20 (amount of leave taken at the establishment) improved considerably over the 2000 survey: 78 percent of respondents answered Q20 on the number of leaves, compared to 45 percent in the 2000 survey.

The questionnaire development process for the 2012 Worksite Survey proceeded in three main phases, described below:

- 1) Updates to Screener
- 2) Updates to Main Questionnaire
- 3) Pilot Surveys

These steps are described in greater detail below. Appendix C includes a question-by-question crosswalk between the 2000 and 2012 questionnaires. Appendix B includes a copy of the Worksite Survey questionnaire.

To facilitate comparisons and identification of trends, we began with the 2000 survey as a base. To ensure that new questions adequately capture the range of issues and experiences with regulatory changes since 2000 and the possibilities for future efforts, we conducted in-depth interviews DOL staff and a listening session with groups of knowledgeable stakeholders.

To update the content of the main survey questionnaire, we gathered information from various sources, including the public comments in response to the DOL's 2006 RFI, peer-reviewed published literature as well as "gray" literature (e.g., relevant newspaper articles, policy papers). We conducted in-person interviews with staff at the Department of Labor. Additionally we held two listening group sessions with representatives from nine employee and four employer stakeholder organizations to elicit their feedback. The questionnaire drafts were reviewed by a Technical Working Group. A discussion of key questionnaire revisions and additions follows below.

2.2.1 Updates to Screener

In addition to collecting the name, address, and phone number of the most knowledgeable person in the firm to answer questions about FMLA administration, the screening survey for the 2012 Worksite Survey was updated to also collect an email address. The 2012 main survey was available as a self-

administered Web survey prior to attempting to collect data by phone. A pre-notification letter was mailed to all screened establishments. The letter explained the purpose and sponsorship of the survey and provided each respondent with a unique URL address to access the survey online.

Two new questions (V9 and V10) were added to the 2012 screening survey. The first new question established that the firm internally maintains (versus outsourcing the administration of FMLA) records of FMLA leave. The second new question identified the method of processing FMLA requests for leave. These data were originally collected in the screener for special consideration of employers who outsource their FMLA requests to a third party, in the event that a separate procedure would be required for contacting third party administrators. Special consideration was not necessary. These questions are asked again in the 2012 main survey questionnaire because the respondent to the screener questionnaire may not have been the same person as the main survey respondent.

2.2.2 Updates to Main Questionnaire

In this subsection, we discuss changes to the main questionnaire.

Reference Period

We conjecture that the high item nonresponse to questions in the 2000 survey about the total amount of leave taken was (in part) because it was difficult to tabulate answers for the non-standard reference period of 18 months. Thus, the 2012 survey allowed employers to choose the most convenient 12-month period on which to report, be it a calendar year, fiscal year, or other 12-month period for which FMLA records could be easily accessed. The questionnaire asked the respondent to select the twelve month reporting period. This eliminates comparability of employer answers with earlier surveys.

Leave Care Designation

The 2012 survey updated the language of questions on coverage for FMLA qualifying conditions, as shown in the revised question wording (Q16) below. Specifically:

- The 2012 Survey removed any references to parents, fathers or mothers from the question stem.
- “Maternity Leave” was changed to “A pregnancy-related reason”
- Two reasons were added to capture information about coverage for military leave entitlements.

The new question format is shown below:

Q16. For employees at this location, does this site's policies allow for family or medical leave for the following reasons?

[INSERT GRID – ROWS]

- A. *For the care of a newborn*
- B. *For an adoption or foster care placement*
- C. *For an employee's own serious health condition (not including maternity-related reasons) [HYPERLINK “serious health condition”]*
- D. *For a pregnancy-related reason*
- E. *For the care of a child, spouse or parent with a serious health condition [HYPERLINK “serious health condition”]*

- F. For care of a parent or spouse who is elderly [[HYPERLINK “elderly”](#)]
- G. For the care of a military service member with a serious injury or illness [[HYPERLINK “care of a military service member”](#)]
- H. For reasons related to the deployment of a military service member [[HYPERLINK “deployment of a military service member”](#)]

Also, a new question was added to ask about coverage for caregivers:

Q16x_1 “Does this site’s leave policies for these types of leave cover guardians and caregivers of a child regardless of their legal or biological relationship to that child?”

Serious Health Condition

The 2000 Establishment Survey included an open-ended question at which respondents were asked to define “serious health condition” as it pertained to their leave policy. The responses collected were coded. Because open-ended questions can be cognitively burdensome for respondents and costly for data processing, the 2012 Worksite Survey did not include this open-ended question. The recent survey instead provided the FMLA definition “serious health condition” as needed.

Other Changes Relative to the 2000 Survey

The 2012 Worksite Survey added several new questions:

- The number of employees who have worked at the organization for at least one year,
- Method of reporting work time,
- Tracking unscheduled absences and family and medical leave,
- Processing FMLA requests,
- Additional questions on the administration of intermittent leave,
- Conditions for taking leave and returning to work including medical and fitness of duty certifications and re-certification,
- Employee misuse of leave,
- Additional questions on leave for care or deployment of military service members.

To maintain approximately constant respondent burden despite the addition of these new questions, the following questions that were included in the 2000 Establishment Survey were dropped from the 2012 Worksite Survey:

- The history of FMLA at the organization,
- Denial of leave,
- Compensating for costs associated with FMLA,
- Recovering benefits from employees who did not return after leave,
- Comparison of the organization’s own medical and family leave policies compared to FMLA,
- Respondent information.

See Appendix C for the complete list of changes between the 2000 and 2012 Worksite Survey.

2.2.3 Pilot Surveys

The objective of the pilot was to test comprehension, timing, and navigation of the 2012 Worksite Survey. As noted, an important change from the 2000 survey is the addition of a Web survey option. The Web survey option was included to help improve overall and item-specific response rates among employers.

Because we expected that some respondents would still prefer a telephone option and significant changes were made to the questionnaire, the pilot focused on execution of the full recruitment and data collection protocol, self-administration of the survey online, and interviewer administration of the survey over the phone.

The pilot began with the respondent screening interviews on April 15, 2011. Once nine respondents were identified, the main survey began by mailing a pre-notification letter with a Web link to the nine sample cases on April 28, 2011. This mailing of the pre-notification letter (with Web link) was followed by up to three email reminders to non-respondents for whom we had an email address, as well as phone follow-up, and a non-contact/refusal conversion letter. The pilot concluded on May 27, 2011, with four completed main interviews, three by phone and one by Web.

Once a survey was completed, project managers attempted to debrief each respondent to get their feedback on the survey experience. These debriefing sessions were open-ended conversations including questions about the relevance of the survey for the company, appropriateness for the chosen respondent comprehension, timing, coverage of topics, and functionality of the Web survey. As a result of the testing, we made several edits to the survey, namely to keep the survey within the proposed time limit.

The 2012 Worksite Survey underwent OMB clearance from July to December 2011. Upon final clearance in January 2012, a second pilot was conducted to test the fully-programmed survey as well as timing. Like the initial pilot, the second pilot began with respondent screening from January 26, 2012 through February 1, 2012, resulting in a total of 140 verified respondents. The main survey pilot began by mailing the re-notification letter with Web link to the verified sample cases on February 2, 2012. These initial attempts were again followed by up to two email reminders to non-respondents for whom we had an email address, as well as phone follow-up, and a non-contact/refusal conversion letter. The pilot concluded on February 29, 2012, with 38 completed main interviews, 12 by phone and 36 by Web. No significant substantive changes to the survey were required as a result of the second pilot.

2.3 Data Collection Procedures

In this section we discuss data collection procedures: Interviewer training, screening interview, and the main survey interview.

2.3.1 Interviewer Training

Senior project staff conducted intensive trainings for field supervisory staff and interviewers to prepare them for administration of the survey. The first training reviewed general interviewing principles and unique study procedures and requirements. It also allowed interviewers access to the CATI equipment, to gain familiarity with the questionnaire and to perform practice interviews. At the start of the training, the project directors explained the purpose and goals of the study. In telephone

surveys, the most critical issue is usually to ensure that the interviewer understands the questionnaire fully, and knows how to ask the questions properly and record the responses accurately. Abt SRBI project staff reviewed important considerations in the questionnaire, including probing, expected respondent questions, and ambiguity. We reviewed the questionnaire, the question-by-question specifications, and questions and problems that interviewers had concerning the questionnaire. Mock interviews were conducted and designed to mimic a variety of interview situations (smaller and larger worksites, covered and non-covered firms). Additional training was conducted to ensure that interviewers were comfortable helping respondents to access the Web version of the survey. Technical questions that interviewers could not answer were transferred immediately to a help-desk or project manager.

The data collection for the 2012 Worksite Survey was conducted by Abt SRBI and featured two main phases. The first phase was a screening interview used to determine eligibility and identify the key informant at the establishment. The second phase included a pre-notification mailing to direct the informant to the survey Website so that they could complete the survey instrument online and a telephone follow-up call to complete the interview via CATI. Each phase is described in detail below.

2.3.2 Screening Interview

In the first phase of the Worksite Survey, telephone interviewers contacted the sample establishments in order to check eligibility for the survey and to identify the key informant needed to respond on behalf of the establishment. In order for an establishment to be eligible, the screening interview needed to confirm that the place contacted was the establishment listed on the sampling frame, that the establishment was in the private sector, and that the establishment had at least one employee.

Given the detailed nature of the main questionnaire, which included some questions possibly requiring reference to company administrative records, it was necessary to identify the human resources director or the person responsible for the company's benefits plan. Interviewers asked for the name of this "key informant" during the screening interview. If the sampled establishment was a branch location, the key informant may have been located at a different location (e.g., company headquarters).

2.3.3 Main Survey Administration

After an establishment had completed the screening interview, the key informant was mailed a package of materials providing background about the project. The package included a letter from the Department of Labor explaining the importance of the survey and inviting the key informant to complete the survey either online (on a secure Web site) or by calling a toll-free number to complete the survey over the phone with an interviewer. The letter was printed on Department of Labor letterhead so that the recipient could clearly distinguish the survey materials from junk mail. The package was sent via priority mail. Email reminders were sent to key informants who had not responded to the Web survey.

All nonresponders to the package mailing and Web survey were contacted by interviewers who then attempted to complete the interview over the phone. The calling protocol for the CATI follow-up effort was 10 calls per phone number available for the informant.

2.4 Response Rates

The response rate for the Worksite Survey was computed in three steps. In the first step, a response rate was calculated for the screening interview. In the second step, a response rate was calculated for the extended interview. In the third step, the two response rates were combined to produce the overall survey response rate. These steps are described in detail below.

2.4.1 Response Rate for the Screener Interview

The final dispositions for the screener (or “verification”) stage are displayed in Exhibit 2.4.1. In total, 12,240 establishments were sampled from the DMI frame, and 8,229 of these completed the screening process. The set of sample units completing the screener (AAPOR code 1.100 in Exhibit 2.4.1 below) included those determined to be eligible for the survey (n=6,943) as well as those determined to be ineligible (n=1,286) (e.g., no longer in business, public institution).

Exhibit 2.4.1 Final Dispositions for the Screener and Extended Interviews

Disposition	AAPOR Code	Screener	Extended
Interview (Category 1)			
Complete	1.100	8,229	1,812
Eligible, non-interview (Category 2)			
Refusal	2.110	1,024	4,359
Break-off	2.120	28	6
Respondent never available	2.210	2	27
Telephone answering device	2.220	251	139
Other	2.300	2	394
Target Respondent physically or mentally unable/incompetent	2.320	16	11
Language barrier	2.330	24	16
Unknown eligibility, non-interview (Category 3)			
Always busy	3.120	49	
No answer	3.130	311	
No screener completed	3.210	1,082	
Not eligible (Category 4)			
Fax/data line	4.200	84	
Non-working/disconnect	4.300	412	
Nonworking number	4.310	391	
Temporarily out of service	4.330	335	
Government/Public organization	4.510		22
No eligible respondent (zero workers)	4.700		157
TOTAL		12,240	6,943

The screener response rates are reported in Exhibit 2.4.2. The first row reports the rates based on the AAPOR RR(3) formula as applied to establishment surveys. The second row reports the rates based on the formula used in the 2000 Establishment Survey Report. The latter formula yields a somewhat higher response rate because it assumes that only 10 percent of the non-locatable sample units are eligible, as compared to the 89 percent estimated as eligible (“e”) under the AAPOR RR(3) formula.

Exhibit 2.4.2 Weighted and Unweighted Screener, Extended Interview, and Overall Response Rates

Response Rate Formula	Unweighted			Weighted		
	Screener	Extended	Overall	Screener	Extended	Overall
AAPOR Response Rate 3* Formula: $I / ((I+P) + (R+NC+O) + e(UH+UO))$	75.8%	26.8%	20.3%	56.8%	36.8%	20.9%
2000 Establishment Response Rate Formula Formula: $C / (C+RB+LP+NA+NM+MC + (.10(NL+NW)))$	80.9%	26.8%	21.7%	61.7%	36.8%	22.7%

* The "e" coefficient in AAPOR RR(3) was computed as $(I+R+NC+O) / ((I+R+NC+O)+NE)$

The unweighted AAPOR screener RR(3) is 75.8 percent. The weighted screener response rate is also relevant for the Worksite Survey because the probability of selection varied across the sampling strata. The weighted screener RR(3) is 56.8 percent. The weighted response rates incorporate the screener base weights. The fact that the unweighted rate is higher than the weighted rate reflects the fact that smaller establishments (which were weighted up to account for their lower selection probabilities) were less likely to respond to the screener than larger establishments. This pattern was also documented in the 2000 Establishment Survey (albeit to a smaller degree).

2.4.2 Response Rate for the Extended Interview

The weighted and unweighted extended interview response rates are presented in Exhibit 2.4.2. The unweighted AAPOR RR(3) for the extended interview is 26.8 percent and the weighted RR(3) is 36.8 percent. This pattern reflects the fact that smaller establishments (which had larger weights) were more likely than larger establishments to complete the interview – conditional upon having completed the screener.

2.4.3 Overall Response Rate

The overall survey response rate is computed as the product of the screener and extended interview rates. The unweighted AAPOR RR(3) overall survey response rate is 20.3 percent, and the weighted response rate is 20.9 percent (Exhibit 2.4.2).

2.5 Analysis of Nonresponse of the 2012 Worksite Survey

The response rate achieved in the 2012 Worksite Survey was noticeably lower than that achieved in the 2000 survey (weighted AAPOR RR3 = 20.9 percent vs. 65.0 percent, respectively). One potential reason for the drop-off is that the 2012 sample for the Worksite Survey was released on a rolling basis, and the field period was noticeably compressed relative to 2000. Thus, the amount of time devoted to verifying the eligibility of each sampled establishment and obtaining contact information for the key informants was not as extensive as in 2000. In the 2000 Establishment Survey, eight weeks were devoted solely to the screening component and an additional 10 weeks were devoted to completing the extended interviews, for a total of 18 weeks of data collection. For the 2012 Worksite Survey, by contrast, the field period was limited to 14 weeks, which allowed less time for both the screening and the interviewing stages of data collection. If future Worksite Surveys are conducted, we recommend a longer field period so that more interviewing effort can be applied and the response rate increased.

A second factor that we strongly suspect led to a lower response rate in 2012 was deterioration in the quality of the sampling frame. The same frame, the Dun's Market Identifiers (DMI) file, was used for both the 2000 and 2012 Worksite Surveys. In 2012, however, the DMI appeared to contain substantially more out-of-business records, relative to 2000. The proportion of sample records with final disposition "Non-working/Disconnected" was 9.0 percent in 2012 as compared to just 1.3 percent in 2000. Similarly, the proportion of sample records with final disposition "Out of business" was 12.7 percent in 2012 as compared to just 1.6% in 2000. It seems likely that at least part of this deterioration in the DMI sampling frame reflects the economic recession that took place (starting in 2008) between the two surveys. Many establishments went out of business during this period. As a result, there was a significantly higher proportion of ineligible records on the DMI in 2012, which hampered the screening/verification effort relative to the 2000 survey.

Neither of these hypothesized mechanisms of nonresponse in 2012 (shorter field period or frame deterioration) would necessarily be expected to systematically bias estimates from the 2012 Worksite Survey. That said, the lower response rate obtained in the 2012 survey does suggest the potential for nonresponse bias to undermine survey estimates. The analysis below provides an empirical investigation of the potential risk posed by nonresponse bias. The approaches used to evaluate nonresponse in the Worksite Survey are a comparison of easier-to-reach versus harder-to-reach establishments and response propensity modeling.¹⁰

2.5.1 Comparison of Easier to Reach vs. Harder to Reach Establishments

In this analysis, establishments that were more difficult to interview in the Worksite Survey are compared to those that were easier to interview. The more difficult cases serve as proxies for the establishments that never completed the extended interview. If the harder to reach cases do not differ from the easier to reach ones, then presumably the sample members never reached also do not differ from those interviewed. If observed differences disappear after controlling for weighting variables, then that would suggest that the weighting protocol has minimized the risk of nonresponse bias with respect to the estimate at hand. As discussed above in the discussion of the Employee Survey nonresponse analyses, support for this “continuum of resistance” model is inconsistent, but it can still be a useful framework for assessing the relationship between level of effort and nonresponse bias.

In this analysis, the easy to reach versus hard to reach dimension was defined as the total number of the calls to the establishment to obtain the completed interview. Unlike in the Employee Survey, no refusal conversion was attempted with cases that refused to participate in the Worksite Survey. The number of call attempts made to Worksite Survey respondent ranged from zero to 21. The n=143 (7.9 percent) respondents with zero attempts represent instances where the extended interview was completed on the Web before the start of the CATI nonresponse follow-up phase. The mean number of attempts for the total responding sample was 4.68.

Exhibit 2.5.1 presents the mean number of call attempts for responding establishments grouped by size (number of employees), industry (NAICS code), FMLA coverage status, workforce gender ratio (% female employees), and whether any of their workforce was unionized. Two types of statistical testing were performed with these group means. First, bivariate tests were conducted (either *t*-tests or *F*-tests depending on the nature of the grouping variable) to test for variation between the groups in the number of attempts required to complete the interview. Due to the positive skew in the distribution of call attempts, the dependent variable used in all significance tests for this analysis was the natural log of 1 plus the number of attempts.¹¹

¹⁰ Preliminary plans for Worksite Survey nonresponse analysis included a comparison of weighted Worksite Survey estimates to external benchmarks. Upon further investigation, however, each of the three estimates available from the key benchmark study (the QCEW) was used in the weighting protocol. There were, thus, no other variables that could be used to support a benchmark comparison analysis.

¹¹ The log transformation reduced the skewness in the attempts distribution from 1.03 to -0.33, where zero represents no skewness.

Exhibit 2.5.1 Mean Number of Attempts by Establishment Characteristics

Characteristic	Mean Number of Call Attempts	Is the difference in means significant in bivariate analysis?	Is the difference in means significant when controlling for size and industry?
Establishment size †		Yes	n/a
9 or fewer employees	4.99		
10-249 employees	4.11		
250+ employees	6.19		
FMLA coverage status		Yes	No
FMLA covered	4.90		
Not FMLA covered	4.42		
Workforce unionization		No	No
Any employees unionized	4.99		
No employees unionized	4.63		
Percent female workforce		Yes	Yes
0%	4.63		
1% – 24.9%	3.58		
25% - 49.9%	4.72		
50% - 74.9%	5.19		
75% +	4.86		
Industry type†		Yes	n/a
Manufacturing	3.87		
Retail	4.67		
Services	5.23		
Other	4.78		

† Indicates that the variable is included in the raking ratio adjustment of the weighting.

The results from the bivariate tests are presented in the middle column of Exhibit 2.5.1. All of the variables except for workforce unionization are significant bivariate predictors. In general, more attempts were required for larger establishments (250+ employees), FMLA covered establishments, establishments with more female workforces, and establishments in the Service sector. While these patterns are informative about the nature of nonresponse in the 2012 Worksite Survey, they do not account for the fact that the survey estimates were weighted and the weighting was designed, in part, to reduce the risk of nonresponse bias. Specifically, the weighting protocol included raking ratio adjustment to Quarterly Census of Employment and Wages (QCEW) population controls for region and size x industry.

The key question with respect to the risk from nonresponse bias is whether or not the statistically significant bivariate patterns remain when controlling for the variables used in the weighting, particularly the cross-classification of size and industry. The far right column of Exhibit 2.5.1 reports the results of multivariate testing. The effect on number of attempts from the grouping variable on the left was tested in the presence of main effects and the interaction term for size and industry. This multivariate testing showed that the relationship between FMLA coverage and number of attempts disappears when controlling for size and industry. This result is not surprising given that FMLA coverage is, in part, a function of establishment size. The association between workforce gender distribution and the number of attempts does remain marginally significant ($p=0.048$) when

controlling for size and industry. This suggests that the weighted survey estimates may somewhat under-represent establishments with workforces that have a relatively high proportion of women. On balance, this level of effort analysis indicates that nonresponse did vary across key establishment subgroups, but there is evidence that the raking ratio adjustment to QCEW control totals likely reduce the potential for nonresponse bias in a number of the survey estimates.

2.5.2 Response Propensity Models for Contact and Cooperation

A different approach for evaluating the potential for nonresponse bias in the Worksite Survey is a response propensity analysis that identifies factors associated with survey response. Many establishment characteristics can influence response propensity. The response propensity model allows the researcher to identify the most powerful predictors of response when all available predictors are tested simultaneously.

In this analysis we consider two different outcomes: contact with the establishment and cooperation with the extended interview conditional upon contact. In the response propensity modeling for the Employee Survey, a model predicting contact was not estimated because essentially no useful information was available for the non-contacted cases. For the Worksite Survey, by contrast, several useful variables are known for both the contacted and non-contacted cases due to the richer sampling frame. These variables are establishment size (measured as number of employees), industry (NAICS code), and Census region.

Exhibit 2.5.2 presents the estimated generalized logistic regression parameters for the model predicting contact based on sampling frame variables for industry, size, and region. The interaction for size and industry was also included in the model based on information from the 2000 Establishment Survey that the relationship between size and response propensity varies across industry groups.¹² The sample (n=11,043) used to estimate the model includes all sample units released for the survey, excluding those determined to be ineligible on the basis that the phone number was disconnected or otherwise not working. Numbers with undetermined eligibility status (e.g., “ring no answer” and “busy all attempts”) are included in the model.

¹² During preliminary analysis, a fully-saturated model was tested but none of the interactions involving region were statistically significant. Consequently, only the main effect of region is included in the model.

Exhibit 2.5.2 Logistic Regression Estimating the Probability of Contact with a Sample Establishment in the Worksite Survey

Parameter	Estimate	s.e.	Wald χ^2	p value
Intercept	2.098***	0.088	566.09	<.0001
Size (Number of employees)	0.222***	0.025	81.98	<.0001
Industry = NAICS Group I	0.126	0.167	0.57	0.451
Industry = NAICS Group II	0.480**	0.169	8.02	<.01
Industry = NAICS Group III	-0.200	0.122	2.70	0.100
Size x Industry = NAICS Group I	-0.035	0.045	0.62	0.429
Size x Industry = NAICS Group II	-0.112*	0.046	6.03	0.014
Size x Industry = NAICS Group III	-0.052	0.036	2.11	0.146
Region = Northeast	-0.018	0.080	0.05	0.821
Region = Midwest	0.219**	0.080	7.53	<.01
Region = South	-0.011	0.065	0.03	0.864

Model Diagnostics:

Area under ROC curve (c) 0.657

-2 Log Likelihood 4,640.7

Sample size 11,043

Source: Worksite Survey. Figures are unweighted.

Reference groups for categorical variables: industry (NAICS group IV), region (West)

The estimated model shows that establishment size had a strong, positive association with probability of contact. This pattern was most noticeable at the lower end of the size spectrum. The contact rate was 87 percent for establishments with fewer than 10 employees, but it was over 95 percent for all other size categories. This result may reflect the fact that small establishments are more likely to have gone out of business than larger establishments, but this could not be determined with certainty during the field period because no one was reached at some of the numbers. This post hoc hypothesis is supported by the fact that the ineligibility rate was by far the highest for this small size group. Some 18 percent of establishments in the smallest size group were determined to be ineligible, versus less than 10 percent of establishments with 10 or more employees.

The model also shows several statistically significant effects associate with industry and region. From a practical standpoint, however, the consequence of these other results does not appear to be very great. The contact rates observed across the industry groups varied within a fairly narrow range – from 91.6 percent (Group III) to 95.6 percent (Group IV) (see Exhibit 2.1.1 for the NAICS codes used for each of the four industry groupings). Likewise, the contact rates observed across the Census region range from 93.1 percent (West) to 95.5 (Midwest). Due to the large model sample size (n=11,043) these effects are statistically significant but the modest differences in contact rates across these groups suggests minimal consequences for data quality. It is also important to bear in mind that the Worksite Survey protocol included a nonresponse adjustment aligning the responding sample to the stratum totals for size x industry group. In additional, the weighting featured a raking ratio adjustment to population controls for Census region and size x industry based on the QCEW. These adjustments would be expected to minimize nonresponse error associated with the significant relationships detected by the model.

A second model estimated for the 2012 Worksite Survey predicts cooperation to the main interview conditional upon eligibility and completion of the screener. Unlike in the Employee Survey, the screener for the Worksite Survey collected little in the way of new information about the sample units. The screener included one question about whether the establishment maintains records of

employee leave and another question about whether FMLA requests are processed by a third party, but the item nonresponse rate for both items was over 40 percent. Presumably, many of the respondents to the screener component did not know the answer to these questions.¹³ In light of these very high item nonresponse rates, these two items were excluded from the cooperation propensity model. The cooperation model, thus, included the same set of predictors as the contact model: size, industry, Census region, and the interaction of size and industry. The estimated cooperation model parameters are presented in Exhibit 2.5.3.

Exhibit 2.5.3 Logistic Regression Estimating the Probability of Cooperation with the Worksite Survey Conditional on Eligibility and Completion of the Screener

Parameter	Estimate	s.e.	Wald χ^2	p value
Intercept	2.151***	0.136	250.26	<.0001
Size (Number of employees)	-0.229***	0.029	63.29	<.0001
Industry = NAICS Group I	-0.006	0.261	0.00	0.981
Industry = NAICS Group II	0.507*	0.256	3.92	0.048
Industry = NAICS Group III	-0.205	0.212	0.94	0.332
Size x Industry = NAICS Group I	0.032	0.054	0.35	0.554
Size x Industry = NAICS Group II	-0.138*	0.055	6.22	0.013
Size x Industry = NAICS Group III	0.021	0.046	0.21	0.644
Region = Northeast	0.050	0.097	0.27	0.602
Region = Midwest	-0.018	0.087	0.04	0.836
Region = South	-0.057	0.077	0.55	0.459

Model Diagnostics:

Area under ROC curve (c) 0.618
 -2 Log Likelihood 2,484.3
 Sample size 2,360

Source: Worksite Survey. Figures are unweighted.

Reference groups for categorical variables: industry (NAICS group IV), region (West)

The cooperation model shows a strong negative association between establishment size and propensity to cooperate with the Worksite Survey main interview, conditional upon screener completions. Interestingly, the direction of this effect is opposite of that documented for the propensity to be contacted shown in Exhibit 2.5.2. Small establishments were significantly less likely to be contacted for the survey but, conditional upon screener completion, they were significantly more likely to cooperate with the main interview, relative to larger establishments. One post hoc explanation for this pattern is that the reporting burden may have been lower for smaller establishments than larger ones. Small establishments may have fewer benefits and policies to report, and they may have only one worksite, which would result in a faster and less challenging interview relative to interviews completed by establishments with complex leave policies and/or multiple worksites.

Neither industry nor Census region has a significant association with cooperation propensity in bivariate analysis. That pattern is largely borne out in Exhibit 2.5.3. There are a few statistically significant terms associated with industry, but the cooperation rate varied by only 2.6 percentage points across the four industry groups (from a high of 78.2 in Group I to a low of 75.6 in Group III).

¹³ One purpose of the screener was to identify a key informant who was often different from the screener respondent.

On balance then, the only effect of consequence in Exhibit 2.5.3 is arguably that associated with establishment size.

As discussed above, differential nonresponse across size groups was accounted for in the weighting in a within-stratum nonresponse adjustment as well as raking ratio estimation. Given this, there is no information gleaned in the contact or cooperation propensity modeling to suggest that Worksite Survey estimates are compromised by residual nonresponse bias. While this piece of analysis yielded no evidence of nonresponse error, we would note that the models were quite limited in scope and, therefore, by no means confirm that nonresponse is of no concern. This response propensity analysis would have been more informative if there had been more relevant information available for the nonresponding and responding establishments sampled for the Worksite Survey. If the contact and cooperation models included more variables related to the survey outcomes, the models would have provided a more robust evaluation of the potential risk posed by nonresponse. This lack of relevant information from sampling frames and other sources is generally what motivates survey data collection in the first place.

2.5.3 Summary of Nonresponse Analysis for the Worksite Survey

The analysis of the Worksite Survey found only fairly limited evidence that nonresponse poses a threat to the survey estimates. In the level of effort analysis, two of the three significant bivariate associations between level of effort and survey outcomes disappeared when the survey weighting variables were accounted for in the test. The association between the workforce gender distribution and the number of attempts did, however, remain significant ($p=0.048$). This suggests that the weighted survey estimates may have somewhat under-represented establishments with workforces that have a relatively high proportion of women. As discussed above, however, this result is not definitive in proving the existence of nonresponse bias because the “continuum of resistance” assumption underlying the level of effort analysis is not necessarily correct.

The primary finding in the response propensity analysis is that establishment size is negatively associated with contact in the Worksite Survey but positively associated with cooperation (conditional on screener completion). While interesting, this finding does not necessarily represent a threat from nonresponse bias because differential nonresponse across size groups was accounted for in the weighting in a within-stratum nonresponse adjustment as well as raking ration estimation.

2.6 Weighting

The weights for the 2012 Worksite Survey were designed to adjust for several key factors: differential probabilities of selection across sampling strata, differential nonresponse across sampling strata, and residual nonresponse addressed through post-stratification.

2.6.1 Base Weight for Probability of Selection

The survey employed a stratified simple random design, and so the base weight for all of the sample units in a given stratum was equal to the stratum population size on the DMI sampling frame divided by the number of sample units in the stratum. The values of the base weights are presented in Exhibit 2.6.1.

2.6.2 Weighting Class Nonresponse Adjustment

The second weighting stage featured a weighting class nonresponse adjustment. The weighting classes were defined as the 28 sampling strata. The adjustment was computed as the ratio of the weighted sum of all eligible establishments in the stratum (responders and nonresponders) divided by weighted sum of establishments that responded to the survey. The values of the weighting class adjustments are presented in Exhibit 2.6.1.

Exhibit 2.6.1 Base Weight and Weighting Class Nonresponse Adjustment Values, by Stratum

Sampling Stratum	Base Weight	Nonresponse Adjustment
Size 1-9 NAICS Group I	6,230.05147	5.02564
Size 1-9 NAICS Group II	6,194.21767	3.54839
Size 1-9 NAICS Group III	6,243.13753	4.82292
Size 1-9 NAICS Group IV	6,253.39446	4.12329
Size 10-19 NAICS Group I	633.77311	4.75000
Size 10-19 NAICS Group II	622.04891	4.55172
Size 10-19 NAICS Group III	624.19155	4.30190
Size 10-19 NAICS Group IV	622.59579	3.90123
Size 20-49 NAICS Group I	278.76842	4.40909
Size 20-49 NAICS Group II	281.89749	4.64474
Size 20-49 NAICS Group III	285.04684	3.63513
Size 20-49 NAICS Group IV	282.68170	4.31035
Size 50-99 NAICS Group I	88.36725	4.07407
Size 50-99 NAICS Group II	88.87248	4.83098
Size 50-99 NAICS Group III	88.85577	4.04348
Size 50-99 NAICS Group IV	88.57568	4.05645
Size 100-249 NAICS Group I	62.02134	5.09805
Size 100-249 NAICS Group II	61.48571	5.04761
Size 100-249 NAICS Group III	61.92419	4.68181
Size 100-249 NAICS Group IV	61.67431	4.26760
Size 250-999 NAICS Group I	19.36486	5.15517
Size 250-999 NAICS Group II	19.66667	5.12727
Size 250-999 NAICS Group III	19.54110	4.82926
Size 250-999 NAICS Group IV	19.57215	4.03226
Size 1,000+ NAICS Group I	3.06393	4.79412
Size 1,000+ NAICS Group II	3.11494	7.42856
Size 1,000+ NAICS Group III	3.11189	4.70731
Size 1,000+ NAICS Group IV	3.08056	3.97753

2.6.3 Post-stratification Ratio Adjustment

To help reduce possible under-coverage errors in the DMI sampling frame and reduce possible residual nonresponse bias, the weights were then post-stratified to reflect the most recent population information available from the 2011 Quarterly Census of Employment and Wages (QCEW). Specifically, the weights were adjusted to population control totals for region and the cross-classification of size (three categories) by industry (four categories). The post-stratification ratio adjustment forced the sum of the weights to agree with the QCEW totals for each of the post strata.

2.7 Variance Estimation

To account for the complex design of the 2012 Worksite Survey, a repeated sampling technique (specifically jackknife delete two repeated replication (Krewski and Rao 1981; Lu, Brick & Sitter 2006)), was used to create replicate weights for this study. The subsamples (replicates) were created using the same sample design, but deleting a portion of the sample, and then weighting each subsample up to the population total. A total of 80 replicates were created by combining telephone numbers to reduce the computational effort. The replicate weights are given by variables RPL01-RPL80. They can be used in SAS with the REPWEIGHTS statement.

Statistical significance tests presented in the report were computed using appropriate complex survey software and procedures.

2.8 Producing Employee Level Estimates from the Worksite Survey

If the sample of worksites is representative of the population that provides employment, then the data on employees can be used to draw inferences on the population of the employees (or at least on the part of this population employed at the target establishments). Thus, in addition to constructing the base and replicate weights that can be used to provide inference for the population of employers, we also developed a methodology to provide employee level estimates from the Worksite Survey described below.

For a worksite i , let e_i be the number of employees, w_i be the sampling weight of the worksite (represented by variable WEIGHT in the deliverable data set), z_i be the worksite-level characteristic of interest (e.g., the number of unionized employees), y_i be the employee-level characteristic of interest (e.g., % of unionized employees). In this example, $z_i = y_i e_i$. The population percentage of unionized employees is then

$$\theta = \frac{\sum_{i \in U} z_i}{\sum_{i \in U} e_i} = \frac{T[z]}{T[e]} \quad (1)$$

$$= \frac{\sum_{i \in U} e_i y_i}{\sum_{i \in U} e_i} = \frac{T[ey]}{T[e]} \quad (2)$$

where U is the population (universe), $T[z]$ is the total of the variable z , etc. This population percentage is estimated with

$$\hat{\theta} = \frac{\sum_{i \in S} w_i z_i}{\sum_{i \in S} w_i e_i} = \frac{t[z]}{t[e]} \quad (3)$$

$$= \frac{\sum_{i \in S} w_i e_i y_i}{\sum_{i \in S} w_i e_i} = \frac{t[ey]}{t[e]} \quad (4)$$

where $t[z]$ is the *estimate* of the total of the variable z , S is the sample, etc. Depending on how the question was asked in the Worksite Survey instrument, and how the data may be presented, the estimate of interest may have the form of (3) or (4). Either way, $\hat{\theta}$ is an estimator of a ratio (Lohr 2009, Sec. 4.1; Korn & Graubard 1999, Sec. 2.4). The linearization estimator of the sampling variance of $\hat{\theta}$ is given by

$$v[\hat{\theta}] \approx \frac{1}{(t[e])^2} v[z_i - \hat{\theta}e_i] \quad (5)$$

where $v[\cdot]$ is an appropriate estimator of the sampling variance of the quantity in the brackets (e.g., a jackknife variance estimator). Alternatively, the delta method for the total estimates $t[z]$, $t[e]$ can be used to obtain

$$v[\hat{\theta}] \approx \frac{1}{(t[e])^2} (v\{t[z]\} + \hat{\theta}^2 v\{t[e]\} - 2\hat{\theta} \text{cov}\{t[z], t[e]\}) \quad (6)$$

(formula (2.4.7) in Korn & Graubard 1999). Computing the point estimates, variances and the design effects for $\hat{\theta}$ is available in complex survey software using ratios (e.g., PROC SURVEYMEANS with RATIO statement in SAS, svyratio() function in R and computed statistic in WesVar).

If the data are available only as a y -type (individual level, per-employee basis) rather than a z -type (worksite-level, per-worksite basis) variable, they need to be scaled up to the worksite level, i.e., a worksite-level variable (the total # of unionized employees in the worksite) needs to be created for the analysis.

When replication variance estimation methods, such as the jackknife, BRR, or the bootstrap, are used with the survey data, a different computational shortcut can be taken with the individual level y -type data. As is easily seen from (4), the estimate $\hat{\theta}$ can be thought of as a weighted mean of y_i with the weights given by the doubly expanded weight product $w_i e_i$. When the replicate values of $\hat{\theta}^{(r)}$ necessary for variance estimation are being computed, the main weight w_i is being replaced by the r -th replicate weight $w_i^{(r)}$. Yet the expression (4) can still be interpreted as the weighted mean, now with the weight given by the product $w_i^{(r)} e_i$. Hence, both the main sampling weight and the replicate weights can be multiplied by the number of employees, and the statistic of interest and its standard error can be computed as the weighted mean of y_i with the doubly-expanded weight, rather as the ratio $t[ey]/t[e]$. Again, this is easily implemented with the complex survey-aware software (svy: mean in Stata; PROC SURVEYMEANS in SAS; svymean in R; mean in WesVar).

The drawback of this computational shortcut is that it gives a wrong DEFF. In terms of the equivalent variance expression (5), the variance that needs to be computed is $v[\hat{\theta}] \approx \frac{1}{(t[e])^2} v[y_i e_i - \hat{\theta}e_i]$. The shortcut procedure essentially factors out the employment e_i from the above procedure. However, when the variance under SRS (i.e., the denominator of the design effect) is being computed, the weights are ignored as not applicable under the SRS, leading to an incorrect expression $\frac{1}{(t[e])^2} v[y_i - \hat{\theta}]$. Another way to explain the problem is to note that using the doubly expanded weight $w_i e_i$ implies a sampling design in which one observation on an employee is taken from each establishment. However, this is an incorrect design specification; a data collection in which a worksite-level variable (number of unionized employees) is being observed should instead be interpreted as a one-stage cluster design in which all employees of the worksite i are observed, and their characteristic of interest y_i (such as union membership) is measured.

References

American Association for Public Opinion Research. 2010. AAPOR Cell Phone Task Force Report: New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers.

American Association for Public Opinion Research (AAPOR). 2010. "New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers," Prepared for AAPOR Council by the Cell Phone Task Force operating under the auspices of the AAPOR Standards Committee.

http://aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force&Template=/CM/ContentDisplay.cfm&ContentID=2818

Bethlehem, J.G. 2002. "Weighting Nonresponse Adjustments Based on Auxiliary Information." In *Survey Nonresponse* (eds R. Groves, D. Dillman, J. Eltinge and R. Little), pp. 275–287. New York: Wiley.

Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December, 2011. National Center for Health Statistics. July 2012

Brick, J.M., Cervantes, Flores, I., Lee, S. and Norman, G. 2011. "Nonsampling Errors in Dual Frame Telephone Surveys." *Survey Methodology* 37:1-12.

Groves, Robert M., *Survey Errors and Survey Costs*. New York: John Wiley & Sons, 1989.

Groves, R.M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 646-675.

Groves, R. and Couper, M. 1998. *Nonresponse in Household Interview Surveys*. Wiley: New York.

Groves, R.M. and Peytcheva, E. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis." *Public Opinion Quarterly* 72: 167-189.

Groves, R.M., Singer, E., and Corning, A. 2000. "Leverage-salience Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64:299–308.

Frankel, Martin R., Battaglia, Michael P., Link, Michael, and Mokdad, Ali H. 2007. "Intergrating Cell Phone Numbers into Random Digit Dial (RDD) Landline Surveys." *ASA Proceedings of the Social Statistics Section*, 3793-3800.

Keeter, S., Miller, C., Kohut, A., Groves, R., Presser, S. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64:125–48.

Keeter, S., Kennedy, C., Dimock, M., Best, J., Craighill, P. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70: 759-779.

Kennedy, C. Everett, S.E., Traugott, M.W. 2009. "Mechanisms of Nonresponse in Cell Phone Surveys." Presented at the Annual Meeting of the American Association for Public Opinion Research.

- Korn, E. L., and Graubar, B. I. (1999). *Analysis of Health Surveys*. John Wiley & Sons, New York , NY.
- Krewski, D. and J.N.K. Rao. 1981. "Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods," *Annals of Statistics* 9: 1010-1019.
- Lavrakas, P.J., Tompson, T.N., Benford, R., and Fleury, C. 2009. "More Research on a Hybrid Respondent Selection Method." Paper presented at the 34th annual conference of the Midwest Association for Public Opinion Research, Chicago.
- Lin, I. and N. Shaeffer. 1995. "Using Survey Participants to Estimate the Impact of Nonparticipation." *Public Opinion Quarterly* 59: 236-258.
- Little, R.J. 1986. "Survey Nonresponse Adjustments." *International Statistical Review* 54: 139-157.
- Lohr, S. (2009) *Sampling: Design and Analysis*, 2nd edition. Duxbury Press, Pacific Grove, CA.
- Lu, W. W., J. M. Brick, and R. R. Sitter (2006). Algorithms for constructing combined strata variance estimators. *Journal of the American Statistical Association* 101 (476): 1680-1692
- Merkle D. and Murray, E. 2002. "Nonresponse in Exit Polls: A Comprehensive Analysis." In Groves Robert M., Dillman Don A., Eltinge John L., Little Roderick J. A. (Eds.). *Survey Nonresponse* (Wiley, New York) pp. 243–58.
- 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. Japac, P. Lavrakas, M. Link, R. Sangster. New York, NY: John Wiley & Sons, Inc.
- Olson, K. 2006. "Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias." *Public Opinion Quarterly* 70: 737-758.
- 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.
- Potthoff, R., Manton, K., and Woodbury, M. (1993), "Correcting for Nonavailability Bias in Surveys by Weighting Based on Number of Callbacks," *Journal of American Statistical Association* 88:1197-207.
- Putnam, R.D. 1995. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.
- Singer, E. 2003. "Exploring the Meaning of Consent: Participation in Research and Beliefs about Risks and Benefits." *Journal of Official Statistics* 19: 273-285.
- Tourangeau, R. 2004. "Survey Research and Societal Change." *Annual Review of Psychology*, 55: 775-801.
- Zimowski, M., T. Tourangeau, and R. Ghadialy. 1997. "Nonresponse in Household Travel Surveys (FHWA-PL-98-029)," prepared for the Federal Highway Administration.