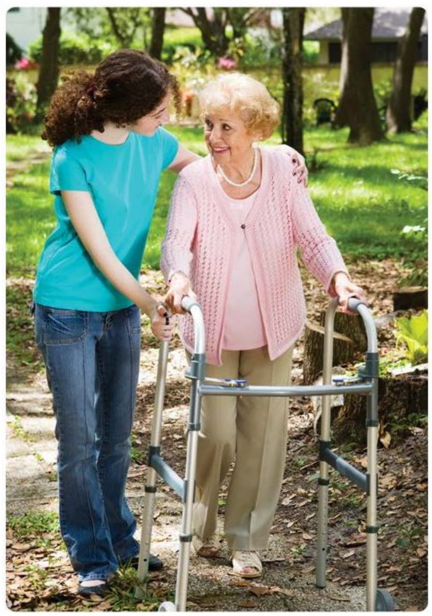
Employee and Worksite Perspectives of the Family and Medical Leave Act:

Methodology Report for the 2018 Surveys

July 2020


Submitted to:

Kuang-chi Chang
U.S. Department of Labor
The Chief Evaluation Office
200 Constitution Avenue, NW
Washington, DC 20210

Submitted by:

Abt Associates
6130 Executive Boulevard
Rockville, MD 20852

Authors:

Julie Pacer
Michelle Kahmann
Stan Hsieh
Stas Kolenikov
Michael Witt
Marci Schalk
Radha Roy

This report was prepared for the U.S. Department of Labor (DOL), Chief Evaluation Office by Abt Associates, under contract number DOLQ129633231/DOL-OPS-16-U-00141. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. government.

Abstract

In 2018, Abt Associates surveyed worksites and employees about experiences with family and medical leave. The 2018 Worksite Survey, a sequential, multi-mode (telephone and web) survey of U.S. firms, includes sites that are covered by the Family and Medical Leave Act (FMLA) and those that are not covered. The 2018 Employee Survey, a multi-mode (telephone and web) survey of the general population of working-age U.S. adults, includes employees who took leave, those who had an unmet need for leave, those with both met and unmet needs for leave, and those with neither. Some of the included employees are eligible for the FMLA and some are not. The 2018 surveys update similar surveys conducted in 1995, 2000, and 2012. This document presents the methods used to design the sampling plan, collect the data, and analyze the results for the Worksite Survey and the Employee Survey. In addition to the Methodology Report, there are a Survey Results Report, Methodology Report Appendices, a Supplemental Results Appendix, and a Public Use File Documentation volume.

BOLD THINKERS DRIVING REAL-WORLD IMPACT

Abt Associates | 6130 Executive Boulevard | Rockville, MD 20852

CONTENTS

Introduction **1**

1. Employee Survey **2**
1.1. Target Population and Sampling Design 2
1.2. Questionnaire Development 7
1.3. Data Collection Procedures 13
1.4. Response Rates 16
1.5. Analysis of Non-Response of the Employee Survey 19
1.6. Weighting 43

2. Worksite Survey **52**
2.1. Target Population and Sampling Design 52
2.2. Questionnaire Development 56
2.3. Data Collection Procedures 62
2.4. Response Rates 66
2.5. Analysis of Non-Response of the Worksite Survey 68
2.6. Weighting 77
2.7. Variance Estimation 80
2.8. Producing Employee-Level Estimates from the Worksite Survey 81

References **83**

List of Exhibits

Exhibit 1.1. Structure of the 2018 Employee Survey ... 8
Exhibit 1.2. Telephone dispositions for the 2018 Employee Survey, by sample ... 16
Exhibit 1.3. AAPOR telephone response rates for screener interview, extended interview, and overall, by sample ... 17
Exhibit 1.4. Legacy telephone response rates for screener interview, extended interview, and overall, by sample ... 17
Exhibit 1.5. AAPOR web survey response rates ... 18
Exhibit 1.6. Unweighted distribution of Employee Survey and NRFU samples, by demographics ... 23
Exhibit 1.7. Comparing Employee Survey and NRFU respondents across several key questionnaire items: unadjusted and regression-adjusted means ... 24
Exhibit 1.8. Regression-adjusted percentages for several key questionnaire items, by the interaction of Employee Survey/NRFU and mode of data collection (telephone/web) ... 28
Exhibit 1.9. Leave-related characteristics of telephone respondents, by level of effort groups ... 31
Exhibit 1.10. Leave-related characteristics of web respondents, by level of effort groups ... 32
Exhibit 1.11. Logistic regression estimates of RDD response propensities, by demographic characteristics, for landline and cell frames ... 33
Exhibit 1.12. Graphical representations of RDD response propensities by age, education, and gender, for landline and cell frames ... 34
Exhibit 1.13. Summary statistics of KnowledgePanel base weights ... 36
Exhibit 1.14. Distribution of Employee Survey and CPS samples, by demographics ... 38
Exhibit 1.15. Union participation (CPS and Employee Survey) ... 40
Exhibit 1.16. Employer size (CPS and Employee Survey) ... 41
Exhibit 1.17. Hours worked (CPS and FMLA Employee Survey) ... 41
Exhibit 1.18. Population control totals ... 43
Exhibit 1.19. Population control totals ... 44
Exhibit 1.20. Base weights, by phone stratum ... 44
Exhibit 1.21. Weight adjustment for working/residential numbers, by phone stratum ... 45
Exhibit 1.22. Cases affected by adjustments for multiple phones ... 45
Exhibit 1.23. Statistics for within-household respondent selection ... 46
Exhibit 1.24. Summary statistics: untrimmed and trimmed weights ... 47
Exhibit 1.25. Weighting summary post integration ... 49
Exhibit 1.26. Weighting summary post population calibration ... 49
Exhibit 2.1. NAICS groupings for Worksite Survey ... 53

Exhibit 2.2. Detailed sampling strata for the Worksite Survey 53
Exhibit 2.3. Worksite Survey sampling stratification by survey wave 56
Exhibit 2.4. Structure of the Wave 4 Worksite Survey 57
Exhibit 2.5. Response option experiments in the Worksite Survey 59
Exhibit 2.6. Contact protocol for the 2018 Worksite Survey 65
Exhibit 2.7. Final dispositions for the screener and extended (main) interview 67
Exhibit 2.8. Unweighted and weighted screener, extended interview, and overall response rates 67
Exhibit 2.9. Mean number of attempts by worksite characteristics 69
Exhibit 2.10. Results from the screening contact propensity logistic model 72
Exhibit 2.11. Results from the interview cooperation response propensity logistic model ... 75

Acknowledgements

The report is a collaboration of the United States Department of Labor's Chief Evaluation Office (CEO), Wage and Hour Division (WHD), and Abt Associates. The authors gratefully acknowledge the efforts of many individuals who assisted in the completion of this report. Christina Yancey and Kuang-chi Chang from CEO provided invaluable guidance throughout the study. In addition, Samuel Rowe from CEO provided insightful feedback on drafts of the report. The team thanks Technical Working Group member Jolene Smyth for her thoughtful review and feedback. The authors acknowledge Scott Keeter's careful review of selected sections of the report.

At Abt Associates, Jacob Klerman and Glen Schneider provided valuable feedback and guidance throughout the study. Puneet Kaur, Jessica Kerbo, and Charmayne Walker provided production and graphic design support. The team also is appreciative of the role of Bry Pollack in editing the report.

The Worksite Survey was fielded under the leadership of Julie Pacer. Michelle Kahmann and Stan Hsieh oversaw the Employee Survey. Ricki Jarmon served as Survey Quality Advisor for data collection. Stas Kolenikov led the technical design and weighting effort. Marci Schalk and Michael Witt also contributed to the technical design and related analyses.

Introduction

This Methodology Report summarizes the methods used by Abt Associates in conducting the 2018 (Wave 4) Family and Medical Leave Act (FMLA) Employee and Worksite Surveys for the United States Department of Labor (DOL).

Employee Survey

Chapter 1 of this report describes the methods for the **2018 Employee Survey**. The survey started as an overlapping, dual frame landline and cell phone random-digit dial (RDD) computer-assisted telephone interviewing (CATI) survey. Initial planning accounted for some but not all of the general decline seen in RDD productivity and response rates and the growth in spam flagging and call blocking technologies (AAPOR, 2019). To address the challenges of low RDD response rates in the first months of data collection, Abt Associates made several refinements to the data collection protocols. After these changes were made, response rates remained below established targets. In consultation with DOL, the study team transitioned the data collection design to a hybrid approach, adding a nationally representative probability-based web panel (Ipsos KnowledgePanel) to supplement the RDD efforts. The final Employee Survey was therefore a multi-mode telephone and web survey.

The Employee Survey featured three sections:

- ***Screener***—identified the target population of U.S. adults aged 18 or older who were employed for pay in the past 12 months (excluding self-employed) and those who needed and/or took family and medical leave. It also identified respondents for the extended interview.
- ***Basic Interview***—answered by respondents who *did not* need or take family and medical leave in the past 12 months.
- ***Extended Interview***—answered by respondents who *did* need or take family and medical leave in the past 12 months. The extended interview was roughly twice the length of the basic interview.

Worksite Survey

Chapter 2 of this report describes the methods for the **2018 Worksite Survey**. The survey was a sequential multi-mode (CATI and web) survey of U.S. firms. The study was conducted to obtain estimates of the use of leave under FMLA and examine the perceived impact on U.S. private business establishments. The sampling frame was drawn from Dun & Bradstreet's Dun's Market Identifiers (DMI) file. The final sample excluded the self-employed without employees, as well as government and quasi-government units (federal, state, and local governments; public educational institutions; and post offices).

Appendices

The separate Methodology Report Appendices volume contains the following:

Appendix A contains the Employee Survey materials, including the phone and web questionnaires, refusal conversion letters, and screen shots of the web survey.

Appendix B contains the Worksite Survey materials, including the informational packet cover letter and project information sheet, the survey questionnaire, and screen shots of the web survey.

Appendix C contains the revision matrices for both the Employee and the Worksite Surveys.

Appendix D contains the Employee Non-response Follow-Up Survey materials, including the invitation letter and questionnaire.

Appendix E contains the Employer Survey Response Option Experiment, Detailed Findings.

1. Employee Survey

This chapter presents the methods employed to design and administer the 2018 FMLA Employee Survey: the survey's target population and sampling design (Section 1.1); development of the survey instrument (Section 1.2); data collection procedures (Section 1.3); response rate calculations (Section 1.4); analysis of non-response (Section 1.5); weighting (Section 1.6); and variance estimation (Section 1.7).

1.1. Target Population and Sampling Design

The Employee Survey sampled 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, as they are not subject to FMLA. Initially, the survey used an overlapping, dual frame random-digit dial (RDD) design, with national samples from the landline and cell phone RDD frames. We added a web survey midway through data collection after low response rates made the RDD alone not viable to meet the required number of completed surveys in the allotted time frame. We estimate the coverage rate provided by this design to be at least 97 percent, based on the most recent estimate of 3.1 percent of phoneless population from the National Health Interview Survey (Blumberg & Luke, 2015). Interviews were conducted in English and Spanish.

Against a target total of 4,000 completed surveys, we collected 4,470 completed extended interviews, including 189 from the landline sample, 550 from the cell phone sample, and 3,731 from the web panel. The data collection was conducted by Abt Associates from March 6, 2018, through February 24, 2019.

The phone samples were provided by Survey Sampling International, LLC according to Abt's 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. Both frames were stratified so as to oversample the states of California, New Jersey, and Rhode Island, which had enacted paid leave legislation by the time the study started. The cell phone frame was additionally stratified to oversample prepaid phones in order to reach lower-income and racial/ethnic minority populations.

The web sample was selected from the Ipsos KnowledgePanel. The details of the panel are discussed later in Section 1.1.3. We selected a stratified sample using the panel profile data, using the same selection criteria as the RDD sample. Unlike the RDD sample (for which cell phone number prefixes and prepaid type are inaccurate proxies of geography and income, respectively), web panel profile information is considered accurate, and strata sample sizes were specified in advance and monitored in data collection.

1.1.1 Screening for the Target Population (Landline)

Screening

The landline sampling design necessitated screening for members of the target population. In the survey screener, interviewers determined whether the household contained at least one person aged 18 or older who had been employed (excluding self-employed) during the past 12 months. Then for all persons in the household meeting these criteria, the interviewer attempted to determine if they had taken, were taking (at the time of the interview), or needed without taking family and medical leave during the reference period.

The screener asked the household informant (aged 18 or older) to report the following information for each adult in the household:

- age;
- gender;
- education;

- worked for pay or profit in the past 12 months (yes/no);
- number of jobs (more than one job, one job, or not working in the past 12 months);
- sector (government, private company, non-profit, self-employed);
- took leave from work in the past 12 months (yes/no); and
- needed but did not take leave from work in the past 12 months (yes/no).

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

Each eligible adult was classified during the screening into one of three family and medical leave groups: ***leave needer*** (defined as an employee who needed to take family and medical leave for a covered reason but did not; that is, "unmet need for leave"); ***leave taker*** (an employee who took family and medical leave); or ***employed only*** (an employee who did not need or take family and medical leave).[1] This classification informed our selection of the within-household respondent.

Within-household selection

The purpose of the within-household selection procedure was twofold: (1) we used the procedure to identify one randomly selected eligible adult per household for the extended interview; (2) we also used it to increase sample sizes for key oversampled subgroups (i.e., the leave needers and the leave takers). For the latter, we assigned each eligible adult in the household a non-zero probability of selection for the extended interview, assigning the leave needers and leave takers higher probabilities of selection than the employed only. It is important to note that this oversampling procedure provides full coverage for the target population, and that the survey weights (described in Section 1.6) adjust for the differential probabilities of selection. In the weighted survey estimates, each of these subgroups is represented in proportion to its actual size.

To accomplish these two objectives, within-household respondent selection was conducted in three stages.

Stage 1. First we took inventory of which family and medical leave *subgroups* were present in that household (i.e., is there at least one leave needer? at least one leave taker? at least one employed only?). We used this information to determine from which family and medical leave *subgroup* the extended interview respondent should be selected. For households in which all eligible adults were classified as belonging to the same subgroup (e.g., employed only), that subgroup was automatically selected. For households in which multiple family and medical leave subgroups were represented, we selected the leave needer and leave taker subgroups at a higher rate than the employed only subgroup because the former populations' incidence rates are significantly lower. The selection rules and rates applied in Stage 1 were as follows:

- If all eligible adults in the household are of the *same* family and medical leave subgroup (i.e., leave taker, leave needer, or employed only), then select that subgroup. If all household member(s) are employed only, then go to Stage 2.

[1] Employees who were reported as both needing and taking leave were classified as leave needers *only for the purpose of the within-household 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, then select the leave needer subgroup with 90 percent probability and the leave taker subgroup with 10 percent probability. Skip to Stage 3.
- If the household has at least one leave needer and at least one employed only, then select the leave needer subgroup with 90 percent probability and the employed only subgroup with 10 percent probability. Skip to Stage 3.
- If the household has at least one leave taker and at least one employed only, then select the leave taker subgroup with 90 percent probability and the employed only subgroup 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, then select the leave needer subgroup with 80 percent probability, the leave taker subgroup with 10 percent probability, and the employed only subgroup with 10 percent probability. Skip to Stage 3.

Stage 2. The next stage applied only to households in which the employed only subgroup was selected. This stage involved subsampling these households in order to focus limited survey resources on the employees of most interest; that is, on leave takers and leave needers. The 2012 FMLA study suggests that about 80 percent of U.S. workers belong to the employed only group (don't need and have not taken family and medical leave). Conducting extended interviews with all such cases would have been expected to yield more than 6,000 interviews, but consistent with the 1995 (Wave 1), 2000 (Wave 2), and 2012 (Wave 3) surveys, we needed for analysis only about 1,800 completed interviews with employed only employees. Consequently, households in which the employed only subgroup 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 cooperating and ended the call. It is important to note that the survey weights (described in Section 1.6) adjust for this subsampling so that in the weighted survey estimates, the employed only subgroup is represented in proportion to its actual size.

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

Stage 3. In this stage we randomly selected an eligible adult from the family and medical leave subgroup identified in Stage 1 to be the extended interview respondent. For households in which there was exactly one employee in the selected family and medical leave subgroup, that employee was automatically selected. In households where there was more than one employee in the selected family and medical leave subgroup, one was randomly selected among those in the subgroup. If the selected employee 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.

1.1.2 Screening for the Target Population (Cell Phone, Web)

Cell phone

The cell phone sampling design did not screen for household members, but instead treated the cell phone as individual (rather than household level). As such, there was no need for household rostering. Stage 2 subsampling was still applied to the employed only respondents.

Web

The web sampling design mimicked the cell phone sampling design in treating the survey as individual (rather than household level). We assumed that a panelist's email address belonged to one person, and that would be the person who could complete the survey. Stage 2 subsampling was still applied to the employed only respondents.

1.1.3 Comparison of the 2012 and 2018 Employee Sampling Designs

The 2018 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 2012 Employee Survey, without threatening objective 1.

One key change in 2018 was the addition of the web survey using a nationally representative probability-based web panel to supplement the RDD phone completes. This became a necessity during the early months of data collection as RDD response rates remained well below established targets. The general and continuing decline in RDD response rates has forced researchers to look to alternatives in collecting representative data, including transitioning from telephone to self-administered web surveys, using a probability-based web panel (AAPOR, 2019). As such, collecting both RDD and web panel data in the 2018 Employee Survey allows it to be the transition wave between modes. Having both the RDD and web samples allows comparison to past waves and sets up the web to be the mode of data collection for future waves. Ensuring that longitudinal comparisons remained intact was thoughtfully considered when we added the web panel to the 2018 Employee Survey.

In this section we highlight the key consistencies and differences between the 2012 and 2018 sampling designs. Revisions to the survey questionnaire are described in Section 1.2.

Consistencies

Many key design elements are consistent between the 2012 and 2018 Employee Surveys. Critically, the target population has essentially remained the same. Both the 2012 and 2018 surveys are designed to make inference to employed adults in the United States.

Another key consistency is the emphasis on the key subgroups of leave needers and leave takers. In 2012 and 2018, the screener is designed to identify leave needers and leave takers (in the household, for the landline sample; in general, for the cell sample and web panel), and then select them for the extended interview at a higher rate than were employed only adults.

The 2012 and 2018 Employee Surveys both feature an overlapping dual frame landline and cell phone design. The 2018 Employee Survey also continues to select only one eligible adult from each sampled household among the landline sample.

In summary, both the 2012 and 2018 Employee Surveys are national, probability-based, high-coverage surveys that screen for the target population and oversample the key subgroups of leave needers and leave takers.

Notable differences

Oversample of low-wage workers. The low-income population is known to use prepaid cell phones at a rate higher than the general population (among prepaid phone users, 53 percent have annual household incomes below $30,000, compared to 24 percent among non-prepaid phone users; McGeeney, 2015). To provide an oversample of low-wage workers, a group that is traditionally difficult to reach and survey, we used prepaid cell phone flags within the RDD telephone sample. We considered but rejected proposing the addition of an income question to the screener. Income is known to be a sensitive survey question that many respondents will not answer, and including it might lead respondents to discontinue the interview (Yan, Curtin & Jans, 2010).

Traditionally, oversampling to reach low-income populations has been achieved through geographic targeting of landline numbers at fine levels such as Census tracts (Kalton & Anderson, 1986). Because cell phone numbers are not easily associated with geographies, the prepaid flag is instead a feasible option for the cell phone RDD frame. Based on McGeeney (2015), we implemented a 33 percent prepaid phone oversample (implemented via subsampling 75 percent of the non-prepaid phones) to increase the expected sample representation of households with incomes below $30,000, and by extension provide a larger sample of respondents earning less than $15 per hour.

Using the panelist profile data, we were able to oversample the low-income web panelists by 25 percent.[2]

Use of activity flags. A standard component of cell phone RDD design is the use of "activity flags" (McGeeney & Kennedy, 2016), which indicate whether a given number has had any recent activity. Given that cell phone numbers flagged as active are approximately five times more productive than inactive numbers, the latter are often removed from the sample.[3] Yet inactive numbers cover 8 percent of the population (McGeeney & Kennedy, 2015).

Thus, to avoid coverage issues, we subsampled inactive numbers rather than excluding them. As a part of the optimal allocation of the phone sample between landline and the various categories of cell phones (active/inactive/unknown; prepaid versus contract), we have determined that the subsampling rate that provides the optimal balance between the higher cost of data collection (higher subsampling rates) versus higher design effects due to subsampling (lower subsampling rates) is 30 percent. Because prepaid cell phones (which were used to oversample low-wage workers) are always flagged active, we applied the 30 percent subsampling rate for inactive numbers on top of the 75 percent subsampling rate of non-prepaid phones.

Oversample of paid leave areas. The Employee Survey oversampled paid leave areas (states of California, New Jersey, and Rhode Island) at a 25 percent rate. Early in data collection, the state of New York passed a new paid leave policy, effective January 1, 2018. Thus, data collected from New York after July 1, 2018, were included in the paid leave designation at the analysis stage. On the landline frame, we achieved geographic targeting by matching the frame entries (landline numbers) to the modal geographies (e.g., states) in which the "related" listed landline numbers in the same 100- or 1000-banks were found. On the cell phone frame, we achieved geographic targeting by pegging the frame entries (cell phone numbers) to "rate centers," which are somewhat arbitrary yet well-defined telecom service areas. Cell frame "undercoverage" and "overcoverage" (in-state employees who have out-of-state cell numbers and out-of-state employees who have in-state cell numbers, respectively) reduce efficiency of oversampling, which we accounted for in our state sample size projections. The use of frame data such as rate centers is standard in RDD surveys (Barron et al., 2015) and Abt has employed it in dozens of statewide, citywide, and other contained-geography surveys.

California and New Jersey (and later New York) were good candidates for oversampling because they are relatively populous (representing 12.1 percent and 3.0 percent of the total U.S. labor force, respectively).[4] Selecting more populous paid leave areas to oversample is advantageous because states that are

---

2 Low income was defined as panelist profile income below $35,000.

3 In a study conducted by Abt based on 2014 data (used in McGeeney & Kennedy, 2016), the cell phone numbers flagged as active resulted in contact with an adult in the residential non-institutionalized population at a rate of 86 percent, whereas those flagged as inactive resulted in contact at a rate of only 15 percent.

4 Abt calculation, based on American Community Survey 2014 Public Use Microdata Sample (PUMS) data.

oversampled must be weighted down when computing national estimates, so that each state is represented in proportion to its population. By selecting the more populous states, we maximize the precision of survey estimates by minimizing the variation in sampling weights.

Hybrid design, including RDD and web panel data collection. The telephone data collection for the 2018 Employee Survey started in March 2018 and immediately faced productivity challenges. Careful internal monitoring conducted over an eight-week period confirmed that the precipitous decline in RDD-based response rates had accelerated since 2014, when we first developed the sampling design and survey budget.

To ensure that established targets were met, we supplemented the RDD-based responses with sample drawn from the Ipsos KnowledgePanel, a representative, probability-based online panel, which makes available rich profile data, including demographics and economic activity data. We drew a stratified sample that used the same principles of oversampling paid leave states and low-income workers.

1.2. Questionnaire Development

Over the four waves of data collection (1996, 2000, 2012, 2018), the FMLA survey questionnaires have as much as possible followed the model implemented in the previous wave's survey questionnaire. This approach preserves comparability, allowing analyses of changes over time. However, the final 2018 Employee Survey questionnaire differs substantially from the 2012 instrument for several reasons. The 2018 survey was changed based on Wage and Hour Division (WHD) comments. In addition, each survey was conducted in a very different economic environment: across the survey's four waves, the unemployment rate was 5.6 percent in 1995, 4.0 percent in 2000, more than 8 percent in 2012, and 3.9 percent in 2018.[5]

Questionnaire development for the 2018 Employee Survey proceeded in four phases, discussed in detail in this section:

1) Revisions to the screener
2) Revisions to the main survey
3) Cognitive testing
4) Piloting

To facilitate comparisons and identification of trends, we began development with the 2012 survey questionnaire as a base. To ensure that new questions would adequately capture the range of issues and experiences with regulatory changes since 2012 and the possibilities for future efforts, we gathered information from various sources. Those included the DOL's Wage and Hour Division, public comments in response to DOL's 2006 Request for Information, peer-reviewed published literature, and "gray" literature such as newspaper articles, policy papers, and research reports. We conducted in-person interviews with DOL staff. We held two listening group sessions with representatives from nine employee and four employer stakeholder organizations to elicit their feedback. The resulting questionnaire drafts were reviewed by a Technical Working Group.

1.2.1 2018 Questionnaire Overview

The survey questionnaire comprises five sets of question items.

---

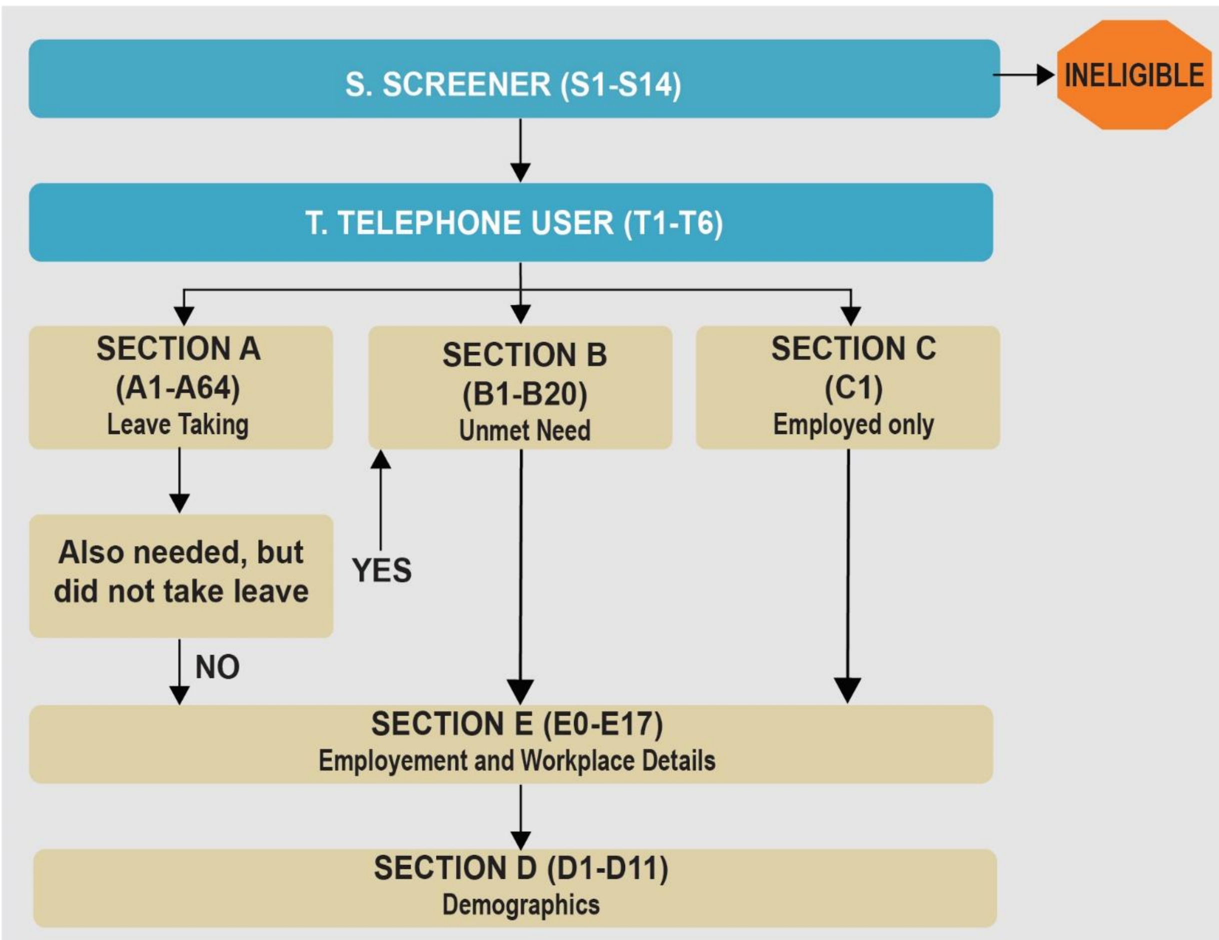
5 These raw monthly unemployment rates are drawn from Bureau of Labor Statistics data available at data.bls.gov/timeseries/LNS14000000. The 1995, 2000, 2012, and 2018 rates are simple averages of the year's 12 monthly values.

The first is the screener, shown as Block S in Exhibit 1.1. As described earlier in Section 1.1, the screener selects the target respondent according to the employee's status as a leave taker, a leave needer, or employed only. Second, eligible respondents are asked a series of telephone usage questions for the purposes of integrating and weighting the cell and landline sample frames (Block T).

Third, respondents are asked questions related to leave that take them to Block A (leave takers) or Block B (leave needers) or Block C (employed only). At the beginning of each, the respondent's classification is confirmed. Respondents classified as leave needers or employed only proceed directly to Block B or Block C, respectively. But leave takers are asked whether they *also* had an unmet need for leave during the reference period. If so, they also are asked the Block B questions.

Last, all respondents are asked questions related to their employment situation and benefits (Block E), followed by standard demographic questions (Block D).

Exhibit 1.1. Structure of the 2018 Employee Survey



Source: Design Report Wave 4 FMLA Surveys

A discussion of questionnaire revisions and additions follows below. Appendix C details changes made from the 2012 to 2018 Employee Surveys. Appendix A provides a copy of the 2018 Employee Survey.

1.2.2 Revisions to the Screener

This subsection discusses revisions to the screener.

Education

The 2018 survey screener added a question asking for highest level of education completed.

Number of jobs

The 2018 survey screener added a question asking whether the respondent had more than one job at the time of the reference point of 12 months prior to the interview.

1.2.3 Revisions to the Main Survey

In this section we describe the substance changes to the main survey questionnaire from the 2012 Survey to the 2018 survey.

Concurrent leave

We added questions asking whether leave was taken by two household members for the same event, and if so, how much of the leave overlapped.

Income sources during leave

We changed the series of questions asking about sources of income while on leave to include different categories with a more up-to-date list.

Job tenure

We added to the employment section a series of questions asking for the start date of their job worked 12 months prior to the interview, number of jobs held 12 months prior, and number of hours worked 12 months prior. If more than one job was held 12 months prior, the same questions were asked about their main job or the job where they worked the most number of hours. This series concludes by asking how many hours they worked at all of their jobs in total at the focal point of 12 months prior.

Absence from work policies

We added to the employment section of the survey a series of questions asking whether the respondent was allowed to take paid leave off work for various reasons, including own illness or medical care, illness of another family member, routine childcare, eldercare, and errands.

Worksite characteristics

We added to the employment section questions about worksite size, industry, occupation, and ZIP code.

Demographics

We added questions asking for the salary at their current job to the demographics section.

Key Changes to the Content of the 2012 Employee Survey for the 2018 Employee Survey

- Expanded focus on awareness of FMLA (e.g., whether it is paid, whether the respondent believes that he/she is covered).
- Increased emphasis on need for leave for medical conditions that do not meet the “serious medical” threshold, such as eldercare.
- Revised questions on employment tenure to identify FMLA coverage status at the start of the recall window.
- Revised questions on access to paid leave (for all respondents), and pay received while on leave (for leave takers).
- Revised focus on a respondent’s “main job,” defined as the one most likely to provide FMLA coverage (for respondents who work multiple jobs).
- Revised questions on family income and added questions on own income. This was done to better identify primary earners.
- Deleted secondary questions about leave taking (e.g., for short, non-recent leaves; for leaves taken during the last 18 months), which had been included in Wave 3 to allow comparability with Wave 2. Retained questions about leave taking during the past 12 months.
- Deleted non-primary leave reasons that were not common in prior waves (e.g., specifics of deployment or military-related leaves, specifics of relationships of non-relative care recipients).
- Deleted secondary questions about the impacts of leave that were not common in prior waves (e.g., specific reasons why employees did not return to the same employer after leave).

Other changes relative to the 2012 Survey questionnaire

In addition to the revisions listed above, some questions were deleted from the 2012 Employee Survey.

> *A5a/B6a. If reason for longest/most recent* leave is to address issues arising from the deployment of a military member:* What type of deployment-related issue did you need to address for this leave?*
>
> *A7/B8. If reason for longest/most recent leave* is because of other NON-relative's health condition:* What is that person's relationship to you?*
>
> *A9/B10. If reason for longest/most recent* leave was to care for someone other than self that is 18+ years old:* Was this leave taken in order to care for a member of the military for a service-related health condition or injury?*
>
> *A9a/B10a. If reason for longest/most recent* leave was to care for a member of the military:* What is that person's relationship to you?*
>
> *A11/B12. For longest/most recent* leave in last 18 months if for own/other serious health condition:* Did [you/your care recipient] require a doctor's care at any time during this leave?*
>
> *A12/B13. If longest leave/most recent* in last 18 months required doctor's care:* [Were/Was] [you/your care recipient] in the hospital overnight at any time during this leave?*
>
> *A19a. If longest/most recent* leave in last 18 months taken to care for a military member:* How much time was needed for the care for the military member?*
>
> *A19d. If screener shows that another adult in household took leave in last 18 months, AND took for same reason as longest/most recent*:* How much time in total did this person take off from work for the same reason you mentioned?*
>
> *A21. 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?*
>
> *A23c. Were you unable to afford an unpaid leave?*
>
> *A23f. Were you able to maintain or pay for health insurance?*
>
> *A29. Why wasn't your medical certification accepted on the first submission?*
>
> *A39. Did you pay out of your own pocket for your medical RE-certifications (for example, a co-pay or portion of the cost)?*
>
> *A47a. Was receiving some of the pay as part of paid time off, or PTO your choice, did your employer require it, or both?*
>
> *A47b. Was receiving some of the pay as part of your sick days or sick leave your choice, did your employer require it, or both?*
>
> *A47c. Was receiving some of the pay as part of your vacation days or vacation leave your choice, did your employer require it, or both?*
>
> *A47d. Was receiving some of the pay as part of paid personal leave your choice, did your employer require it, or both?*

A47e. Was receiving some of the pay as part of maternity leave your choice, did your employer require it, or both?

A47f. Was receiving some of the pay as part of paternity leave your choice, did your employer require it, or both?

A48a. Was some of the pay you received part of… Temporary disability insurance?

A48b. Was some of the pay you received part of… State-paid family leave?

A48c. Was some of the pay you received part of… State-paid disability leave?

A49. When you received pay during your leave, was it the same amount as your regular pay or only part of your pay?

A50. Over the entire time you were on leave, about how much of your regular pay did you receive in total?

A61. Why didn't you return to work [at the same employer]?

B5a. Were all the times you needed leave but did not take it since [INSERT 18 MONTH PERIOD] for the SAME reason or condition, or were they for DIFFERENT reasons or conditions?

B5b. For how many TOTAL reasons or conditions did you need leave from work, but not take it, since [INSERT 18 MONTH PERIOD]?

B14. How many different times, since [INSERT 18 MONTH PERIOD], did you need leave for the REASON OR CONDITION you mentioned?

B14a. And how many different times did you need leave for this reason or condition, IN THE LAST YEAR [12 MONTHS, INSERT DATE]?

B19. Why were you denied leave?

E4. At your place of employment, is there a notice posted that explains the federal Family and Medical Leave Act?

E5. Now I'm going to read you some questions about your current employment situation. Since [INSERT 18 MONTH PERIOD], have any co-workers where you work taken leave for family or medical reasons?

E6. As a result of these co-workers taking leave, did you work more hours than you usually do? Work a shift that you do not normally work? Take on additional duties? Take on different job responsibilities?

E7. I'm going to read a list of benefits that some employers offer to their employees. Are you eligible to receive any of these benefits? Flextime? Flexplace or telecommuting? Job sharing? Paid family leave? Paid vacation? Paid sick time? Paid time off? Break time for mothers who are breastfeeding?

E8. Does your employer have an attendance policy that includes penalties for absences?

1.2.4 Cognitive Testing

The objectives of cognitive testing are to identify problems that respondents are likely to have with any part of the response process and to help eliminate sources of response error, by trying out the material with up to nine persons similar to the target respondents, observing and discussing the items with them.

First, we used cognitive testing to aid in the development of new survey items and to test the appropriateness of published survey questions for use in this context. Specifically, we sought to ascertain whether the wording of individual questions and response categories adequately captured the range of respondent experiences with taking leave and needing to take leave, particularly on an intermittent basis. Second, we attempted to identify recall issues for questions pertaining to multiple leave occasions or multiple conditions over a 12-month period. 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 new questionnaire design under a variety of different respondent conditions (leave takers, leave needers, intermittent and long-term leave, for one's own serious health condition and to care for others).

We conducted cognitive testing on the 2018 Employee Survey in February 2017. We conducted nine interviews and recruited a convenience sample of respondents for the testing, ensuring diversity across a range of characteristics. Specifically, we included respondents employed in different types and sizes of worksites (e.g., professionals at large firms, unskilled workers, small business operators). To detect variation in response quality (comprehension, retrieval recall, and response), we also included respondents across age, gender, and education levels.

Because the 2012 Employee Survey was already field tested, we focused the 2018 Employee Survey cognitive testing probes on the new or revised survey questions. For those questions, the testing focused on the following:

- Are all the words understood?
- Do respondents interpret the question in the same way?
- Are all response choices appropriate?
- Are the range of response choices actually used?
- Do respondents correctly follow directions?

This testing took place in person in Abt's Chicago facility. Interviewers introduced the study and the tasks associated with cognitive testing (e.g., "thinking aloud"). They also administered specific, scripted probes for select survey questions identified by expert review as more likely than others to be ambiguous or difficult to answer. Other times the interviewer simply asked generic probes ("What were you thinking?") if the respondent seemed to have difficulty answering.

Cognitive testing was led by two professional staff. One person administered the questionnaire while the other took notes to observe any nonverbal cues, adding additional probes and follow-up questions as necessary.

1.2.5 Survey Pre-Test

In addition to cognitive testing, we conducted nine pre-tests prior to submission of the study's OMB Paperwork Reduction Act package. One goal of the pre-testing was to determine overall survey burden (i.e., the number of minutes to complete the survey). We conducted the 2018 Employee Survey pre-tests using the same protocol and setting as the planned survey. In addition to assessing burden, the pre-testing helped determine whether respondents were interpreting questions as intended and whether the order of questions would influence responses. Doing this was especially important for testing sections where changes were more substantive. Similarly, new questions about sources and amount of pay during leave were pre-tested to ensure respondents understood and could provide answers.

Senior survey and project staff monitored the interviews as they took place. Additionally, the interviewers requested feedback from respondents about the survey's questions during a debriefing.

1.3. Data Collection Procedures

The 2018 Employee Survey underwent OMB clearance from March 2017 to January 2018. Shortly after final clearance, we began data collection. Interviewing for the Employee Survey was conducted using both computer-assisted telephone interviewing (CATI) and computer-assisted web interviewing (CAWI). Both modes were conducted in English and Spanish.

1.3.1 Telephone Interviewing

Interviewer training

Interviewers received intensive training to prepare them to administer 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, we explained the purpose and goals of the study. In telephone surveying, the most critical training 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. In the training, we 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. We also conducted mock interviews across all of the interview types (leave taker, leave needer, dual leave takers/needers, and employed only).

Call design

Telephone survey administration took place from Abt's centralized call centers. We set different calling rules for landline and cell phones, because what may seem like a moderate number of calls to a household's landline can seem excessive to a person's cell phone. As in 2012, we initially attempted to reach respondents on landlines a maximum of 15 times, and on cell phones a maximum of eight times. However, we eventually lowered these numbers to eight for landlines and six for cell phones (for non-contacts and callbacks), as more calls did not yield more results. For callbacks, more calls were permitted if the interviewer made contact with an eligible household but was asked to call back later.

Interviewers placed phone calls from 5:00 pm to 9:00 pm on weekdays, from 10:00 am to 6:00 pm on Saturdays, and from noon to 9:00 pm on Sundays. Daytime calling during the week was used periodically to attempt to reach non-contacts. We analyzed production data to determine the best days and times to contact study members and avoid refusals. In addition, we made special arrangements to accommodate other times of the day if a respondent requested (i.e., outside those regular calling hours). 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.

Consistent with the Telephone Consumer Protection Act (47 U.S.C. 227), 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, until the telephone number was determined to be incorrect or out of service, or until the maximum number of attempts was reached. We left a voicemail message two or three times to introduce the study and mention the incentive, as appropriate, in English or Spanish as appropriate. For cell phones, we left messages on the first non-contact; for landlines, we left messages on the third non-contact. We varied the timing of voicemail messages in an effort to increase response rates. The protocol was tailored for landlines and cell phones based on best practices for each mode.

For respondents completing the questionnaire on a cell phone, we issued a $15 incentive to compensate for per-minute carrier charges. (Landline respondents received no incentive.) Using pre-translated instruments, the interview was conducted in English or Spanish, depending on the respondent's preference.[6]

The average length of completed interviews was 21.9 minutes for landline, 22.4 minutes for cell phone.

Refusal conversion

We attempted to re-contact all respondents who declined to be interviewed, except for those respondents who were hostile during their initial refusal to participate in the interview. Our refusal conversion approach consisted of three steps. First, we coded refusals by type and by when they occurred in the survey. This information was used to adjust our protocols. Second, we tracked refusals in real time, so interviewers who were generating a much higher than average number of refusals could be identified and retrained. Finally, we selected a subset of households to receive a refusal conversion letter.

As RDD data collection began winding down in June 2018, we identified the refusals that we wanted to target with a refusal conversion letter. We were able to match addresses to phone numbers for 3,560 of 5,960 refusals and sent letters to the sample that was matched. The letter explained the survey and addressed specific reasons for refusal, provided information to validate the legitimacy of the survey, offered an incentive, and provided a toll-free number that the respondent could use to call and do the interview. It also said that an interviewer would be contacting them again soon. A series of up to eight outbound calls was made to the address-matched sample.

For cell phone cases, non-respondents (those who refused before completing the screener) were offered a $20 incentive ($5 more than the standard $15), and non-cooperative respondents (those who qualified and started the survey, but refused to complete the survey) were offered $40. Landline cases were now offered an incentive: the same $20 for non-respondents or $40 for non-cooperatives.

This refusal conversion effort yielded an additional 20 landline and six cell phone survey completes.

1.3.2 Web Survey

Web survey addition

The first two months of RDD data collection yielded less than 8 percent of the target total of 4,000 completed interviews. Our initial planning had assumed some decline in response rates from Wave 3, partially offset by other changes in survey methods. However, review of the experience over the first two months confirmed that our initial planning targets did not adequately account for the actual decline in RDD productivity and response rates and the growth in spam flagging and call blocking technologies.

To address these challenges, we made several modifications to our data collection protocols:

- increased the selection rate of the employed only population;
- decreased the number of follow-up attempts on each piece of sample;
- systematically dialed sample, paused dialing, and then resumed dialing;

[6] In the 2018 Survey, 0.3 percent of all numbers dialed were recorded as non-interviews because the respondent spoke neither English nor Spanish.

- modified the survey introduction and lead-in; and
- pre-screened all landline sample for disconnected numbers.

These changes, however, produced only very minor benefits, and response rates stayed low. In consultation with DOL, we transitioned the data collection to a hybrid approach, adding a nationally representative probability-based web panel (Ipsos KnowledgePanel[7]) to supplement the RDD efforts.

CATI survey adaptations to web survey

Once the use of a web survey was approved, the interview-administered phone interview was adapted so it could be programmed as a self-administered web survey. In adapting the phone survey for web administration, we kept the phone and web versions as consistent as possible. We made only the following changes, which reflect best practices when adapting a phone survey to web (Dillman et al., 2014):

- Removed all "Refused" response options and most of the "Don't know" response options. The web questionnaire was programmed so that if a respondent left a key question blank, a message would appear alerting the respondent that an answer was missing. The respondent then could either give an answer or continue without answering. Questions left blank were coded in the data as "Refused."
- Changed the introduction, from the RDD interviewer screening household members to identify the key informant in the household, to asking respondents to complete the self-administered web survey. (KnowledgePanel is made up of individuals, not households.)
- Removed selected demographic questions, as the web panel profile data provided that information.
- Changed probes read by the RDD interviewer to follow-up confirmation questions.
- Presented definitions that RDD interviewers were expected to read aloud as hover-over text definitions.
- Displayed for all web panel respondents clarifying text that RDD interviewers were to use as needed.

Web panel provider Ipsos programmed the web survey. In conjunction with Ipsos, we reviewed and tested the survey thoroughly and checked the test data prior to the survey launch. Selected screen shots from the web survey can be found in Appendix A of the separate Methodology Report Appendices volume.

Web survey administration

Once assigned to the Employee Survey, web panel members received a notification email letting them know the survey was available for them to complete. This email contained a link that sent them directly to the survey. No login name or password was required. After three days, all non-responding panel members in the sample received a reminder email. Additional email reminders were sent as needed. To assist panel members with their survey taking, each individual member also could access the survey via a personalized member portal.

In addition to Ipsos's regular incentive to participate in the panel, respondents who completed the Employee Survey received points, which represented a $5 incentive.

---

7 https://www.ipsos.com/en-us/solutions/public-affairs/kowledgepanel

1.4. Response Rates

For each survey mode, we computed the response rate separately for the screener, extended interview, and overall. The web mode also includes a cumulative response rate, which takes into account several aspects for the recruitment by Ipsos of the panel prior to the survey effort. Unless otherwise noted, the response rates in this report are computed according to current American Association for Public Opinion Research (AAPOR) Standard Definitions of case codes and outcome rates (AAPOR, 2016).

We also computed an alternative set of telephone response rates using the formulas from the 1995 and 2000 Employee Surveys. These formulas are outdated, especially because the 2018 Employee Survey included a cell phone sample in addition to a landline sample, but are included for comparison purposes.

1.4.1 Final Call Outcomes

The final outcomes of call attempts for the screener and the extended interview are presented in **Exhibit 1.2**. These outcomes are presented separately for the landline and cell phone samples as well as combined (unweighted).

Exhibit 1.2. Telephone dispositions for the 2018 Employee Survey, by sample

	AAPOR	Total RDD Sample		Landline Sample		Cell Sample	
Disposition	**Code**	**Screener**	**Extended**	**Screener**	**Extended**	**Screener**	**Extended**
Interview (Category 1)							
Complete	1.100	7,631	739	2,715	189	4,916	550
Eligible, non-interview (Category 2)							
Refusal	2.110	20,872	427	7,274	199	13,598	228
Respondent never available	2.210	99		53		46	
Telephone answering device	2.220	39,871		9,481		30,390	
Physically or mentally unable	2.320	533		301		232	
Language problem	2.330	540		175		365	
Unknown eligibility, non-interview (Category 3)							
Always busy	3.120	5,676		1,001		4,675	
No answer	3.130	21,256		11,780		9,476	
Call blocking	3.150	914		114		800	
Not eligible (Category 4)							
Fax/data line	4.200	3,485		3,246		239	
Disconnected number	4.320	52,181		35,657		16,524	
Temporarily out of service	4.330	5,638		1,117		4,521	
Cell Phone in LL Sample	4.420	18		18		0	
Business, gov't office, other org.	4.510	7,214		5,353		1,861	
No eligible respondent	4.700	2,161		64		2,097	
Total Numbers Dialed		**168,089**	**1,166**	**78,349**	**388**	**89,740**	**778**

Source: 2018 Employee Survey

1.4.2 Telephone Response Rates

Exhibit 1.3 shows the screener, extended interview, and overall AAPOR response rates for telephone interviews by sample frame and overall. This AAPOR Response Rate 3 formula uses an eligibility coefficient (e) that is computed as $(I+R+NC+O)/((I+R+NC+O)+NE)$, where I denotes completed interviews, R is refusals, NC is non-contacts, O is other non-responses, and NE is not eligibles. All of the Employee Survey response rates reported here are based on unweighted data because within each frame, all telephone numbers had the same probability of selection.

Exhibit 1.3. AAPOR telephone response rates for screener interview, extended interview, and overall, by sample

	Landline Sample			Cell Sample			Total Sample		
Response Rate Formula	**Screener**	**Exten.**	**Overall**	**Screener**	**Exten.**	**Overall**	**Screener**	**Exten.**	**Overall**
AAPOR Response Rate 3*	11.3%	48.7%	5.5%	8.3%	70.7%	5.8%	9.1%	65.1%	5.8%
Formula: I/((I+P) + (R+NC+O) + e(UH+UO))									

Source: 2018 Employee Survey
Key: AAPOR = American Association for Public Opinion Research; I = completed interview; P = partial interview; R = refusal; NC = non-contact; O = other non-response; UH = unknown if household; UO = other unknown eligibility; NE = not eligible.
* The *e* coefficient in AAPOR RR(3) was computed as (I+R+NC+O)/((I+R+NC+O)+NE).

Exhibit 1.4 shows response rates computed for the 2018 Employee Survey using formulas from the 1995 and 2000 Employee Survey reports. The "Lower" and "Higher" response rate formulas reflect a range of assumptions regarding the eligibility of telephone numbers where no respondent ever answered the telephone. The "Lower" response rate is based on the standard formula used by Westat to compute response rates for RDD surveys at the time of the 1995 survey. This response rate assumes that 27 percent of telephone numbers that were never answered are in fact eligible, and that 60 percent of telephone numbers where the call was answered by a machine are eligible. The "Higher" response rate formula is similar to the response rate formula used by the University of Michigan for the 1995 Survey of Employees. This response rate formula excludes telephone numbers that were never answered and assumes that all calls that reached an answering machine were eligible.

Exhibit 1.4. Legacy telephone response rates for screener interview, extended interview, and overall, by sample

	Landline Sample			Cell Sample			Total Sample		
Response Rate Formula	**Screener**	**Exten.**	**Overall**	**Screener**	**Exten.**	**Overall**	**Screener**	**Exten.**	**Overall**
1995 / 2000 FMLA "Lower" RR Formula C/(C+R+.27NA+.6M+LP+MC+ONR)	13.8%	48.7%	6.7%	11.9%	70.7%	8.4%	12.5%	63.4%	7.9%
1995 / 2000 FMLA "Higher" RR Formula C/(C+R+M+LP+MC+ONR)	13.6%	48.7%	6.6%	9.9%	70.7%	7.0%	11.0%	63.4%	7.0%

Source: 2018 Employee Survey
Key: C = completed interview; R = refusal; NA = no answer; M = answering machine; LP = language problem; MC = maximum calls; ONR = other non-response

Landline RDD response rates

As shown on **Exhibit 1.3 (left panel)**: For the landline RDD sample, the screener response rate is 11.3 percent. The extended interview response rate of 48.7 percent represents the proportion of landline sample interviews that were completed among those respondents eligible and selected for the extended interview. We computed the overall landline sample response rate as the product of the screener and extended interview response rates. The overall landline sample response rate is 5.5 percent.

Cell RDD response rates

As shown on **Exhibit 1.3 (middle panel)**: The screener response rate for the cell RDD sample is 8.3 percent. The extended interview response rate of 70.7 percent represents the proportion of cell sample interviews that were completed among those respondents eligible and selected for the extended interview. We computed the overall cell sample response rate as the product of the screener and extended interview response rates. The overall cell sample response rate is 5.8 percent.

Total sample response rate

The total sample telephone response rates shown in **Exhibit 1.3 (right panel)** combine the landline and cell phone sample response rates in a weighted fashion according to the method recommended in the 2016 AAPOR Standard Definitions Report.[8] Response rates for each sample frame were combined using weights proportional to the share of extended interviews completed from each respective sample frame.

1.4.3 Web Response Rates

We computed four response rates for the web mode, as shown in **Exhibit 1.5**: (1) a screener response rate, (2) an extended interview response rate, (3) an overall web response rate, and (4) a cumulative web response rate. The screener and extended interview response rates are based on the AAPOR Response Rate 3 formula.

Exhibit 1.5. AAPOR web survey response rates

Response Rate Formula	Screener	Extended	Overall	Cumulative
AAPOR Response Rate 3	82.7%	91.0%	75.2%	5.5%
Formula: I/((I+P) + (R+NC+O) + e(UH+UO))				

Source: 2018 Employee Survey
Key: I = completed interview; P = partial interview; R = refusal; NC = non-contact; O = other non-response; *e* coefficient; UH = unknown if household; UO = other unknown eligibility

The response rate for the screener section of the web survey is 82.7 percent. To compute the web screener response rate, we first estimated the eligibility rate of cases with unknown eligibility (those who never started the web survey) using the proportion of screened cases that were determined to be eligible for the survey. The extended interview web response rate of 91.0 percent represents the proportion of web interviews that were completed among those respondents eligible and selected for the extended interview. The overall response rate for web mode of 75.2 percent is computed as the product of the screener and extended interview response rates.

Because web respondents were recruited from an online panel, we also computed a cumulative web response rate, which incorporates the average panel recruitment rate of 12.0 percent and the household profile rate of 60.7 percent. The final cumulative web response rate is computed as the product of the overall web response, the average panel recruitment rate, and the household profile rate. The cumulative web response rate for the Employee Survey is 5.5 percent.

1.4.4 Summary

These overall response rates for the Employee Survey are in line with the response rates for similar surveys. The AAPOR Task Force on the Future of U.S. General Population Telephone Survey Research reported decline in response rates from 16 percent on landline and 12 percent on cell phones in 2009, to 9 percent on landline and 7 percent on cell phones in 2015, using their sample of surveys that maintained their methodology over time (Dutwin & Lavrakas, 2016). Kennedy & Hartig (2019) reported typical response rates in its telephone RDD surveys were 6 percent in 2018. There is a limited number of large-scale and comparable federal or state-level phone surveys. The 2019 NYC Community Health Survey had a combined response rate of 7 percent across both RDD landline and cell (Ruther & Sokolowski, 2020). Similarly, the 2017 California Health Interview Survey reported a 7 percent response rate with a design that was primarily RDD with an addressed-based sampling oversample (Dutwin et al., 2018). On the web survey side, the Survey of American Family Finances conducted in 2014 for The Pew Charitable Trusts

8 https://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx

(2015) using the response rates of Ipsos KnowledgePanel[9] have been relatively stable, with a 10 percent recruitment rate and about a 60 percent participation rate, for a cumulative web response rate of 6 percent for both federal and commercial clients after taking into account the recruitment, household profile rate, and study completion rate (Hays, Liu, & Kapteyn, 2015).[10] More recently, the Federal Reserve Board conducted the Survey of Household Economics and Decisionmaking in 2017 using KnowledgePanel, with a cumulative response rate of 4 percent.[11]

1.5. Analysis of Non-Response of the Employee Survey

The RDD response rate achieved in the 2018 Employee Survey (5.8 percent; see **Exhibit 1.3**) was noticeably lower than the 15.1 percent achieved in the 2012 Employee Survey. In this section, we examine the potential reasons for this drop-off in the response rate, discuss the implications of non-response from a theoretical perspective, and examine the potential threat posed by non-response to Employee Survey estimates from an empirical perspective based on four non-response analyses.

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

Decline in response rates has been observed across practically all surveys, all modes, all sponsors, and all topics over the past several decades whenever compatible response rates could be combined. Williams and Brick (2018) analyzed nine national in-person surveys and found declines in response rates of 0.5 percent to 1 percent per year between 2000 and 2014, with acceleration of the downward trend starting around 2006. The AAPOR Task Force Report on "The Future of U.S. General Population Telephone Survey Research" (2017; see also Dutwin & Lavrakas, 2016) documented a drop in response rates in phone surveys by about one-third from 2008 to 2015. Going back further, a meta-analysis by de Leeuw and de Heer (2002) of response rates to principal government surveys across 16 countries generally covering the period from 1981 to 1998 found that response rates declined by about 0.2 percent per year, with both non-contact and refusal rates increasing approximately 0.3 percent per year each.

We have identified two factors that likely account for much of the drop in the response rate: societal changes and changes in cell phone technology.

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, Singer, & Corning, 2000), increased concern about privacy and confidentiality (Singer, 2003), rising hostility toward telemarketers, the possibility of identity theft, and increase in the volume of unwanted telemarketing calls (despite the attempts by the Federal Communications Commission to curtail them). In addition, shifts in the demographic composition of the U.S. population are likely compounding non-response. 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. (In the 2018 FMLA study, this is indirectly reflected in weights, where the average weights of Hispanics is 37,993 versus the average weight of non-Hispanic whites is 28,324). There have been many changes in cell phone usage since the 2012 Survey. Cell phones are now the main (if not the only) type of phone for a majority of households. Recent years have seen an increase in availability of call blocking and spam flagging apps and an increase in the use of messaging apps (AAPOR, 2017).

---

9 Ipsos acquired KnowledgePanel in 2018; at the time of the Pew study, it was known as GfK KnowledgePanel.

10 Board of Governors of the Federal Reserve System website: https://www.federalreserve.gov/econresdata/2015-economic-well-being-of-us-households-in-2014-appendix-1.htm

11 Board of Governors of the Federal Reserve System website: https://www.federalreserve.gov/consumerscommunities/shed.htm

This constellation of factors has led to the general and continuing decline in survey response rates over the past six years, particularly in RDD surveys. According to the Kennedy & Hartig (2019) report cited earlier, since 2012, after a brief plateau, response rates in its surveys declined from 9 percent in 2012 to 6 percent in 2018.

1.5.2 The Nature of Non-Response Bias

Non-response bias is a systematic difference between a population figure and the sample-based statistic that is caused by unit non-response; that is, failure to obtain an interview from a sampled unit. Non-response bias is generally very difficult to quantify, except for a very limited number of situations where an external "golden standard" is available (some of which include voter studies where the fact of voting is reflected in the voter registration data, or studies of college students where self-reports are compared with university administrative data). The existing survey methodology literature features two approaches to conceptualizing non-response bias.

- Whether the person responds or not can be considered as their innate characteristics; non-response is thus deterministic.

In this framework, non-response bias of a sample mean *y* where *R* is the response rate, $\bar{y}_N$ is the mean of non-respondents and $\bar{y}_R$ is the mean of respondents:

$$\text{NR Bias}[\bar{y}] = (1 - R)(\bar{y}_N - \bar{y}_R)$$

According to this formula, non-response bias is absent either when there is no non-response (*R*=100 percent) or when non-respondents are identical to respondents (and hence $\bar{y}_N = \bar{y}_R$). This formulation highlights the importance of response rate as a potential determinant of non-response bias. Though response rate can generally be understood as the fraction of the sampled units who completed the survey, there is a multitude of fine points in who is considered eligible to take the survey, and what to do with cases of unknown eligibility. Calculation of response rates in surveys follows a strict protocol in accordance with AAPOR *Standard Definitions* (2016).

- A more flexible, stochastic framework of non-response (Bethlehem, 2002) views non-response as a random event, in which a unit *i* responds with propensity ρ_i. Then non-response bias can be found to be

$$\text{NR Bias}[\bar{y}] = \frac{Cov(\,y,\rho)}{\bar{\rho}} = \frac{\frac{1}{n}\sum_i (y_i - \bar{y})(\rho_i - \bar{\rho})}{\bar{\rho}}, \bar{\rho} = \frac{1}{n}\sum_i \rho_i, \bar{y} = \frac{1}{n}\sum_i y_i$$

It provides an alternative view of non-response: When all individuals have the same response propensity, there is no variability in rho, and hence the numerator is 0. Hence a diagnostic of the potential risk of non-response bias is the variability of response propensities – for example, between different demographic groups.

When assessing the risk from non-response bias, two key properties of non-response are particularly relevant. First, non-response bias can be negligible for some survey estimates and large for other estimates. In other words, non-response bias is an estimate-specific phenomenon. Non-response 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 & Peytcheva, 2008). A second, closely related property of non-response is that response rates alone are 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 non-response bias. (A majority of the studies in his paper came from medicine where frame data could be used to assess bias in unweighted estimates; also these studies generally had relatively high response rates, median 70 percent. Groves (2006) also cautions against misinterpreting studies on non-response rates as "implying that there is rarely, if ever, a reason to worry about non-response 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, Miller et al., 2000; Keeter, Kennedy et al., 2006; Merkle & Edelman, 2002; Pew Research Center, 2012). Tourangeau (2017) and Brick and Tourangeau (2017) suggested that the link between response rates and non-response biases is stronger, although the analysis producing higher correlations between response rates and non-response biases hinges on weighting the study data by the sample sizes.

One reason why estimates from surveys with low response rates can still be accurate is the ability to apply statistical weighting to correct for differential non-response across demographic subgroups (Chang & Kott, 2008; Haziza & Lesage, 2016; Lundstrom & Sarndal, 1999). 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. The 2018 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.

1.5.3 An Empirical Assessment of Non-Response in the 2018 Employee Survey

We used four methods to evaluate non-response to the 2018 Employee Survey. First, a direct contact with non-respondents—a non-response follow-up survey—may be able to provide direct evidence for the reasons of non-response, and for the differences between respondents and non-respondents (hence providing an insight to the y_N-y_R term in the deterministic model of non-response bias). Second, assuming that it takes longer, on average, to reach a lower-propensity respondent than a high-propensity respondent in the "continuum of resistance" model of non-response, a comparison of easier-to-reach versus harder-to-reach respondents is another angle to obtain information on the differences between respondents and non-respondents. Third, response propensity modeling addresses directly the research question of "who is more versus less likely to respond." Fourth, a comparison of survey estimates with external benchmarks is also capable of indirectly addressing the differences between respondents and non-respondents.

Each of these methods provides a different perspective on the potential risk of non-response in the 2018 Employee Survey. We used the same four methods in the non-response analysis effort for the 2012 Employee Survey.

(1) Non-response follow-up survey (NRFU)

The purpose of the NRFU was to collect information on employees who failed to respond to the Employee Survey. Results from the NRFU provide some insight into whether some of the non-respondents differ from respondents on selected characteristics of interest, particularly on several of the more important characteristics collected in the Employee Survey. Note that the NRFU will not provide additional information on *all* non-respondents, because some didn't respond to the NRFU and others weren't selected, but it will provide some insight on differences between respondents and non-respondents for some of the non-respondents.

For the NRFU, we sampled from the original RDD and web samples both non-contact non-responding households and non-cooperative non-responding households:

- ***Non-contact sample.*** These are households/people with whom we attempted contact for the survey but were unable to speak with (in RDD survey) or who did not start the survey at all (in web panel survey).

- ***Non-cooperative sample*.** These are people who previously completed the screener but not the extended interview. We interpreted this to include people who completed the screener and were classified as a leave taker/needer/dual respondent but did not complete the survey for some reason.

The available pool of potential respondents in the non-cooperative sample is very small compared to the non-contact sample. As such, virtually all of the NRFU respondents were from the non-contact sample.

The NRFU pursued randomly selected non-respondents to both the RDD Employee Survey (landline and cell phone components) and the web Employee Survey. NRFU interviewers asked selected respondents to complete a shortened version of the Employee Survey instrument (see Appendix D). Selected respondents were also offered a larger incentive for their participation ($40 for RDD or web panel respondents). The NRFU was conducted from March 11 to April 12, 2019, for RDD and April 12 to April 29, 2019, for web. It was completed by 553 people, 222 from among the RDD Employee Survey non-respondents and 331 from among the web survey non-respondents. Response rates for these groups (AAPOR Response Rate 3) was 2.5 percent for the RDD NRFU employee survey and 3.0 percent of the web NRFU survey.

Results presented in this section are unweighted. Imputation was done to address item non-response to a few demographic items used in this analysis for some of the NRFU respondents (8 percent). The imputation procedures used in this analysis were equivalent to those used when creating variables for Employee Survey weight adjustments for those variables that were defined similarly between the two tasks.

The demographic and geographic variables considered in this analysis are equivalent to those used in the sample weight adjustment process. Two additional variables were added to this set for the NRFU analysis because they tended to be significant predictors of the outcome measures considered in this analysis. These were number of children (range: 0-7; where 7 = seven or more children) and job type (government, private company, or non-profit organization).

We began by examining whether there were any differences in the distribution of several important demographic variables between the Employee Survey and NRFU respondents. The results of this analysis are summarized in **Exhibit 1.6** below. This table shows that the NRFU respondents tended to be younger, particularly in the aged 30-39 subgroup (Employee Survey=20.7 percent, NRFU=28.4 percent). More NRFU respondents had at least one child (Employee Survey=36.6 percent, NRFU=45.0 percent) and a household income of less than $35,000 (Employee Survey=25.0 percent, NRFU=31.5 percent).

Exhibit 1.7 below shows distributions for several key questionnaire items for the Employee Survey and NRFU respondents. This table shows the simple (unadjusted) percentages as well as regression-adjusted means (adjusted percentages). The adjusted percentages were derived from a multinomial logistic model that was fit for each dependent measure (i.e., key item) in the table. All of the demographics listed in **Exhibit 1.6** were included in the model, so the adjusted percentages are showing what the estimates would look like if both the Employee Survey and NRFU respondents were distributed across the demographic variables in exactly the same way. In other words, the adjusted percentages are controlling for differences in the distributions of the Employee Survey and NRFU samples across the **Exhibit 1.6** demographics.

One thing to notice about the results displayed in **Exhibit 1.7** is that the sample size varies greatly between items. This is generally due to the conditional nature of some items; that is, some of the items were asked only of a subgroup of respondents. For example, among the 553 NRFU respondents who answered the leave taker item (yes/no), only the 245 who took leave were routed to the item that asked about the reasons for taking leave.

Exhibit 1.6. Unweighted distribution of Employee Survey and NRFU samples, by demographics

Characteristic	Percentage		
	Employee Survey	NRFU	Diff.
Total	100.0	100.0	0.0
Age Group			
18-29	11.8	13.0	1.3
30-39	20.7	28.4	7.7
40-49	18.2	19.2	1.0
50-59	24.6	19.9	−4.7
60+	24.7	19.5	−5.1
Gender			
Male	44.3	42.7	−1.6
Female	55.7	57.3	1.6
Education			
High school or below	17.4	18.6	1.2
Some college/Associates	33.7	35.3	1.6
Bachelor's degree	29.1	27.7	−1.4
Graduate/professional degree	19.8	18.4	−1.4
Number of Children			
0	63.4	55.0	−8.4
1 or 2	27.9	35.3	7.4
3 or more	8.7	9.8	1.0
Marital Status			
Not married	43.4	39.2	−4.2
Married or live with partner	56.6	60.8	4.2
Household Income			
Under $35,000	25.0	31.5	6.4
$35,000 to $100,000	46.1	40.7	−5.4
$100,000+	28.8	27.8	−1.0
Region			
Northeast	7.6	6.9	−0.8
South	30.2	33.3	3.1
Midwest	22.3	18.8	−3.5
West	12.3	10.1	−2.1
California, New Jersey, Rhode Island	21.0	21.7	0.7
New York	6.6	9.2	2.6
Job Type			
Government	20.6	16.3	−4.4
Private	64.2	68.7	4.5
Non-profit	15.1	15.0	−0.1

Source: 2018 Employee Survey
Note: Sample size for Employee Survey was 4,470. Sample size for NRFU was 553.

Exhibit 1.7. Comparing Employee Survey and NRFU respondents across several key questionnaire items: unadjusted and regression-adjusted means

Item	Sample Size		Unadjusted Percentage			Adjusted Percentage		
	Employee Survey	NRFU	Employee Survey	NRFU	Diff.	Employee Survey	NRFU	Diff.
Total								
Total	4,470	553	100.0	100.0	0.0	100.0	100.0	0.0
Leave Needer/Taker Group								
Leave taker	1,430	167	32.0	30.2	−1.8	32.2	28.4	−3.9
Leave needer	513	72	11.5	13.0	1.5	11.6	12.4	0.8
Employed only	2,128	231	47.6	41.8	−5.8	47.3	44.4	−2.9
Dual taker/needer	399	83	8.9	15.0	6.1	8.9	14.9	5.9*
Leave Taker								
Yes	1,829	250	40.9	45.2	4.3	41.2	43.2	2.1
No	2,641	303	59.1	54.8	−4.3	58.8	56.8	−2.1
Leave Needer								
Yes	912	155	20.4	28.0	7.6	20.5	27.2	6.7*
No	3,558	398	79.6	72.0	−7.6	79.5	72.8	−6.7*
Total Reasons/Conditions for Taking Leave in Past 12 Months								
1	1,201	125	66.5	51.0	−15.5	66.4	51.8	−14.6*
2 or more	605	120	33.5	49.0	15.5	33.6	48.2	14.6*
Reason for Taking Most Recent Leave								
Own illness	1,018	133	56.1	54.3	−1.8	55.9	55.1	−0.8
Pregnancy or child-related reason	435	55	24.0	22.4	−1.5	24.4	20.2	−4.2*
Other person, non-child reason	361	57	19.9	23.3	3.4	19.7	24.7	5.0
Nature of Health Condition for Most Recent Leave								
One time health matter	601	72	40.8	35.0	−5.9	40.8	35.2	−5.6
Treatment that now requires routine schedule care	254	35	17.3	17.0	−0.3	17.3	16.6	−0.8
Ongoing health condition	328	50	22.3	24.3	2.0	22.5	22.8	0.3
Eldercare	52	15	3.5	7.3	3.7	3.5	7.4	3.9*
Other reason	237	34	16.1	16.5	0.4	15.9	18.0	2.1
Time Taken Continuously for Most Recent Leave								
One continuous block of time	1,244	142	69.1	57.3	−11.9	69.1	57.7	−11.4*
Separate occasions	555	106	30.9	42.7	11.9	30.9	42.3	11.4*
What Happened After Recent Leave								
Went back to same employer	1,338	220	92.7	88.0	−4.7	92.4	89.9	−2.5
Went to new employer/did not return to work	106	30	7.3	12.0	4.7	7.6	10.1	2.5

Item	Sample Size		Unadjusted Percentage			Adjusted Percentage		
	Employee Survey	NRFU	Employee Survey	NRFU	Diff.	Employee Survey	NRFU	Diff.
Total Reasons/Conditions for Needing to Take Leave in Past 12 Months								
1	410	52	45.8	37.1	−8.7	45.6	38.6	−6.9
2 or more	485	88	54.2	62.9	8.7	54.4	61.4	6.9
Reason for Needing to Take Most Recent Leave								
Own illness	466	76	51.5	50.0	−1.5	51.6	49.4	−2.2
Pregnancy or child-related reason	164	30	18.1	19.7	1.6	18.2	19.4	1.2
Other person, non-child reason	274	46	30.3	30.3	0.0	30.2	31.2	1.0
Nature of Health Condition for Most Recent Leave Need								
One-time health matter	202	41	24.9	30.1	5.2	25.0	29.2	4.2
Treatment that now requires routine schedule care	110	20	13.6	14.7	1.1	13.5	14.8	1.2
Ongoing health condition	312	48	38.5	35.3	−3.2	38.3	36.4	−1.9
Eldercare	57	14	7.0	10.3	3.3	7.0	10.7	3.7
Other reason	130	13	16.0	9.6	−6.5	16.2	8.9	−7.3*
Ever Heard of FMLA								
Yes	3,549	428	79.9	78.2	−1.6	79.6	80.7	1.2
No	894	119	20.1	21.8	1.6	20.4	19.3	−1.2
Can Take Paid Leave for Own Illness or Medical Care (Among Those Currently Employed)								
Yes	2,724	336	68.7	66.1	−2.6	68.8	66.0	−2.7
No	864	136	21.8	26.8	5.0	21.9	25.5	3.5
Depends on circumstances	375	36	9.5	7.1	−2.4	9.3	8.5	−0.8
Can Take Paid Leave for Illness or Medical Care of Another Family Member (Among Those Currently Employed)								
Yes	2,187	248	55.6	50.0	−5.6	55.8	48.9	−6.9*
No	1,168	188	29.7	37.9	8.2	29.9	36.2	6.3*
Depends on circumstances	576	60	14.7	12.1	−2.6	14.3	14.9	0.6
Can Take Paid Leave for Routine Childcare Other Than Illness (Among Those Currently Employed								
Yes	1,224	144	31.4	29.8	−1.6	31.6	28.6	−3.0
No	2,014	286	51.7	59.2	7.6	51.9	57.2	5.2*
Depends on circumstances	661	53	17.0	11.0	−6.0	16.5	14.3	−2.2
Can Take Paid Leave for Eldercare (Among Those Currently Employed)								
Yes	1,392	152	35.8	31.2	−4.6	36.1	29.2	−6.9*
No	1,769	270	45.5	55.4	10.0	45.7	53.6	7.9*
Depends on circumstances	731	65	18.8	13.3	−5.4	18.3	17.2	−1.0

Item	Sample Size		Unadjusted Percentage			Adjusted Percentage		
	Employee Survey	NRFU	Employee Survey	NRFU	Diff.	Employee Survey	NRFU	Diff.
Can Take Paid Leave for Errands or Personal Reasons (Among Those Currently Employed)								
Yes	1,095	155	27.8	30.7	2.9	28.1	28.1	0.0
No	2,226	270	56.5	53.5	−3.1	56.7	51.7	−5.0*
Depends on circumstances	616	80	15.6	15.8	0.2	15.2	20.2	5.0*
Can Take Paid Leave for Any Reason Other Than Errand/Personal Reason (Among Those Currently Employed)								
Yes	2,888	365	73.1	71.4	−1.6	73.1	71.0	−2.1
No	757	110	19.2	21.5	2.4	19.2	20.9	1.7
Depends on circumstances	308	36	7.8	7.0	−0.7	7.7	8.1	0.4

Source: 2018 Employee Survey and Employee NRFU Survey
* Indicates difference is statistically significant at the .05 level of significance.
Note: The adjusted percentages displayed in this table were derived from a multinomial logistic model that was fit for each dependent measure in the table. All of the demographics listed in **Exhibit 1.6** were included in the model so the adjusted percentages are showing what the estimates would look like if both the Employee Survey and NRFU respondents were distributed across the demographic variables in exactly the same way. Significance testing was not done for differences in the unadjusted percentages.

Exhibit 1.7 suggests there are some differences between the Employee Survey and NRFU respondents across the key items that were collected in both surveys.[12] Looking at the adjusted percentages, notable findings include these:

Dual Leave Takers/Needers

- NRFU respondents were more likely to be dual leave takers/needers (Employee Survey=8.9 percent, NRFU=14.9 percent).

Leave Taker

- NRFU respondents were not more likely to be a leave taker, but NRFU respondents were more likely to have more than one reason for taking leave in the past 12 months (Employee Survey=33.6 percent, NRFU=48.2 percent).
- NRFU respondents were less likely to take leave for pregnancy or some other child-related reason (Employee Survey=24.4 percent, NRFU=20.2 percent).
- NRFU respondents were more likely to take leave for eldercare (Employee Survey=3.5 percent, NRFU=7.4 percent).
- NRFU respondents were more likely to take leave on more than one occasion (Employee Survey=30.9 percent, NRFU=42.3 percent).

Leave Needer

- NRFU respondents were more likely to be a leave needer (Employee Survey=20.5 percent, NRFU=27.2 percent)
- NRFU respondents were more likely to have more than one reason for needing to take leave (Employee Survey=54.4 percent, NRFU=61.4 percent).

Paid Leave Awareness

- NRFU respondents were more likely to believe that they cannot take paid leave for illness of a family member (Employee Survey=29.9 percent, NRFU=36.2 percent).
- NRFU respondents were more likely to believe that they cannot take paid leave for routine childcare (Employee Survey=51.9 percent, NRFU=57.2 percent).
- NRFU respondents were more likely to believe that they cannot take paid leave for eldercare (Employee Survey=45.7 percent, NRFU=53.6 percent).

So on balance, these results suggest the following:

- Respondents to the NRFU did not necessarily take more leave. Their leaves, however, were more complex than those of the Employee Survey respondents (care for others/elderly, more reasons, and separate occasions during the year).
- Respondents to the NRFU were more likely to need leaves, and to have a leave need that was more complex (more reasons than just one).

[12] There are 34 independent comparisons implicitly made in the table; though no formal control over Type I error, or adjustment for multiplicity testing, was made, we should expect one or two comparisons to be significant just by chance alone.

- Respondents to the NRFU had lower awareness of paid leave policies. This could have affected their decision, or ability, to take leave in the past 12 months; it also could have affected their perception of the survey as salient to them.

Because the set of Employee Survey respondents is the union of two independent samples (RDD, web panel) and data were collected from each sample using a different data collection mode (telephone, web), we thought it would also be interesting to investigate a potential Employee Survey/NRFU interactive effect with telephone/web. The adjusted means for this interaction across the same set of dependent measures are displayed in **Exhibit 1.8**.[13] Some of the interesting findings in this table:

- Employee Survey telephone respondents had a higher percentage of leave takers and a lower percentage of dual leave takers/needers.
- Employee Survey telephone respondents had a lower percentage of leave needers.
- NRFU web respondents had the lowest percentage of employees who can take paid leave for medical care of a family member.
- Employee Survey telephone respondents had a higher percentage of employees who can take paid leave for routine childcare.

Mode differences are further discussed in a separate methodology memorandum.

Statistically significant differences between the estimates generated from the Employee Survey and NRFU samples, such as those noted in the discussion above about **Exhibit 1.7**, suggest there could be some non-response bias in the final estimates of the Employee Survey. The sample weights were designed to correct for some potential bias. This analysis does not shed any light on the magnitude of any potential *residual* non-response bias. An alternative analysis that could be conducted to measure the impact of non-response would be to combine the Employee Survey and NRFU respondents, re-weight the samples, and then compare weighted estimates from the combined Employee Survey/NRFU samples with weighted estimates from the Employee Survey samples.

Exhibit 1.8. Regression-adjusted percentages for several key questionnaire items, by the interaction of Employee Survey/NRFU and mode of data collection (telephone/web)

Item	Adjusted Percentage			
	a. Employee Survey, Telephone	b. Employee Survey, Web	c. NRFU Telephone	d. NRFU Web
Leave Needer/Taker Group				
Leave taker	42.8[bcd]	29.8[a]	25.7[ad]	33.6[ac]
Leave needer	10.3	11.9	13	11.7
Employed only	41.9[bc]	48.4[ad]	49.9[ad]	39.5[bc]
Dual leave taker/needer	5.0[bcd]	10.0[ad]	11.3[a]	15.2[ab]
Leave Taker				
Yes	47.6[bc]	39.7[ad]	37.1[ad]	48.9[bc]
No	52.4[bc]	60.3[ad]	62.9[ad]	51.1[bc]
Leave Needer				
Yes	15.2[bcd]	21.9[ad]	24.4[a]	26.9[ab]

[13] There are 252 independent comparisons implicitly made in the table; though no formal control over Type I error, or adjustment for multiplicity testing, was made, we should expect about 13 comparisons to be significant just by chance alone; the actual number is greater, at 58.

Item	Adjusted Percentage			
	a. Employee Survey, Telephone	b. Employee Survey, Web	c. NRFU Telephone	d. NRFU Web
No	84.8[bcd]	78.1[ad]	75.6[a]	73.1[ab]
Total Reasons/Conditions for Taking Leave in Past 12 Months				
1	64.9[cd]	66.8[cd]	50.0[ab]	52.4[ab]
2 or More	35.1[cd]	33.2[cd]	50.0[ab]	47.6[ab]
Reason For Taking Most Recent Leave				
Own illness	57.8	55.4	55.3	55.3
Pregnancy or child-related reason	22.0	25.0[c]	18.6[b]	20.6
Other person, non-child reason	20.2	19.6	26.1	24.2
Nature of Health Condition for Most Recent Leave				
One-time health matter	47.3[bc]	39.0[a]	29.4[a]	39.3
Treatment that now requires routine schedule care	19.5	16.7	21.7	14.5
Ongoing health condition	21.9	22.7	17.4	25.4
Eldercare	4.3	3.3	9.8	6.6
Other reason	7.0[bcd]	18.3[a]	21.7[a]	14.2[a]
Time Taken Continuously for Most Recent Leave				
One continuous block of time	68.1[d]	69.4[cd]	57.8[b]	57.4[ab]
Separate occasions	31.9[d]	30.6[cd]	42.2[b]	42.6[ab]
What Happened After Recent Leave				
Went back to same employer	91.8	92.6	90.3	89.7
Went to new employer/did not return to work	8.2	7.4	9.7	10.3
Total Reasons/Conditions for Needing to Take Leave in Past 12 Months				
1	40.2	46.5	39	37.4
2 or more	59.8	53.5	61	62.6
Reason for Needing to Take Most Recent Leave				
Own illness	59.0	50.2	53.9	49.2
Pregnancy or child-related reason	13.5	19.2	16.6	19.6
Other person, non-child reason	27.6	30.6	29.6	31.3
Nature of Health Condition for Most Recent Leave Need				
One-time health matter	30.9	23.8	34.8	28.4
Treatment that now requires routine schedule care	13.8	13.5	13.3	15.5
Ongoing health condition	34.1	39.2	29	38.8
Eldercare	6.3	7.1	9.8	10.9
Other reason	14.9	16.4[d]	13.1	6.6[b]
Ever Heard of FMLA				
Yes	78.5	79.8	83.2	79
No	21.5	20.2	16.8	21
Can Take Paid Leave for Own Illness or Medical Care (Among Those Currently Employed)				
Yes	75.8[bd]	67.0[a]	72.8[d]	64.3[ac]
No	22.3	21.8	25.4	25.6
Depends on circumstances	1.8[bd]	11.2[ac]	1.8[bd]	10.1[ac]
Can Take Paid Leave for Illness or Medical Care of Another Family Member (Among Those Currently Employed)				
Yes	66.3[bd]	53.3[acd]	61.5[bd]	45.2[abc]
No	29.7	29.8[cd]	36.5[b]	36.0[b]
Depends on circumstances	4.0[bd]	16.8[ac]	2.0[bd]	18.9[ac]

Item	Adjusted Percentage			
	a. Employee Survey, Telephone	b. Employee Survey, Web	c. NRFU Telephone	d. NRFU Web
Can Take Paid Leave for Routine Childcare Other Than Illness (Among Those Currently Employed)				
Yes	39.9[bcd]	29.9[a]	30.2[a]	30.2[a]
No	56.6[bc]	50.8[ac]	66.2[abd]	53.5[c]
Depends on circumstances	3.5[bd]	19.3[ac]	3.6[bd]	16.3[ac]
Can Take Paid Leave for Eldercare (Among Those Currently Employed)				
Yes	50.8[bcd]	33.1[a]	39.5[ad]	28.1[ac]
No	44.3[cd]	45.9[cd]	55.7[ab]	52.2[ab]
Depends on circumstances	4.9[bd]	21.0[ac]	4.8[bd]	19.7[ac]
Can Take Paid Leave for Errands or Personal Reasons (Among Those Currently Employed)				
Yes	37.4[bd]	26.0[ac]	37.7[bd]	26.4[ac]
No	58.7[d]	56.1[d]	57.3	49.6[ab]
Depends on circumstances	3.8[bd]	17.9[acd]	5.0[bd]	24.0[abc]
Can Take Paid Leave for Any Reason Other Than Errand/Personal Reason (Among Those Currently Employed)				
Yes	79.7[bd]	71.4[ac]	79.8[bd]	68.0[ac]
No	19.1	19.2	19.8	21.6
Depends on circumstances	1.2[bd]	9.4[ac]	0.5[bd]	10.4[ac]

Source: 2018 Employee Survey and 2018 NRFU Survey

[a] Indicates difference between estimate and the estimate in column a is statistically significant at the .05 level of significance.
[b] Indicates difference between estimate and the estimate in column b is statistically significant at the .05 level of significance.
[c] Indicates difference between estimate and the estimate in column c is statistically significant at the .05 level of significance.
[d] Indicates difference between estimate and the estimate in column d is statistically significant at the .05 level of significance.

Note: The adjusted percentages displayed in this table were derived from a multinomial logistic model that was fit for each dependent measure in the table. All of the demographics listed in **Exhibit 1.6** were included in the model so the adjusted percentages are showing what the estimates would look like if each of the four groups were distributed across the demographic variables in exactly the same way. The four groups refer to:
Employee Survey telephone respondents
Employee Survey web respondents
NRFU telephone respondents
NRFU web respondents

(2) Comparison of easier-to-reach versus harder-to-reach respondents

The second technique used to assess the risk of non-response 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 non-respondents who never completed the extended interview. If the harder-to-reach respondents do not differ from the easy-to-reach ones, then it seems reasonable that the sample members never reached would also not differ from those interviewed. Support for this "continuum of resistance" model is inconsistent (Lin & Schaeffer, 1995; Montaquila et al., 2008), but it can still be a useful framework for assessing the relationship between level of effort and non-response bias. Groves (2006) has argued that its utility lies more in understanding the way data quality changes during the data collection process rather than providing insight into non-response bias. Despite its limitations, analyzing level of effort is a standard approach to evaluate non-response bias (Halbesleben & Whitman, 2013; Maitland et al., 2017; Montaquila & Olson, 2012; McFarlane et al., 2007).

For respondents reached in the RDD telephone mode, the level of effort is defined by the number of calls required to complete the interview. The number of call attempts required to complete an interview ranged from one to 16, with a median of three. About one in 29 respondents (3.4 percent) was a converted refusal. However, due to the small number of respondents who were converted refusals (n=25), this dimension was not taken into account in the analysis. **Exhibit 1.9** shows how easy-to-reach and harder-to-reach telephone respondents compare with respect to FMLA group, employer coverage, employee eligibility for FMLA, and familiarity with FMLA. The first set of columns defines harder-to-reach respondents as those who completed the survey after the third call attempt; the second set of columns restricts harder-to-reach respondents to those who completed after the fourth call attempt. On all four of these measures, we find no statistically significant differences between easy-to-reach respondents and harder-to-reach respondents.

Exhibit 1.9. Leave-related characteristics of telephone respondents, by level of effort groups

Characteristic	Easy to Reach 1 2 attempts	Harder to Reach 3+ attempts	Easy to Reach 1 3 attempts	Harder to Reach 4+ attempts
	%	%	%	%
FMLA Group				
Leave taker	45.9	42.3	43.7	44.4
Leave needer	17.2	17.5	19.8	13.3
Employed only	36.9	40.3	36.5	42.3
Employer Is...				
Not covered by FMLA	22.8	24.0	22.8	24.5
Covered by FMLA	77.2	76.0	77.2	75.5
Employee Is...				
Not eligible for FMLA	46.4	50.3	47.3	50.4
Eligible for FMLA	53.6	49.7	52.7	49.6
Heard of FMLA				
Yes	73.6	69.5	73.0	69.5
No	26.4	30.5	27.5	30.4
Minimum sample size	**308**	**346**	**408**	**246**

Source: 2018 Employee Survey, figures are unweighted
[a] Indicates that the chi-square test for the difference between 1-2 versus 3+ attempts is statistically significant at the .05 level of significance.
[b] Indicates that the chi-square test for the difference between 1-3 versus 4+ attempts is statistically significant at the .05 level of significance.

For the 3,731 extended interview respondents from the web mode, the level of difficulty in reaching the respondent is considered with respect to the number of email reminders sent prior to survey completion. Half of respondents (50.9 percent) completed the web survey prior to receiving an email reminder. An additional 6 percent completed after receiving the first reminder, and another 15 percent completed after receiving the second reminder.

Exhibit 1.10 below presents leave-related characteristics of web respondents based on two definitions of level of effort. In the first set of columns, harder-to-reach web respondents are defined as those who completed the survey after the second email reminder. These harder-to-reach respondents are significantly less likely to report being employed only, compared to those who completed before the second reminder (45.2 percent versus 52.5 percent, chi-square p = <.0001). Using this definition of level of effort, no statistically significant differences were found between level of effort and employer coverage, employee eligibility for FMLA, or whether the respondent had heard of FMLA.

Exhibit 1.10. Leave-related characteristics of web respondents, by level of effort groups

	Easy-to-Reach 0-1 reminders	Harder–to-Reach 2+ reminders	Easy-to-Reach 0-2 reminders	Harder-to-Reach 3+ reminders
	%	%	%	%
FMLA Group [a]				
Leave taker	27.5	30.1	29.0	27.7
Leave needer	20.0	24.7	21.1	24.3
Employed only	52.5	45.2	49. 9	48.1
Employer Is...				
Not covered by FMLA	23.8	22.5	23.8	21.8
Covered by FMLA	76.2	77.5	76.2	78.2
Employee Is...				
Not eligible for FMLA	41.5	42.7	41.9	42.2
Eligible for FMLA	58.5	57.3	58.1	57.8
Heard of FMLA [b]				
Yes	82.3	80.6	82.4	79.5
No	17.8	19.4	17.7	20.5
Minimum sample size	**1980**	**1491**	**2,490**	**981**

Source: 2018 Employee Survey, figures are unweighted
[a] Indicates that the chi-square test for the difference between 0-1 versus 2+ reminders is statistically significant at the .05 level of significance.
[b] Indicates that the chi-square test for the difference between 0-2 versus 3+ reminders is statistically significant at the .05 level of significance.

The significant difference in FMLA group membership by level of effort disappears when the harder-to-reach category is restricted to those who completed after the third reminder, as shown in the second set of columns. Upon closer examination, the employed only group is significantly lower among those who completed the web survey after the first reminder and before the third reminder (39.9 percent), compared to those who completed the survey before receiving a reminder (54.0 percent) or after the third reminder (48.1 percent) (chi-square p = <.0001). Web respondents who completed the survey after the third reminder are slightly less likely to have heard of FMLA, compared to those who completed earlier (79.5 percent versus 82.4 percent, chi-square p = .0446). No statistically significant relationship was found between level of effort and employer coverage and employee eligibility. Because the differences in FMLA group membership are associated with a specific level of effort rather than with a higher level of effort, it does not suggest the potential for non-response bias. The negligible differences observed for these other measures suggest that other survey variables are likely to be unrelated to this level of effort dimension.

(3) Response propensity modeling

Response propensity modeling is an integral part of both weighting procedures and non-response analysis. These statistical models have a 0/1 indicator of the response (where 1 stands for response; i.e., completed interview) and incorporate demographic variables as predictors to help identify the characteristics that are associated with higher or lower response propensities. It should be noted, however, that such analyses are limited to the variables that are available for both respondents and non-respondents.

RDD sample

For the RDD sample, no information is available upfront in the sample, short of geography represented by the phone number and the prepaid flag used for oversampling. Hence, propensity modeling is limited to the later stage of the response process, after the initial screening is completed, which included questions on gender, age, and education. Thus, this analysis compares 739 respondents who completed the FMLA survey on the phone versus 2,482 who completed only the screener (up to question S12, where question S12 is the last item in the screener, after which the determination of the HHFLG variable is calculated). Among the respondents who qualified for the extended interview (i.e., leave takers, leave needers, and a subsample of employed only), the best fitting response propensity logistic regression model was an interaction of frame (cell phone versus landline) with those demographic variables. (Higher-order interactions were also explored, such as frame by race by gender, but were found to be insignificant.)

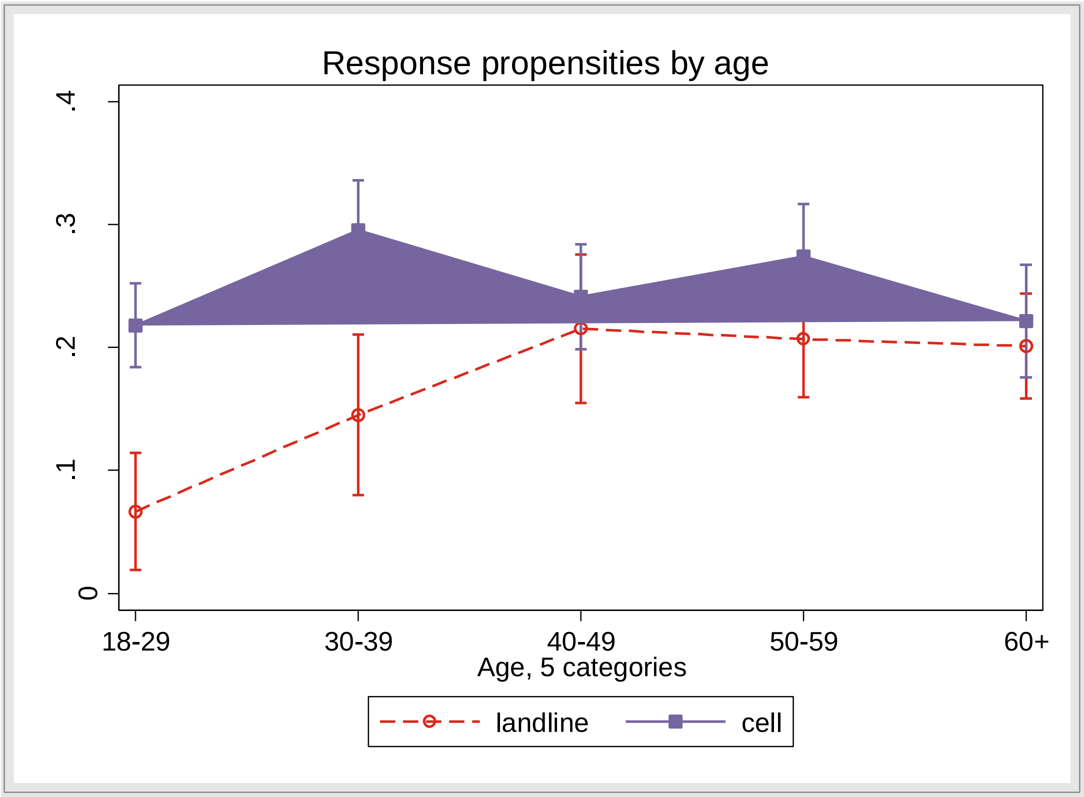
Exhibit 1.11 shows the logistic regression coefficient estimates for RDD response propensities by age, education, and gender, for the landline and cell frames, respectively. The table reports the estimates of the parameters of the logistic regression, with standard errors. These estimates are used to produce the marginal predictions of response propensities. The table makes it clear that response propensities have generally been higher on the cell frame; that the younger respondents were far less willing to respond via the landline; and though there were no differences between response rates by gender within the landline frame, women responded at higher rates than men did on cell phones. As this propensity model was used to define one of the intermediate weighting factors, those response patterns have been fully incorporated into the RDD weights. Graphical representations of these propensities are shown in **Exhibit 1.12** in absolute units (i.e., marginal response propensities).

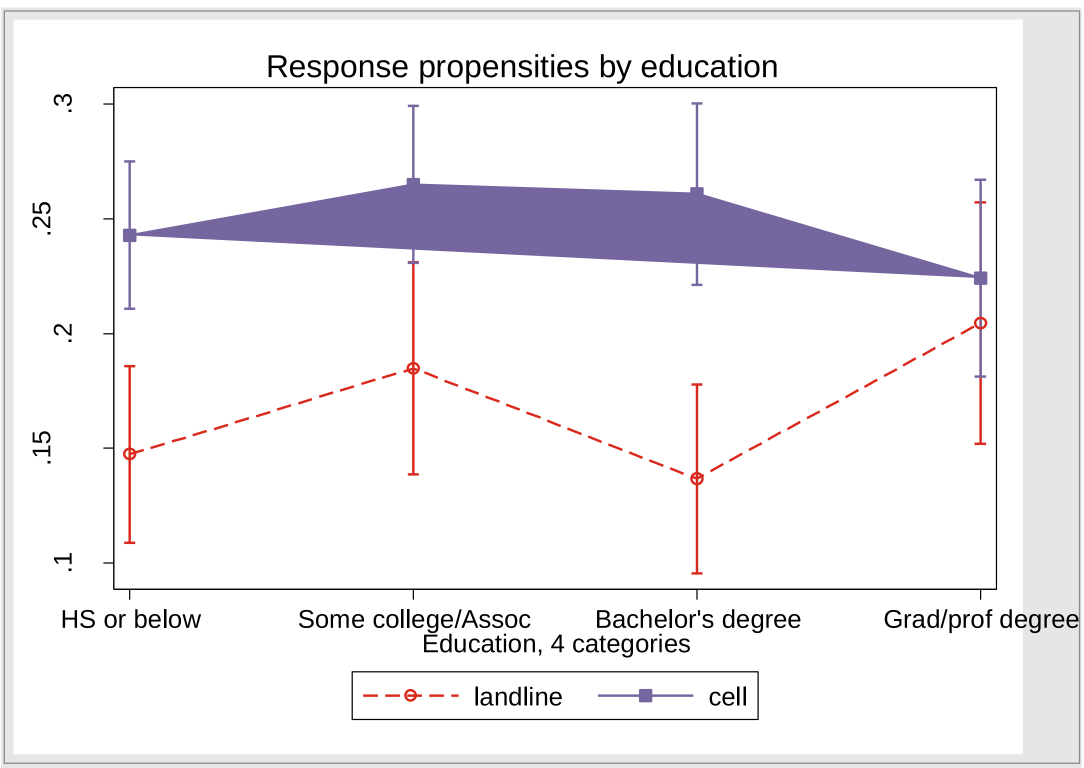
Exhibit 1.11. Logistic regression estimates of RDD response propensities, by demographic characteristics, for landline and cell frames

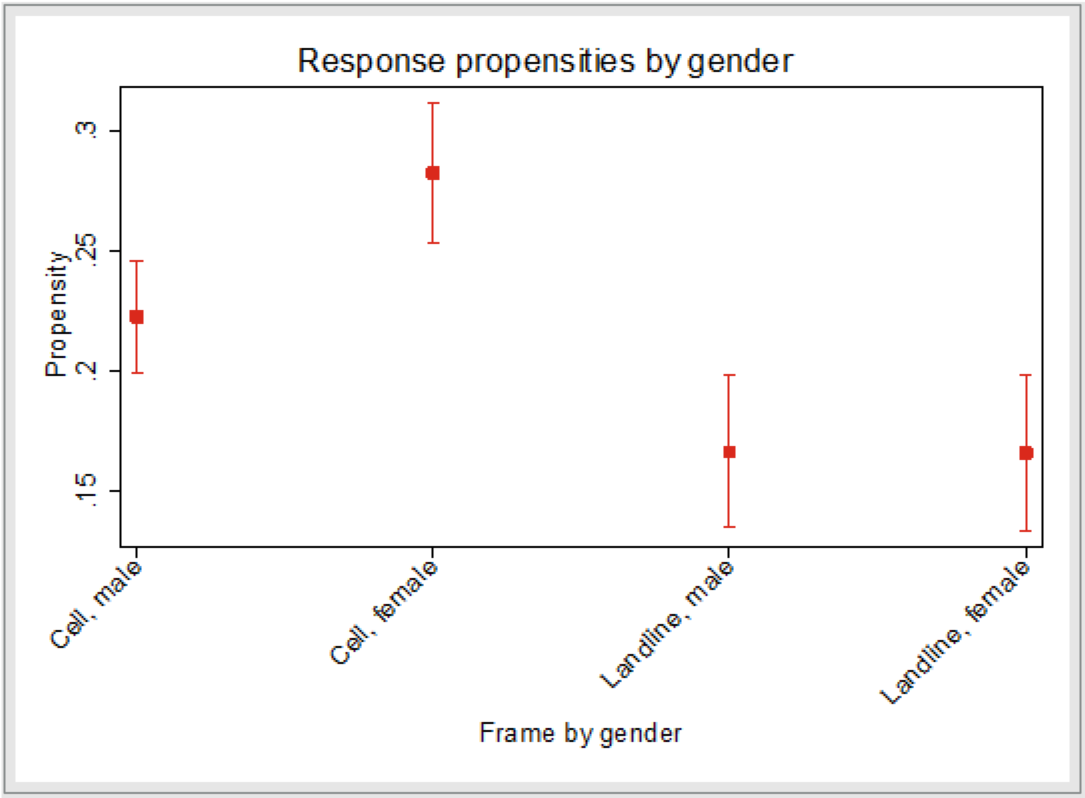
Characteristic	Landline Frame		Cell Frame	
	Estimate	Std Error	Estimate	Std Error
Age				
18-29	Reference group		Reference group	
30-39	0.869	0.477	1.535	0.458***
40-49	1.352	0.435**	1.255	0.459**
50-59	1.301	0.420**	1.429	0.461**
60+	1.266	0.416**	1.144	0.467*
Education				
High school or below	−0.408	0.227	0.106	0.156
Some college/associates	−0.129	0.226	0.224	0.156
Bachelor’s degree	−0.496	0.243*	0.202	0.165
Graduate/professional degree	Reference group		Reference group	
Gender				
Male	Reference group		Reference group	
Female	−0.003	0.163	0.320	0.100**
Intercept	−2.376	0.429***		

Source: Abt calculations based on the original data
****p*<.001; ***p*<.01; **p*<.05. Based on 3,221 participants who completed the screener treating 739 eventual respondents as the “success” outcome in logistic regression.

Exhibit 1.12. Graphical representations of RDD response propensities by age, education, and gender, for landline and cell frames







Source: Abt calculations based on the original data
Note: Based on 3,221 participants who completed the screener treating 739 eventual respondents as the "success" outcome in logistic regression.

KnowledgePanel sample

The data set provided by KnowledgePanel contains cases who completed the interview, were screened out as ineligible (e.g., did not work for pay in the past 12 months, were self-employed), or were randomly terminated (subsample of employed only). Indirect information on the response process, combining all of the panel recruitment, panel attrition, and response to this particular survey, were obtained from the distribution of the base weights. Higher weights are associated with the groups that have lower propensities to respond.

Exhibit 1.13 reports the summary statistics of the KnowledgePanel base weights, where the weights were standardized to have a mean of 1 for easier comparisons. Within each variable, the category with the largest mean weight is bolded, indicating a demographic group that was hardest to reach. The group with the highest unequal weighting design effect is also bolded (compared to the overall design effect of 2.54), indicating the greatest variability of response propensities with that group.[14,15]

[14] According to the motivation of a modern measure of sample representativeness, the *R*-indicator, higher variability of response propensities may indicate greater risk of differential nonresponse bias (Schouten, Cobben, & Bethlehem, 2009).

[15] Small cells with fewer than 50 observations are removed.

Exhibit 1.13. Summary statistics of KnowledgePanel base weights

Variable	Category	*N*	Minimum Relative Weight	Mean Relative Weight	Maximum Relative Weight	Unequal Weighting Design Effect
Overall		26,140	0.045	1.000	33.844	2.542
Geography	Northeast	2,368	0.098	0.942	9.275	1.851
	South	5,804	0.105	0.944	33.844	2.112
	Midwest	8,562	0.120	**1.142**	28.447	2.391
	West	3,093	0.073	1.021	17.708	2.217
	CA+NJ+RI	4,686	0.045	0.832	33.318	3.688
	NY	1,627	0.082	0.982	29.347	**3.716**
Income	Under $35,000	8,388	0.045	0.690	18.814	**2.543**
	$35,000 to $100,000	10,900	0.073	1.047	20.040	2.261
	$100,000+	6,852	0.197	**1.306**	33.844	2.508
Education	High school or below	5,775	0.188	**1.797**	33.844	**2.206**
	Some college/associates	8,948	0.063	0.876	33.318	2.156
	Bachelor's degree	6,658	0.045	0.679	10.094	1.933
	Graduate/professional degree	4,759	0.047	0.715	18.973	1.880
Age	18-29	1,872	0.271	**2.856**	33.844	**2.158**
	30-39	3,372	0.070	1.280	12.112	1.734
	40-49	3,642	0.051	1.132	23.260	1.767
	50-59	5,118	0.047	0.965	10.629	1.677
	60+	12,136	0.045	0.611	14.958	1.693
Work Status	Paid employee	12,475	0.047	1.173	33.318	2.280
	Self-employed	2,312	0.047	0.887	28.447	**3.134**
	On temporary leave	84	0.073	1.186	12.112	2.861
	Looking for job	1,001	0.051	**1.374**	33.844	2.987
	Not working - retired	7,929	0.063	0.622	8.493	1.714
	Not working - disabled	1,028	0.045	1.049	18.814	2.466
	Not working - other	1,311	0.051	1.507	30.296	2.559
# of Household Members Aged 18+	1	7,900	0.047	0.751	28.447	2.491
	2	13,339	0.045	1.010	30.296	2.068
	3	3,171	0.050	1.304	33.318	2.812
	4	1,283	0.053	1.461	33.844	3.050
	5	316	0.051	1.503	16.902	2.476
	6	90	0.086	**1.997**	29.347	**4.461**
# Adolescents Aged 13-17	0	23,258	0.045	0.958	33.318	2.445
	1	2,184	0.053	1.317	24.482	2.403
	2	614	0.051	**1.408**	33.844	**3.834**
	3	76	0.182	1.353	9.275	1.924
# Children Aged 6-12	0	22,990	0.045	0.956	33.844	**2.646**
	1	2,076	0.070	**1.326**	18.619	2.095
	2	884	0.086	1.314	9.509	1.695
	3	163	0.050	1.139	6.754	1.722
# Children	0	24,277	0.045	0.970	33.844	**2.631**

Variable	Category	*N*	Minimum Relative Weight	Mean Relative Weight	Maximum Relative Weight	Unequal Weighting Design Effect
Aged 2-5	1	1,496	0.076	**1.398**	13.455	1.769
	2	330	0.102	1.374	9.750	1.742
# Children Aged 0-1	0	25,213	0.045	0.983	33.844	**2.572**
	1	789	0.098	**1.511**	18.619	1.910
Housing	Owned or being bought	19,583	0.045	0.967	33.844	2.497
	Rented for cash	6,120	0.047	1.095	30.296	2.517
	Occupied without payment	437	0.073	**1.141**	28.447	**3.917**
MSA Status	Non-Metro	3,332	0.073	0.961	13.716	2.146
	Metro	22,808	0.045	**1.006**	33.844	**2.594**
Marital Status	Married	14,574	0.045	1.006	28.447	1.918
	Widowed	1,972	0.073	0.592	9.644	1.984
	Divorced	3,838	0.047	0.665	9.509	2.022
	Separated	502	0.050	0.950	7.561	1.833
	Never married	4,148	0.050	1.341	33.844	**3.483**
	Living with partner	1,106	0.063	**1.556**	29.347	2.444
Household Size	1	7,194	0.047	0.714	28.447	2.484
	2	10,017	0.045	0.904	23.260	2.120
	3	3,649	0.053	1.231	33.318	2.389
	4	2,964	0.053	1.367	30.296	2.426
	5	1,310	0.095	1.446	22.763	2.401
	6	544	0.050	**1.582**	33.844	**3.439**
Gender	Male	11,391	0.053	**1.108**	33.844	**2.915**
	Female	14,749	0.045	0.916	22.763	2.071
Race and Ethnicity	White, Non-Hispanic	19,011	0.053	0.877	25.979	2.180
	African American, Non-Hispanic	2,091	0.132	1.464	29.347	2.128
	Other, Non-Hispanic	769	0.217	**2.308**	33.844	2.550
	Hispanic	3,693	0.050	1.134	28.447	2.685
	Multi Races, Non-Hispanic	576	0.045	0.787	30.296	**5.267**

Source: Abt calculations based on the original data

The weight summaries presented in **Exhibit 1.13** demonstrate for the KnowledgePanel sample the following:

- Employees with higher incomes were less likely to respond.
- Employees with education of high school or below were substantially less likely to respond.
- Younger adults were substantially less likely to respond.
- Sampled panel members from larger households and households with children were less likely to respond.
- Never-married employees and employees living with partners were less likely to respond.
- Males were slightly less likely to respond, and exhibited greater variability in response propensities than females.

- Some racial/ethnic minorities had lower response propensity, especially Other, Non-Hispanic (i.e., not white or African American). On the other hand, multiracial employees had the greatest design effects / variability of response propensities.

Note that some of these patterns are similar to those faced in the RDD sample analysis presented above (e.g., lower response rates of young adults, higher response rates of women).

(4) Comparisons with external benchmarks

The benchmark analysis compares the results of the Employee Survey with results from an external survey, to gauge a degree of mismatch that would at least partially be related to the difference in response rates. Specifically, we compare the weighted final respondent estimates from the Employee Survey versus those based on the Current Population Survey (CPS)'s Annual Social and Economic Supplement (ASEC, or March CPS). It provides detailed information on employment, workplace, and income, and it is the most appropriate data set to use. The CPS is considered a "gold standard" survey due to its rigorous protocol (e.g., area-probability sampling with in-person interviewing). By virtue of its more rigorous design, the estimates from the March 2018 CPS are assumed to contain less non-response bias than estimates from the 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 2018 Employee Survey. One weakness of this approach is that the set of variables collected in both the Employee Survey and the CPS is very limited, and nearly all of them are already used in weight calibration (so the two surveys agree on them perfectly by the study design).

Another weakness is that the measurements collected in the 2018 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, whereas the Employee Survey used phone and web modes. Furthermore, the question wording for the comparison questions varies 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 non-sampling errors differ between the Employee Survey and CPS due to very different designs and implementation of the two surveys (including non-response error, rotation biases in CPS, recall biases in the Employee Survey, etc.).

We computed CPS-weighted estimates based on the population of adults aged 18 and older who were employed for pay within the past 12 months (excluding self-employed). This matches the target population of the Employee Survey. (Note that this is the CPS subpopulation used in weighting the Employee Survey data.) The benchmark analysis compared responses for union membership, employer size, and hours worked.

Exhibit 1.14. Distribution of Employee Survey and CPS samples, by demographics

Characteristic	Employee Survey, unweighted	Employee Survey, pre weighted	Employee Survey, weighted	CPS, weighted
Total	100.0	100	100.0	100.0
N	4,470	4,470	4,470	75,529
Age Group				
18-29	11.8	22.7	24.5	24.5
30-39	20.7	20.9	22.7	22.7
40-49	18.2	18.1	20.6	20.6
50-59	24.6	23.7	19.7	19.7
60+	24.7	14.6	12.5	12.5

Characteristic	Employee Survey, unweighted	Employee Survey, pre weighted	Employee Survey, weighted	CPS, weighted
Gender				
Male	44.3	52.1	52.2	52.2
Female	55.7	47.9	47.8	47.8
Education				
High school or below	17.4	27.9	33.4	33.4
Some college/Associates	33.7	32.9	28.9	28.9
Bachelor's degree	29.1	23.4	24.0	24.0
Graduate/professional degree	19.8	15.8	13.7	13.7
Number of Children				
0	63.4		62.9	59.8
1 or 2	27.9		29.0	32.5
3 or more	8.7		8.0	37.7
Marital Status				
Not married	43.4	44.7	48.6	48.6
Married or live with partner	56.6	55.3	51.4	51.4
Household Income				
Under $35,000	25.0	13.8	11.4	11.4
$35,000 to $100,000	46.1	44.4	44.0	44.0
$100,000+	28.8	41.7	44.5	44.5
Region				
Northeast (except RI, NY, NJ)	7.6	8.8	8.5	8.5
South	30.2	23.1	21.8	21.8
Midwest	22.3	35.2	36.8	36.8
West (except CA)	12.3	12.0	11.7	11.7
California, New Jersey, Rhode Island	21.0	14.8	15.1	15.1
New York	6.6	6.1	6.1	6.1
Job Type				
Government	20.6		17.7	15.9
Private (including non-profit)	64.2		82.3	84.1

Source: Abt calculations based on the original FMLA data and publicly available CPS data

The "pre-weighted" distribution is the distribution with an intermediate weight that is the input to the ultimate weight calibration step. It is a combination of the calibrated weight of the RDD sample and the calibrated weight of the KnowledgePanel sample, with the compositing factor proportional to their respective effective sample sizes. (See Section 1.6. Weighting.)

Union membership

As shown in **Exhibit 1.15**, Employee Survey respondents have somewhat higher rates of unionization than CPS respondents.[16,17] (Note that the answer choice of being covered by a union without being a member is not offered in the Employee Survey.)

Exhibit 1.15. Union participation (CPS and Employee Survey)

Source	Status Category	Estimate %	Std. Error %	Confidence Interval %
CPS	No union coverage	88.3	0.4	87.6, 89.1
	Member of labor union	10.4	0.4	9.7, 11.1
	Covered by union but not a member	1.3	0.1	1.1, 1.6
Employee Survey	Not in union	85.9	1.0	84.1, 87.8
	In union	13.4	0.9	11.6, 15.2
	Missing	0.7	0.2	0.3, 1.1

Source: Abt calculations based on the original FMLA data and publicly available CPS data

Employer size

The CPS and the Employee Survey both ask about the size of the employer, but their concepts are different. The CPS asks to report the total employment for all locations,[18] whereas the Employee Survey asks for employer size information for sites within 75 miles.[19] Also the two surveys have different employer size breakdowns. The CPS breaks down the top sizes into two categories (500-999 and 1,000+ employees), whereas the Employee Survey uses only one (500+). In the mid-sizes, the single CPS category (100-499 employees) corresponds to two Employee Survey categories (100-249 and 250-499). For the smallest sizes, the Employee Survey breaks are 10-19, 20-29, 30-39, and 40-49 employees, whereas the CPS breaks are 10-24 and 25-49.[20]

Though the counts in low categories (1-9 and 10-49) generally match, as they may be picking up single-site firms, the estimates diverge at or above the employment size of 50, where firms are more likely to have multiple sites. **Exhibit 1.16** below shows the employer size information for the CPS and the Employee Survey.

---

16 CPS asks about union membership only in two out of eight rotation groups, so the sample sizes are reduced and standard errors are increased: https://cps.ipums.org/cps-action/variables/UNION.

17 The confidence intervals of the unionization estimates overlap just slightly.

18 https://cps.ipums.org/cps-action/variables/FIRMSIZE

19 The Employee Survey has some item non-response; the CPS figures are imputed by the data providers (Bureau of Labor Statistics).

20 If respondents have to use estimation strategies rather than being able to obtain an exact figure from a reputable source, distributions of answers are known to depend on the breaks (Smyth, Dillman, & Christian, 2009). Put differently, respondents derive information from the response categories provided, and they do not like to gravitate towards the extremes of the scales.

Exhibit 1.16. Employer size (CPS and Employee Survey)

Source	Employment Category	Estimate %	Std. Error %	Confidence Interval %
CPS	1 to 9	12	0.4	11.2, 12.7
	10 to 49	15.1	0.4	14.3, 15.8
	50 to 99	7.7	0.3	7, 8.3
	100 to 499	13.2	0.4	12.4, 13.9
	500+	52.1	0.5	51.1, 53.2
Employee Survey	1 to 9	10.4	1.3	7.8, 12.9
	10 to 49	14.4	1.0	12.5, 16.4
	50 to 99	13.6	1.2	11.2, 16
	100 to 499	27.1	1.4	24.3, 29.9
	500+	32.3	1.4	29.6, 35
	Unknown	2.1	0.5	1.1, 3.1

Source: Abt calculations based on the original FMLA data and publicly available CPS data

Hours worked

Both the CPS and the Employee Survey ask for hours worked, but in a different way. The CPS version of the question does not have a specified time frame and is asked for all jobs.[21] The Employee Survey asks about employment for the past 12 months, so it may contain a larger recall error. Also for employees with multiple jobs, there are separate questions about the main job and the total across all jobs.

As shown in **Exhibit 1.17**, CPS respondents are more likely to be employed 35 to 40 hours than are their Employee Survey counterparts. Conversely, Employee Survey respondents are more likely to be employed in the lower and higher ends of hours worked (0-34 hours and 40+ hours). The N/A rows correspond to respondents who do not currently work although they had a job in the past 12 months (they cannot be referred to as "unemployed" as we do not know if they are looking for a job); the estimates for this group are consistent between the two surveys.

Exhibit 1.17. Hours worked (CPS and FMLA Employee Survey)

Source	Category	Estimate %	Std. Error %	Confidence Interval %
CPS	0 to 34	14.4	0.4	13.6, 15.2
	35 to 40	60.5	0.6	59.4, 61.7
	40+	20.2	0.5	19.3, 21.2
	N/A	4.8	0.3	4.3, 5.3
Employee Survey	0 to 34	18.8	1.3	16.2, 21.3
	35 to 40	48.5	1.6	45.5, 51.6
	40+	27.4	1.4	24.5, 30.2
	N/A	5.3	0.8	3.7, 6.9

Source: Abt calculations based on the original FMLA data and publicly available CPS data

[21] https://cps.ipums.org/cps-action/variables/UHRSWORKT

1.5.4 Summary of Non-Response Analysis for the 2018 Employee Survey

In summary:

- NRFU respondents are more likely to be aged 30-39, have children, be married, have income under $35,000, and work for private firms. These differences are largely immaterial, as they are corrected by weighting.
- Respondents to the NRFU did not necessarily take more leave. Their leaves, however, were more complex than those of the Employee Survey respondents (care for others/elderly, more reasons, and separate occasions during the year).
- Respondents to the NRFU were more likely to need leaves and to have a leave need that was more complex (more reasons than just one).
- Respondents to the NRFU had lower awareness of paid leave policies. This could have affected their decision to take a leave in the past 12 months and their perception of the survey as salient to them.

Comparisons of the easier-to-reach versus harder-to-reach respondents did not show notable differences.

Modeling of response propensities on the RDD sample demonstrated lower response propensity for younger adults on landline, and higher response propensity for females. Analysis of weights (implicit response propensities) in the KnowledgePanel demonstrated the following:

- Employees with higher incomes were less likely to respond.
- Employees with education of high school or below were substantially less likely to respond.
- Younger adults were substantially less likely to respond.
- Sampled panel members from larger households and households with children were less likely to respond.
- Never- married employees and employees living with partners were less likely to respond.
- Males were slightly less likely to respond, and exhibited greater variability in response propensities than females.
- Some racial/ethnic minorities had lower response propensity, especially Other, Non-Hispanic other single race (i.e., not white or black/African American). On the other hand, multiracial employees had the greatest design effects / variability of response propensities.

The impact of these differences is ameliorated through weighting procedures.

Comparison to external benchmarks, though complicated owing to the questions being asked differently, revealed higher unionization rates and lower estimates of full time employment (35 to 40 weeks hours) on the Employee Survey compared to CPS.

Overall, it appears that the Employee Survey had greater difficulties reaching respondents in more dire circumstances and with greater demands on their time—families with children, racial and ethnic minorities, people who needed multiple leaves. Lacking any external data, it is impossible to gauge what the impact of non-response is on the survey estimates. To the extent that the variables of interest (e.g., the number and reasons for leaves) are strongly associated with variables used in weighting (in particular, with age and marital status), weighting would reduce non-response biases.

1.6. Weighting

The weighting process for the Employee Survey involved several steps. First we developed weights for the RDD and web panel samples separately. We then combined the weights for the two samples into an integrated set of analysis weights. Last, we created a set of bootstrap replicate weights to facilitate correct variance estimation.

1.6.1 Target Population

The target population of the 2018 Employee Survey was adults aged 18 or older who live in the United States and have been employed for pay (private or public sector) in the 12 months prior to the interview.

Population control totals

The 2018 CPS ASEC is the most appropriate source of control totals. Those control totals include detailed information on age, education, race/ethnicity, marital status, employment, workplace, geography, and income. This data set was subset to those aged 18+ employed for wage/salary in private or public sector (excluding self-employed) in 2018.

The CPS ASEC data set contains 75,529 observations. The estimated population size (sum of weights) is 140,590,319. **Exhibit 1.18** shows the relevant population totals.

Exhibit 1.18. Population control totals

Group	Total	Std. Error
Age 18-29	34,445,192	215,869.0
Age 30-39	31,871,845	158,464.9
Age 40-49	28,982,879	160,926.0
Age 50-59	27,682,668	192,849.8
Age 60+	17,607,735	185,817.5
Education high school or below	46,929,536	369,550.7
Education some college/associates degree	40,661,366	331,625.0
Education bachelor's degree	33,747,103	362,036.9
Education graduate/professional degree	19,252,314	245,484.7
Race/ethnicity Non-Hispanic white	87,236,589	324,722.3
Race/ethnicity Non-Hispanic African American	16,937,544	136,682.3
Race/ethnicity Non-Hispanic Other/Mixed	12,029,106	116,575.5
Race/ethnicity Hispanic	24,387,079	164,243.3
Not married	68,330,488	418,852.3
Married	72,259,831	456,797.4
Region Northeast	11,976,407	144,606.5
Region South	30,601,200	189,292.2
Region Midwest	51,780,178	304,478.3
Region West	16,394,462	144,429.0
Paid leave states CA+NJ+RI	21,231,701	180,136.2
Paid leave state NY	8,606,371	134,520.1
Income under $35,000	16,069,349	198,730.3
Income $35,000 to $100,000	61,888,768	439,954.7
Income $100,000	62,632,201	516,505.5

Source: Abt calculations based on the publicly available CPS data

For the RDD data collection, control totals additionally included phone service. We obtained the phone use population control totals from the 2017 National Health Interview Survey (NHIS) (**Exhibit 1.19**). This subset of the data contains 37,950 person-level observations obtained as follows. First, the household-level information on phone use was unambiguously determined (i.e., omitting households with missing data on phone use questions, unweighted 1.5 percent of households; additional 2.9 percent of households did not have phone service). Second, a person aged 18+ reported having worked in the past year (note that NHIS does not break this population down to self-employed versus employed for pay).

Exhibit 1.19. Population control totals

Group	Proportion	SE
Landline only	0.029117	0.0015
Cell phone only	0.603501	0.0058
Dual use	0.367382	0.0057

Source: Abt calculations based on the publicly available NHIS data

1.6.2 RDD Survey Data

Base weights

We computed the base weights for the phone mode of the Employee Survey as the ratio of the number of phone lines in the universe, as reported by the sample provider, to the number of phone lines used in the survey. **Exhibit 1.20** shows the base RDD weights by telephone sample type. For the set of completed interviews, the unequal weighting design effect due to base weights is 1.098.

Exhibit 1.20. Base weights, by phone stratum

Phone Stratum	Frame Count	Numbers Released	Base RDD Weight
Landline oversample	41,526,100	21,231	1,955.918
Landline rest of country	245,929,200	57,118	4,305.634
Cell phone prepaid oversample	11,092,763	2,835	3,912.791
Cell phone prepaid rest of country	54,051,978	17,101	3,160.750
Cell phone non-prepaid oversample	64,488,737	17,524	3,680.024
Cell phone non-prepaid rest of country	377,555,622	58,026	6,506.663

Source: Frame counts were provided by sample provider Marketing Systems Group (MSG) in the sample reports; Abt production figures and internal calculations

Frame adjustments

In computing the sample weights, we performed a number of intermediate adjustments:

(1) ***Correction for working/residential numbers:*** Because many numbers, especially in the landline frame, are non-working numbers, the first adjustment is to correct for these. This adjustment brings the weights in line with the population of *working* numbers (versus all U.S. phones), and corrects for biases between frames associated with the different working number rates. **Exhibit 1.21** shows the RDD weights after this adjustment.

Exhibit 1.21. Weight adjustment for working/residential numbers, by phone stratum

Stratum	Base Weight	Working Numbers Adjustment	Weight After Adjustment
Landline - oversample	1,955.918	4.423125	8,651.271
Landline - rest of country	4,305.634	2.209594	9,513.702
Cell phone prepaid - oversample	3,912.791	1.211572	4,740.627
Cell phone prepaid - rest of country	3,160.750	1.204173	3,806.090
Cell phone not prepaid - oversample	3,680.024	1.366112	5,027.323
Cell phone not prepaid - rest of country	6,506.663	1.393211	9,065.157

Source: frame counts were provided by sample provider Marketing Systems Group (MSG) in the sample reports; Abt production figures and internal calculations.

The design effect after this correction becomes 1.092—that is, goes *down* because the very low landline weights are brought up to be more in line with the cell phone weights.

(2) ***Correction for multiple phones:*** Survey respondents who report having multiple phones have higher probabilities of selection, so we divided the weights by the number of phones (in the frame the respondent was reached at). We made adjustments for multiple cell phones (question T2) for cell phone interviews, and for multiple landlines (question T5) for landline interviews. We capped the adjustments at three (i.e., for the number of phones reported greater than three, we still used the factor of three). **Exhibit 1.22** shows the number of cases affected by the adjustments.

Exhibit 1.22. Cases affected by adjustments for multiple phones

Frame	Multiple Phones Adjustment	*n*
Landline	1	180
Landline	2	4
Landline	3+	5
Cell phone	1	412
Cell phone	2	99
Cell phone	3+	39

Source: Abt calculations based on the original data

After this adjustment, the unequal weighting design effect increases to 1.163.

(3) ***Adjustment for respondent selection within landline households:*** Because landlines are treated as household devices, we performed within-household selection in the landline interviews. The within-household selection rules depended on the composition of the household, and in particular on the presence and the number of leave takers, leave needers, and employed only. The selection rules were as follows:

- In households with one person, no selection needed, as the selected respondent is the one on the line. Weight adjustment factor = 1.
- In households with more than one person, if all potentially eligible respondents (aged 18+, worked for pay last year) are of the same FMLA category (all leave takers, all leave needers, all employed only), select one participant at random. Weight adjustment factor = number of eligible adults in the FMLA category.

- In households where all eligible adults are employed only, generally subsample 20 percent for extended interview, terminate interview for others. (The adjustment factor for that is built at a later stage; see adjustment item 4 below.) Weight adjustment factor determined later.
- In households with both leave takers and leave needers, assign the household to the leave needer interview with a probability of 90 percent or to the leave taker interview with a probability of 10 percent. Weight adjustment factor = 1.11 × number of leave needers if a leave needer is sampled; = 10 × number of leave takers if a leave taker is sampled.
- In households with both leave takers and employed only, assign the household to the leave taker interview with a probability of 90 percent or to the employed only interview with a probability of 10 percent. Weight adjustment factor = 1.11 × number of leave takers if a leave taker is sampled; = 10 × number of employed only if employed only is sampled.
- In households with both leave needers and employed only, assign the household to the leave needer interview with a probability of 90 percent or to the employed only interview with a probability of 10 percent. Weight adjustment factor = 1.11 × number of leave needers if a leave needer is sampled; = 10 × employed only if employed only is sampled.
- In households that have leave needers, leave takers, and employed only, assign the household to the leave needer interview with a probability of 80 percent, to the leave taker interview with a probability of 10 percent, or to the employed only interview with a probability of 10 percent. Weight adjustment factor = 1.25 × number of leave needers if a leave needer is sampled; = 10 × number of leave takers + number employed only if otherwise.
- If there are multiple household members of the assigned type, select one of them at random.

We capped the adjustment factor at five. **Exhibit 1.23** shows the adjustment statistics.

Exhibit 1.23. Statistics for within-household respondent selection

Within Household Selection Rates	*n*
1.00	654
1.25	37
2.00	37
2.50	1
3.00	6
5.00	4

Source: Abt calculations based on the original data

This adjustment corrects for the over-representation of FMLA target subgroups of leave takers and leave needers relative to employed only. (This over-representation is induced by the study design, to better balance the sample sizes of FMLA subgroups and employed only.)

After this adjustment, the unequal weighting design effect increases to 1.448.

(4) ***Correction for employed only subsampling:*** To account for subsampling of the employed only respondents in employed only households, we computed an adjustment factor as the ratio of the employed only respondents subsampled for an extended interview over all respondents screened

as employed only in employed only households. Unlike other subsampling rates, we adjusted this rate during the field period to optimize the sample yield, and the aggregate over the whole field period was used. The adjustment factor is equal to 4.929.

After this adjustment, the unequal weighting design effect increases to 2.455.

(5) ***Correction for dual landline and cell phone users:*** To account for the elevated probabilities of selection of dual landline and cell phone users, we used a simple frame count composite factor integration approach: we multiplied dual users' weights by 0.5 to correct the bias of over-representation of the dual users.

After this adjustment, the unequal weighting design effect decreases to 2.275.

(6) ***Adjustment for non-response between screener completion and extended interview completion:*** To account for non-response between the screener completion and the extended interview completion, we used a logistic regression model to fit to the set of survey respondents who completed the screener (3,228 cases), with completion of the extended interview as the dependent variable, and using age, education, and gender, interacted with the frame (landline versus cell) as predictors available upon the screener completion. The adjustment factor is the inverse predicted probability based on this model. This adjustment corrects for the differences in non-response between the demographic groups, and importantly, for the different behaviors of the landline versus cell respondents. (Fit separately in the landline and cell phone frames, the models only require the main effects of demographics, but are very different from each other. Therefore, frame interaction was required in conjunction with demographic fit.)

After this adjustment, the unequal weighting design effect increases to 2.649.

Calibration

The last stage of adjustment, weight calibration (*raking*), involves iterative adjustment of weights to multiple weighting targets. We calibrated the main weights to the control totals defined in Section 1.6.1 (the subsection Population Control Totals). In the first pass of raking, weights were unrestricted; in the second pass, the weights were trimmed at the upper end to not exceed the 99th percentile of the unrestricted weights. **Exhibit 1.24** displays the summary statistics for the untrimmed and trimmed weights.

Exhibit 1.24. Summary statistics: untrimmed and trimmed weights

Summary Statistics	Untrimmed Weight	Trimmed Weight
Min	1,567	1,594
Median	60,239	60,453
Mean	190,244	190,244
Max	3,162,546	2,170,891

Source: Abt calculations based on the original data

The untrimmed weights have an unequal weighting effect of 5.026. This is a large design effect, due to the differential non-response between population groups. The trimmed weights have a reduced unequal weighting effect of 4.604.

1.6.3 KnowledgePanel Weights

The Ipsos KnowledgePanel weighting process consisted of three major steps: base/design weights accounting for probability of selection into the panel; calibration to the U.S. population for those who completed the screener; and an additional calibration step for the employed only population.

Ipsos weighting methodology started with the computation of design weights for all panel participants who were sent invitations to the Employee Survey. The design weights reflect the selection probabilities for the invited respondents. In the next step, design weights for all respondents, prior to any screening or subsampling, were raked to geodemographic distributions of U.S. adults, with finer demographic adjustments within:

- initial paid leave states (California, New Jersey, Rhode Island),
- New York, and
- Rest of the United States.

Moreover, an additional raking variable was included to correct the share of respondents across the following seven geographic locales:

- New Jersey and Rhode Island;
- California;
- New York;
- Western region, excluding the above;
- Southern region, excluding the above;
- Midwestern region, excluding the above; and
- Northeastern region, excluding the above.

Ipsos used the 2017 American Community Survey data to create population benchmarks, which included the following variables:

- age;
- gender;
- race/ethnicity;
- education;
- geographic regions as defined above;
- English language proficiency for Hispanics; and
- household income.

The above interim weights, which can be used for prevalence estimation, were labeled screener_wt and included a total of 26,140 respondents. The unequal weighting design effect across all respondents was 1.912.

Weighting of qualified respondents

Because the Employee Survey respondents (leave takers and needers) were not subject to any screening requirements, their final KnowledgePanel weights (variable weight2) were set equal to the above screener weights (screener_wt). The unequal weighting design effect within that group was 2.177. However, the final weights for the employed only respondents had to be created in two steps. In the first step, Ipsos used the screener weights for all 10,188 eligible respondents to create geodemographic benchmarks for this cohort. In the second step, the above benchmarks were used to create the final weights for all qualified respondents in the employed only domain. In doing so, finer demographic adjustments were applied within the three areas. The unequal weighting design effect of WEIGHT1 was 2.481.

Finally, extreme weights were trimmed and the resulting weights were scaled back to their respective number of qualified respondents in each stratum. These final weights were labeled UNTRIMWEIGHT1 and UNTRIMWEIGHT2, respectively.

1.6.4 Integrated Weights

Once we developed the weights for each of the phone and web components, we combined them into the overall analysis weight. This weight accounts for the possibility of an employee to be selected for the survey through either phone or web. This is the same nature of multiplicity that arises with frame composition of landline and cell phone samples in the RDD surveys, except that the nature of the frame overlap is much simpler: the RDD frame is fully contained within the Ipsos web panel sampling frame of mailable addresses.

Each of the RDD and KnowledgePanel samples had weights that sum up to the estimated eligible population of employees. In order for these resulting weights to continue to sum up to that same population size, the two samples need to have their weights multiplied by the factor that sums up to 1. We implemented factors proportional to the effective sample sizes (the ratio of the nominal sample size to the unequal weighting design effect): based on these effective sample sizes, the RDD weights were multiplied by 0.1444 and KnowledgePanel weights by 0.8556. This step brings the RDD weights (which are initially much higher due to lower sample size)[22] closer to the KnowledgePanel weights. **Exhibit 1.25** shows the weighting summary after we combined the RDD and KnowledgePanel weights.

Exhibit 1.25. Weighting summary post integration

Sample	Nominal *n*	Design Effect	Effective *n*	Compositing Factor	Average Weight
RDD	739	4.320	171.0	0.1444	27,466
KnowledgePanel	3,731	3.680	1,013.8	0.8556	29,715

Source: Abt calculations based on the original data

The resulting weights are calibrated to the target population benchmarks based on data from the American Community Survey and the CPS.

The unequal weighting design effect of the combined weights is 4.053. **Exhibit 1.26** presents the weighting summary after calibration to population benchmarks. Design effects and effective sample sizes will be specific for each particular analysis to be performed on the combined data set or its subsamples.

Exhibit 1.26. Weighting summary post population calibration

Sample	Design Effect	*Nominal n*	Effective *n*
RDD	4.678	739	157.95
KnowledgePanel	3.924	3,731	950.85
Employed only	2.747	2,128	774.99
Leave takers, leave needers	4.787	2,342	489.21
Paid leave states including New York	4.908	1,210	246.52
Not a paid leave state	3.803	3,260	857.22

Source: Abt calculations based on the original data

22 Recall that an average weight is the population size divided by the sample size, and both the RDD and the KnowledgePanel samples represent the same population.

The impact of this final step of combining the weights between the two samples can be assessed by comparing the unequal weighting design effects for the sample-specific weights in **Exhibit 1.25** of 4.320 for the RDD sample and 3.680 for the online sample, versus the values in **Exhibit 1.26** of 4.678 for the RDD sample and 3.924 for the online sample.

1.6.5 Replicate Weights and Variance Estimation

We created a set of 200 bootstrap replicate weights to facilitate correct variance estimation, by incorporating the sampling error in the weight adjustment factors.[23] The weight variables are cmb_bsrw1, cmb_bsrw2, …, cmb_bsrw200. For each replicate weight, the weighting steps were repeated as follows:

(1) Base weights in both the RDD and the KnowledgePanel samples were transformed into scaled bootstrap replicate base weights by multiplying the original base weight by a scaled bootstrap frequency (Rao, Wu, & Yu, 1992).

(2) The RDD qualification rates and adjustment factors (working numbers rate; subsampling of employed only rate; modeled response propensity) were recomputed using the new weights.

(3) RDD weights were recalibrated using the same control totals as the main weights.

(4) A qualification propensity model was fit with the weighting variables as predictors. The qualification propensity from the model was used as an additional weighting factor.[24]

(5) The predictions of the weights from the complementary frame were obtained, as detailed above, in Section 1.6.4 (Integrated Weights).

(6) The weights were calibrated to the CPS control totals.

Analyses with replicate weights

To incorporate the calibrated weights in an analysis, the analyst should use the following software syntax. (In the below, the weights variable names are accurate, whereas the variables `fmla_status, male, log_wage,` and `age` are to be understood as generic survey variables.)

In Stata, the acts of declaring a survey design and estimation are isolated:

```
* settings
svyset [pw=combo_trimmed_weight], vce(bootstrap) bsrw(cmb_bsrw*) mse
* analysis for the full sample
svy : tab fmla_status
* analysis in a subsample of males
svy, subpop(if male==1) : tab fmla_status
* regression analysis in a subsample;
svy, subpop(if male==1) : regress log_wage age
```

[23] Due to the complex nature of the 2018 Employee Survey, formulas commonly used in less complex 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, this set of 200 bootstrap replicate weights were created.

[24] Based on sample size targets in different states and FMLA groups, Ipsos KnowledgePanel had a complicated qualification routine to select employed only respondents who were to receive the extended interviews. Its last step of weight calibration accounted for this process, but its internal data were not available to Abt. Hence, we fit a qualification propensity model with the weighting variables as predictors, and qualification propensity from it was used as an additional weighting factor (for replicate weights only).

In SAS, specification of the sampling design and replicate weights is incorporated into PROC syntax itself:

```
* analysis for the full sample;
PROC SURVEYFREQ data=FMLA18 VARMETHOD = brr;
  WEIGHT combo_trimmed_weight;
  REPWEIGHT cmb_bsrw1--cmb_bsrw200;
  TABLES fmla_status;
RUN;
* analysis in a subsample of males: include in TABLES statement;
PROC SURVEYFREQ data=FMLA18 VARMETHOD = brr;
  WEIGHT combo_trimmed_weight;
  REPWEIGHT cmb_bsrw1--cmb_bsrw200;
  TABLES male * fmla_status;
RUN;
* regression analysis in a subsample;
PROC SURVEYREG data=FMLA18 VARMETHOD = brr;
  WEIGHT combo_trimmed_weight;
  REPWEIGHT cmb_bsrw1--cmb_bsrw200;
  DOMAIN male;
  MODEL log_wage = age;
RUN;
```

2. Worksite Survey

This chapter presents the methods used to design and administer the 2018 FMLA Worksite Survey: the target population and sampling design (Section 2.1); questionnaire development (Section 2.2); data collection procedures (Section 2.3); response rate calculations (Section 2.4); analysis of non-response (Section 2.5); weighting (Section 2.6); variance estimation (Section 2.7); and employee-level estimates (Section 2.8).

2.1. Target Population and Sampling Design

The Worksite Survey was a multi-mode (web and CATI) survey of 2,206 U.S. worksites. The field period was March 6, 2018, through February 19, 2019. A total of 1,891 surveys were completed via the web (86 percent of completes) and 315 interviews were completed by phone (14 percent of completes). The target population for the survey was U.S. private-sector worksites, excluding self-employed persons without employees and also excluding government and quasi-government units (federal, state, and local governments; public educational institutions; and post offices). This design deliberately includes both worksites covered by the FMLA and those not covered.

As in the previous survey waves, a worksite was defined as the "single physical location [or address] where business is conducted or where services or industrial operations are performed."[25] Data were collected and analyzed with respect to this worksite, even if the employer has other worksites. The sample universe for the Worksite Survey differs from that for the Employee Survey, which also includes those working for public-sector employers. Also as in previous Worksite Surveys, the Dun & Bradstreet Dun's Market Identifiers (DMI) file served as the sampling frame. The DMI database includes more than 22 million firm listings and is considered the most comprehensive commercially available firm list. It includes variables on worksite size, industry, and location that are necessary for stratifying the sample, as well as information that could be used to identify and remove some out-of-scope worksites.

The target respondent for each worksite was the human resources director or the person responsible for the company's benefits plan. Since the 2012 survey, the DMI offers additional contact information that could aid in identifying a "key informant" within a worksite's human resources department. This added information had the potential to allow an interviewer to more precisely target and pursue a prospective respondent.

2.1.1 Stratification

The sampling design for the Worksite Survey was similar to the design used for previous waves of the survey. 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. The Worksite Survey differed from earlier waves in that it oversampled paid leave areas with parameters identical to those in the Employee Survey (California, New Jersey, and Rhode Island, oversampled so that the base rate of selection was 25 percent higher for worksites in those states). Stratification allows the statistician to allocate resources for data collection so as to address the complex and sometimes conflicting needs of the sampling design. For instance, in order to make inferences on the population of *worksites* (e.g., a proportion of worksites that offer paid leave), a good sampling design would be a self-weighting design, such as a simple random sample or a proportional stratified sample. However in order to make inferences on the population of *employees* (e.g., the proportion of employees taking leaves), a good sampling design would be the design with probabilities proportional to the employment size. The sampling design of the 2018 FMLA Worksite Survey balances these two competing demands on precision by stratifying, among

[25] Previous reports used the term *establishment*.

other frame variables, by establishment employment size, and drawing larger establishments with higher probabilities, as explained below.

The Worksite Survey used the following employment size classes: (1) 1-49; (2) 50-249; (3) 250-999; (4) 1,000+. 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**. The Worksite Survey used two geographical groupings, paid leave states (California, New Jersey, Rhode Island) and non-paid leave states (all other states). The sample allocated so as to give the worksites in the states of CA, NJ and RI 25 percent higher probability of selection than that of their counterparts in the same industry-by-size class in the rest of the country. (The State of New York passed paid leave legislation after the sample was drawn and was thus not oversampled. Data collected from New York after July 1, 2018, was included in the paid leave designation at the analysis stage.) Cross-classifying four employer size groups, four industry groupings, and two state groups yields 32 sampling strata (four size categories × four NAICS groups × two state groups). See Exhibit 2.2 for the detailed strata.

Exhibit 2.1. NAICS groupings for Worksite Survey

Group	NAICS Codes
NAICS Group 1	Agriculture, Forestry, Fishing and Hunting (11); Mining, Quarrying, and Oil and Gas Extraction (21); Construction (23); Manufacturing (31-33)
NAICS Group 2	Utilities (22); Wholesale Trade (42); Retail Trade (44-45); Transportation and Warehousing (48-49)
NAICS Group 3	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 4	Educational Services (61); Health Care and Social Assistance (62); Arts, Entertainment, and Recreation (71); Accommodation and Food Services (72); Other Services (81)

Source: Design Report Wave 4 FMLA Surveys

Exhibit 2.2. Detailed sampling strata for the Worksite Survey

Stratum	Geography	NAICS Group	Employment Size
1	Paid leave states	1	1-49
2	Paid leave states		50-249
3	Paid leave states		250-999
4	Paid leave states		1,000+
5	Paid leave states	2	1-49
6	Paid leave states		50-249
7	Paid leave states		250-999
8	Paid leave states		1,000+
9	Paid leave states	3	1-49
10	Paid leave states		50-249
11	Paid leave states		250-999
12	Paid leave states		1,000+
13	Paid leave states	4	1-49
14	Paid leave states		50-249
15	Paid leave states		250-999
16	Paid leave states		1,000+

Stratum	Geography	NAICS Group	Employment Size
17	Non-paid leave states	1	1-49
18	Non-paid leave states		50-249
19	Non-paid leave states		250-999
20	Non-paid leave states		1,000+
21	Non-paid leave states	2	1-49
22	Non-paid leave states		50-249
23	Non-paid leave states		250-999
24	Non-paid leave states		1,000+
25	Non-paid leave states	3	1-49
26	Non-paid leave states		50-249
27	Non-paid leave states		250-999
28	Non-paid leave states		1,000+
29	Non-paid leave states	4	1-49
30	Non-paid leave states		50-249
31	Non-paid leave states		250-999
32	Non-paid leave states		1,000+

Source: Design Report Wave 4 FMLA Surveys

Following standard Abt practice, the sample was allocated proportionally to the square root ($\sqrt{X_h}$) of the total employment in the stratum as a compromise allocation for computing both per-worksite and per-employee estimates from a Worksite Survey. This was an intermediate-value approach given that many tabulations from the Worksite Survey are weighted to a "per employee" basis and interest in the impact of the paid leave policies. This allocation method allowed worksites with a large number of employees to be selected at a higher rate than worksites with fewer employees. This ensured that enough large worksites were available for the analysis, and hence the per-employee measures were of adequate accuracy.

To demonstrate the advantage of the $\sqrt{X_h}$–proportional allocation, it may be helpful to contrast it with simple proportional allocation. Under simple proportional allocation, all worksites would be selected with the same probability. A worksite of three employees would have the same probability of being selected as a worksite with 30,000 employees. When the purpose of the study is to make inference to the population of *worksites*, the proportional allocation would be the preferred approach. In the Worksite Survey, however, many key variables (e.g., number or percentage of employees covered by FMLA) are related to *employment* size.[26] Large worksites are more important for estimating such variables than are smaller worksites, and a greater number of large worksites helps stabilize the estimates.

That is, the variance in an employment-related variable is concentrated more in the strata of large worksites than in the strata of smaller worksites. To reduce the variance (i.e., increase the precision) of an estimate of such a variable, the strata of large worksites were sampled with a higher probability than the strata of smaller worksites. This way, there were more large worksites in the sample, and the influence of large worksites on the variance of the estimate could be reduced.

[26] See the Abt Methodology Report for the 2012 FMLA survey (Daley et al., 2013). Our report also points out the difficulties in variance estimation and design effect calculations associated with these two goals.

All estimates in the Family and Medical Leave Act 2018 Survey Results Report are adjusted for the oversampling by weighting worksites by the inverse of their probability of selection. See also Section 2.8 below on the statistically appropriate ways to produce estimates of employee characteristics.

2.1.2 Consistency across Previous Worksite Surveys

The sampling design in the 2000, 2012, and 2018 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 worksites in the industry group. The definition of the target population was kept consistent across the surveys.

2.1.3 Notable Differences from Previous Worksite Surveys

The 2012 Worksite Survey was the first to allow the respondent to complete the survey online. In previous Wave 1 and Wave 2, all interviews were conducted using CATI only. Some 86 percent of the interviews completed in 2018 (Wave 4) were completed via the web and 14 percent were completed by phone (CATI). This is a notable increase in web completes since the 2012 survey, when 35 percent of interviews were completed via the web and 65 percent by phone. The web version gave respondents flexibility in the time, location, and pace of completing the survey, especially with regard to consulting administrative records. Consistent with this consideration, both the web and phone surveys were designed to allow the respondent to leave and re-enter the survey as frequently as they wished and at any time.

The changes in Worksite Survey stratification across the survey waves are summarized in **Exhibit 2.3**.

Some of the changes made pertain to the sampling strata. The idea of stratification by employment size was consistent with the 2000 and 2012 surveys, but we updated the definition of the classes. Employee size classes in 1995 and 2000 were 1-10, 11-24, 25-49, 50-99, 100-250, 251-999, and 1,000+; in 2012 they were 1-9, 10-19, 20-49, 50-99, 100-249, 250-999, and 1,000+.[27] Our experience with the 2012 Worksite Survey, however, showed us that the detailed breakdown of the classes resulted in far more complex survey management, without providing a justifiable improvement in precision of the survey estimates. In 2018 we simplified the classes to just four: 1-49, 50-249, 250-999, 1,000+.

In 2018, we added the paid leave and non-paid leave state groupings to the stratification to allow improved standalone estimates for the selection of states with paid leave policies. This would allow insight into paid leave policy and its influence on the changing domain of FMLA. The industry codes remained the same as in 2012, when they had changed to reflect the conversion from Standard Industrial Classification (SIC) codes to NAICS.[28]

---

[27] The employee size classes were changed for 2012 to match the size classifications used in the Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW).

[28] Attempts to reproduce the classification of the 1995 and 2000 surveys in NAICS codes yielded unbalanced sample sizes between groupings. Instead, we defined a fresh set of NAICS groupings for the 2012 survey.

Exhibit 2.3. Worksite Survey sampling stratification by survey wave

<table>
<tr><th></th><th>Wave 1 (1995)</th><th>Wave 2 (2000)</th><th>Wave 3 (2012)</th><th>Wave 4 (2018)</th></tr>
<tr><td colspan="5">Completed Interviews by Mode (%)</td></tr>
<tr><td>Web</td><td>NA</td><td>NA</td><td>35</td><td>86</td></tr>
<tr><td>CATI</td><td>100</td><td>100</td><td>65</td><td>14</td></tr>
<tr><td colspan="5">Sampling Strata</td></tr>
<tr><td rowspan="7">Employee Size</td><td>1-10</td><td>1-10</td><td>1-9</td><td rowspan="3">1-49</td></tr>
<tr><td>11-24</td><td>11-24</td><td>10-19</td></tr>
<tr><td>25-49</td><td>25-49</td><td>20-49</td></tr>
<tr><td>50-99</td><td>50-99</td><td>50-99</td><td rowspan="2">50-249</td></tr>
<tr><td>100-250</td><td>100-250</td><td>100-249</td></tr>
<tr><td>251-999</td><td>251-999</td><td>250-999</td><td>250-999</td></tr>
<tr><td>1,000+</td><td>1,000+</td><td>1,000+</td><td>1,000+</td></tr>
<tr><td>Industry</td><td colspan="2">Standard Industrial Classification (SIC) codes</td><td colspan="2">North American Industry Classification System (NAICS) codes</td></tr>
<tr><td rowspan="2">Location</td><td>NA</td><td>NA</td><td>NA</td><td>Paid leave state</td></tr>
<tr><td>NA</td><td>NA</td><td>NA</td><td>Non-paid leave state</td></tr>
</table>

Source: Design Report Wave 4 FMLA Surveys

Finally, the response rate for the 2018 survey (7.0 percent) is noticeably lower than that reported for the 2012 Worksite Survey (20.9 percent). A decline in response rates is consistent with industry-wide experience (see e.g., https://www.bls.gov/osmr/response-rates/). Like household surveys, worksite surveys are also subject to non-response concerns and their potential for bias (IGEN, 1998; Petroni et al., 2004). A transition of the primary mode from active (outbound phone calling) in 2012 and earlier to passive (web self-response) in 2018 may also have contributed to the response rate decline. We conducted an extensive non-response analysis and discuss the potential implications in Section 2.5 below.

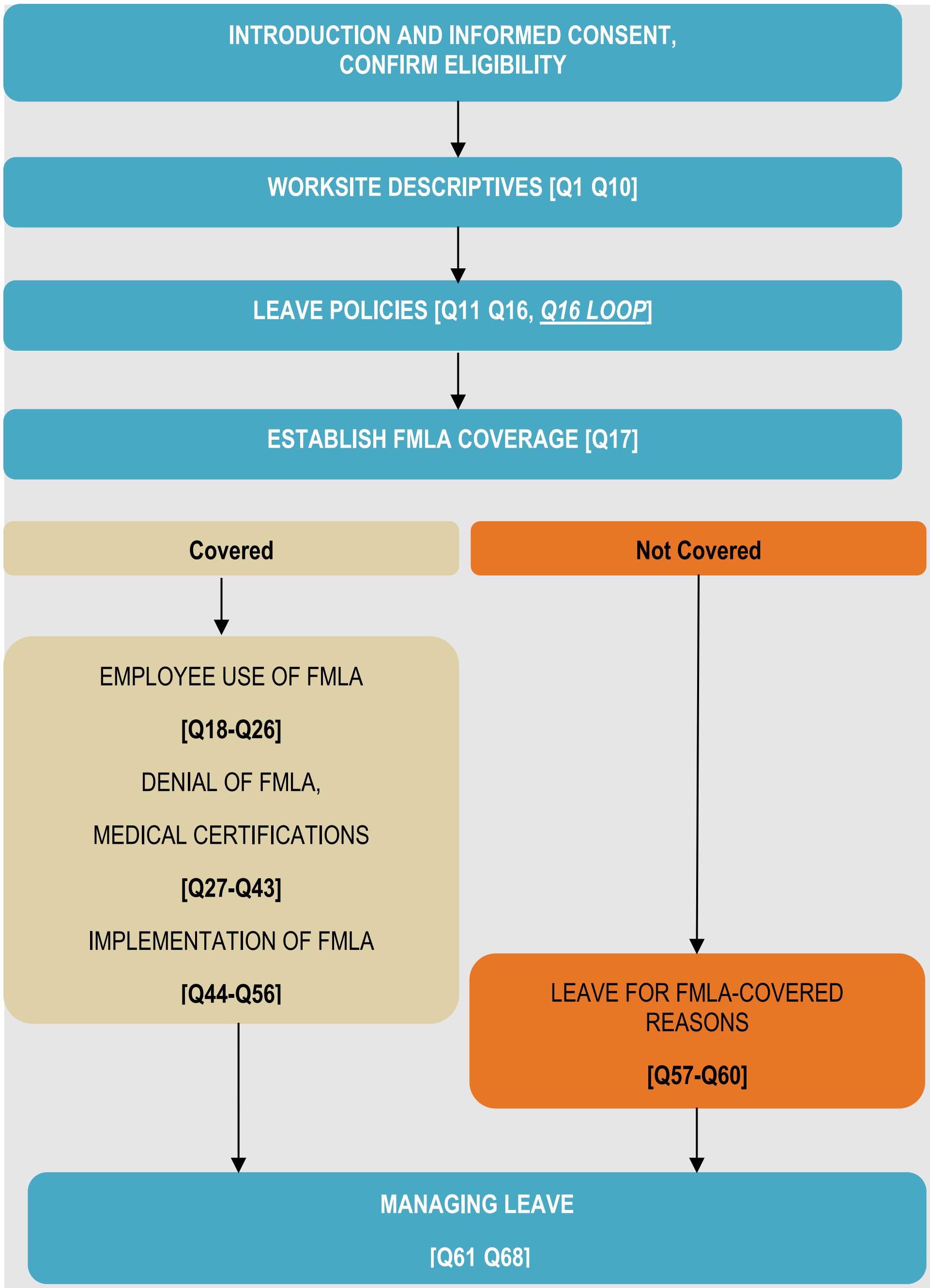
2.2. Questionnaire Development

The 2018 Worksite Survey retained the main structure of previous waves.

2.2.1 Questionnaire Overview

The questionnaire for the Worksite Survey consisted of two primary components: a screening questionnaire and an extended interview. Screening confirmed that sampled worksites still existed and then obtained the correct contact information for the most knowledgeable person at the firm to answer the extended interview. The screener also confirmed that the worksite was neither a quasi/government nor other non-profit organization. The extended interview consisted of questions to describe the worksite and its leave policies in general. Within these questions, it established coverage by FMLA. Covered worksites received a series of questions about their employees' use of FMLA and detailed questions about the implementation of FMLA. **Exhibit 2.4** provides a representation of the Worksite Survey structure.

Exhibit 2.4. Structure of the Wave 4 Worksite Survey



Source: Design Report Wave 4 FMLA Surveys

2.2.2 Survey Revisions

Part of our process to prepare the survey instrument for 2018 was to systematically reconsider the 2012 instrument and data collection results. We classified each 2012 question according to the following criteria:

- importance of comparability across waves (e.g., whether question was asked since Wave 1 in 1995);

- response variation (i.e., whether a large majority of respondents gave the same response); and
- data utility (e.g., how findings were used or presented in prior waves).

Though cross-wave consistency and emerging policy priorities primarily shaped the redesign decisions, these three criteria provided second-tier input that helped with the final set of decisions about the utility of a particular question or variable. Ultimately, we grounded our decisions about survey revisions suggested to DOL in the established set of priorities described below.

Questions or topics given the *most weight*

- reflected the expressed interests of DOL;
- captured changes since 2012 in the leave policy landscape, including the increasing incidence of state and local paid leave laws and the amended definition of "spouse" for FMLA;
- were raised by multiple participants in the listening sessions with key stakeholders; and
- have been consistently tracked since 1995 and 2000 (Waves 1 and 2).

Conversely, questions or topics given the *least weight* were those

- not raised or not of interest during listening sessions with key stakeholders;
- answered by very few respondents in the 2012 survey; and
- whose responses varied little in the 2012 survey.

These criteria were applied against fixed length and burden parameters.

We generated a set of detailed recommendations to DOL for question-by-question updates, grouped by question domain in a matrix. Within each question domain, this matrix further grouped survey questions for 2018 into those we:

- retained from 2012;
- revised from 2012;
- deleted from 2012; and
- added (new).

Appendix C provides the matrix containing the full set of changes made to the 2018 Worksite Survey. The following brief list describes key changes to the 2012 survey for the 2018 survey:

- Expanded questions on reasons for employee leaves taken.
- For employers that offer benefits to a subset of employees, expanded questions to capture additional detail on variation in benefits by employee type.
- Revised questions on worksite characteristics (e.g., prevalence of low-wage workers) and firm characteristics (e.g., whether firm has worksites in multiple jurisdictions, firm revenue).
- Revised questions on medical certifications to focus on most commonly required certification (i.e., initial medical certification) and deemphasize less common secondary certifications, re-certifications, and fitness-for-duty certifications.

- Combined questions on leaves related to a family member in military service (related to preparing for deployment, or to care for a family member injured during military service).

Appendix B provides the final 2018 Worksite Survey instrument and example screen shots of the web version. The remainder of this section highlights key experiments/tests that shaped the revision of the Worksite Survey for 2018.

2.2.3 Response Option Experiments

The 2018 Worksite Survey explored the difference that certain response options might make to the results, looking at two instances in which we revised the response options for added distinction. **Exhibit 2.5** summarizes the response option experiment conditions.

Most/Some. Both the 2018 and 2012 Worksite surveys asked the employer to estimate how many employees it provided with various types of leave. In 2012, about the same number of respondents answered "Most" and "Some," suggesting that employers may not have consistently differentiated between these response options. The 2018 experiment occurred in the question 11 series. It was repeated in Q14a, Q25, and Q27, because they presented the same response options. Half the employers saw the set of response options where "Most" included the quantifier "half or more" and "Some" included the quantifier "less than half." The other half of the employers saw a set of response options that presented *only* the quantifiers "Half or more" and "Less than half."[29] (Shown in the top panel of Exhibit 2.5.)

Neither/Nor. Both the 2018 and 2012 surveys also asked for the employer's opinions on the effects of complying with FMLA. In 2012, these questions included the response option "No noticeable effect," presented in the last position. Review of these questions by the Technical Working Group at the 2018 survey design phase resulted in changing the wording of this response to a substantive, neutral option: "Neither [easy/positive] nor [difficult/negative]." The 2018 experiment explored whether the position of the revised neutral response option among the other response options would differentiate response. Questions Q52 and Q56 varied the presentation of the revised response option, from endpoint to midpoint. Half the employers received this option as the endpoint, as it was presented in 2012. The other half of the employers received it as the midpoint, which is typical of a substantive, neutral option. (Shown in the bottom panel of Exhibit 2.5.)

Exhibit 2.5. Response option experiments in the Worksite Survey

Variable	N	Question	Experimental Condition 1	Experimental Condition 2
Most/Some				
Q11_a	2,206	How many employees are provided with paid sick leave?	1 All 2 Most (half or more) 3 Some (less than half) 4 None 9 REF	1 All 2 Half or more 3 Less than half 4 None 9 REF
Q11_b	2,206	How many employees are provided with paid disability leave?		
Q11_c	2,206	How many employees are provided with paid vacation?		
Q11_d	2,206	How many employees are provided with paid maternity leave?		
Q11_e	2,206	How many employees are provided with paid paternity leave?		

[29] Note that neither option in 2018 is exactly comparable to the 2012 data because in 2012 options were "Most" and "Some" (no quantifiers).

Variable	N	Question	Experimental Condition 1	Experimental Condition 2
Q11_f	2,206	How many employees are provided with paid leave for another family member's illness or medical care?		
Q11_g	2,206	How many employees are provided with paid leave for eldercare?		
Q11_h	2,206	How many employees are provided with flex time?		
Q11_i	2,206	How many employees are provided with other paid time off, excluding paid holidays?		
Q11a	936	How many employees are provided paid time off or PTO?		
Q14a	1,937	How many of your hourly workers earn an hourly wage below $15.00 per hour?		
Q25	1,128	About how many leaves taken under FMLA are given with notice from the employee that is consistent with your company's policies?		
Q27	1,546	How many FMLA leave applications were denied [from [INSERT 12-MONTH REFERENCE PERIOD]] for ANY reason?		
Neither/Nor				
Q52	1,546	In general, how easy or difficult has it been for this location to comply with FMLA?	1 Very easy 2 Somewhat easy 3 Somewhat difficult 4 Very difficult 5 Neither easy nor difficult 9 REF	1 Very easy 2 Somewhat easy 3 Neither easy nor difficult 4 Somewhat difficult 5 Very difficult 9 REF
Q56	1,546	Thinking about employee productivity, absenteeism, turnover, career advancement and morale, as well as the business' profitability, what effect has complying with FMLA had on this location?	1 Very positive 2 Somewhat positive 3 Somewhat negative 4 Very negative 5 Neither positive nor negative 9 REF	1 Very positive 2 Somewhat positive 5 Neither positive nor negative 3 Somewhat negative 4 Very negative 9 REF

Source: 2018 Worksite Survey

There were significant differences in response for only three questions, Q11_b, Q11_g, and Q27. There was no significant difference in response in Q52 or Q56. For analysis, data were combined for all but these three questions.[30] See Appendix E for the detailed findings from these experiments.

2.2.4 Cognitive and User Testing

The 2018 Worksite Survey was subject to *user testing*, which focuses more on the experience of the respondent with respect to, for instance, willingness to complete, time burden, and ease of navigation through the instrument. We considered *cognitive testing* (to identify sources of response error, often due to language) as unnecessary, as the survey interviewed respondents who were well versed in the concepts and language it used. In addition, both modes (phone, web) of the Worksite Survey provided detailed references and definitions accessible as needed.

The user tests took place in February 2017. Respondents were a convenience sample of Abt human resources staff and non-Abt individuals identified through our personal networks. The user test sample was diverse with respect to the type of company, respondent position within the company, and whether the company processed FMLA requests internally or outsourced to another company. A total of 11 prospects were sent an email invitation and link to the web survey, with the goal of getting nine completes of the user tests. The email invitation included a detailed explanation of the study and the information respondents would need to reference in order to answer the survey questions.

A total of five sample members responded within the time period for user testing. All five respondents completed a debriefing interview after finishing the online questionnaire. The debriefing interview was an open-ended conversation about the overall ease of completing the questionnaire, retrieval of administrative data for answering the survey, comprehension of terms and topics, scope of the survey topics, and applicability of the topics for their firm. As a result of the user tests, we edited the survey to clarify certain survey questions and instructions.

2.2.5 Pre-Testing

In May 2017, we conducted a pre-test of the 2018 Worksite Survey, primarily to test the logistical, operational, and procedural aspects of the survey. Specifically, pre-testing helped us understand the process for obtaining and cleaning the sample file; the process for screening sample members, identifying the correct respondent, and determining the respondent's willingness to complete the survey; differences between phone and web administration; and survey length.

We conducted the pre-test with a subset of the Worksite Survey target population: U.S. private-sector firm worksites, excluding self-employed persons without employees and also excluding quasi/government and non-profit organizations. We included worksites covered by the FMLA and those not covered. We used the DMI file as the sampling frame.

The pre-test employed a modified version of the actual survey administration to accommodate its smaller sample size and shorter time frame. First, because we were limited to nine completed surveys for the pre-test due to OMB policy, we stratified the sampling frame by worksite size only, compared to the actual survey's cross-classification of worksite size and industry and oversampled paid-leave states. Second, the

---

[30] Analysis of these data were performed initially with Rao-Scott chi-square tests and additionally analyzed with corrections for multiple testing using Benjamini-Hochberg false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995). This procedure considers tests in the order of their *p*-values, and determines which ones should be considered significant given the number of tests and context of other small *p*-values. FDR is routinely applied in most statistical applications that have to deal with multiple parallel tests; for example, in genomics where test of the differences in the levels of gene expression between experimental arms need to be run for each of ten thousand or more genes in modern microarrays (i.e., the researcher deals with 10,000 tests; here we deal with 15 tests).

pre-test sample included a name and email address for each sampled worksite. As a result, we did not test our approach to screening for sampled worksites without a listed sample member. Finally, the timeline for the pre-test was compressed, and therefore the data collection protocol was also compressed.

We completed nine pre-test interviews. We learned we would need to request the sample file from Dun & Bradstreet well in advance of data collection because of its turnaround time. However, we did not get an accurate assessment of the quality of the sample because the amount of sample needed for the pre-test was much smaller than for the actual survey and because we selected a non-random sample to simplify screening and identifying a respondent, to fit into the compressed pre-test schedule.

We learned respondents were more willing to complete the web version of the survey than the phone version. Of the nine completed pre-tests, seven were via the web and two by phone. As a result, we revised the survey to promote the web version by offering the web survey after the telephone screening was complete. The pre-test survey timings aligned with our expectations for a 25-minute survey on average, with the web survey requiring an average of 20 minutes and the phone survey requiring 32 minutes. We did not obtain an accurate assessment of the level of effort for the phone data collection protocol because we were unable to pre-test this full protocol due to time constraints.

2.2.6 Pilot Survey

Like cognitive and user testing and pre-testing, the objective of a pilot survey is to test comprehension, timing, and navigation, but it adds the execution of the full recruitment and data collection protocol. This includes testing the self-administration of the web survey, interviewer administration of the phone survey, and timing of non-response follow-up efforts. We conducted no formal pilot survey for the 2018 Worksite Survey due to limitations in time and budget. This is worth noting because the original data collection protocol underwent modifications during the field period. These changes are described in detail in Section 2.3.4.

2.3. Data Collection Procedures

The 2018 Worksite Survey underwent OMB clearance from March 2017 to January 2018. Shortly after final clearance, we began data collection. In this section, we discuss data collection procedures: interviewer training, initial data collection protocol, challenges in data collection, and adjusted protocol.

2.3.1 Interviewer Training

We conducted intensive trainings for telephone supervisory staff and interviewers to prepare them for administration of the survey in early March 2018. The training reviewed general interviewing principles and unique study procedures and requirements. It also allowed interviewers access to the CATI equipment, to become familiar with the questionnaire and to perform practice interviews. At the start of the training, we 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. In the training we 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. We conducted mock interviews that we designed to mimic a variety of interview situations (smaller and larger worksites, covered and non-covered firms). Additionally, we conducted the training to ensure that interviewers were comfortable helping respondents to access the web version of the survey. We instructed interviewers to transfer technical questions they could not answer to a help desk or project manager.

Abt conducted data collection for the Worksite Survey by administering a screening interview, followed by an extended interview. Our experience administering the 2012 Worksite Survey informed our data collection protocol in several key ways. The 2018 survey administration also featured some key enhancements. We deployed the original data collection protocol in the first weeks of the 2018 survey and a revised protocol midway through the data collection period. The initial protocol and enhancements are described below.

2.3.2 Initial Data Collection Protocol

In the screening portion of the 2018 Worksite Survey, telephone interviewers contacted the sampled worksites to check eligibility for the survey and identify the respondent needed to respond on behalf of the worksite. For a worksite to be eligible, the worksite had to be the employer (and location) listed on the sampling frame, be in the private sector, and have at least one employee.

In the 2012 survey, at the larger worksites the key informant or person who answered the phone was not always prepared to address questions about the firm’s benefit plan. Given the detailed nature of the extended interview, which included some questions possibly requiring reference to company administrative records, it was necessary to identify the target respondent—a human resources director or the person responsible for the company’s benefits plan. Answering the extended interview questions often required a handoff from the key informant to another party, leading to breakoffs. With the aid of the enhanced DMI contact data for the 2018 survey, we attempted to avoid this handoff by trying to reach the target respondent during the screening interview.

In the 2012 survey, we mailed an informational packet after screening and identifying a respondent. For the 2018 survey, we attempted to collapse these two phases whenever possible and with the consent of the identified respondent. In the screening call, our interviewer offered to send the informational packet via email to the respondent. Again this required that we reach the target respondent in the screening interview.

The informational packet included a cover letter from the Chief Evaluation Office on DOL letterhead introducing the study, explaining its importance, and providing unique login information to a secure website to complete the web survey and a toll-free number to complete the survey by phone with an interviewer. The packet also included a project information sheet on Abt Associates letterhead introducing Abt as the data collection unit and describing the detailed information needed to complete the survey. See Appendix B for the Worksite Survey informational packet materials.

If we did not identify a respondent in the initial call attempts, we sent a generic version of the informational packet (excluding the unique login information) to the worksite name and address listed in the sample. We sent this informational packet via USPS Priority Mail so the recipient could clearly distinguish the survey materials from junk mail. This version did not allow access to the web survey because the survey respondent had not yet been identified; rather, this packet acted as a pre-notification to the follow-up screening call.

The original plan was to field this survey for six months as a mixed-mode phone and web survey. The screening portion was originally set up to be administered by phone only. A telephone interviewer would first get the respondent on the phone to verify and screen. If the respondent was eligible, we would email the informational packet, including the link to access and complete the extended interview online. The interviewer would also offer to help the respondent access the web survey or complete by phone. We expected most respondents would opt to complete via web, based on our pre-test. We planned to send four email reminders to non-respondents before attempting to follow-up again by phone. We would make 10 follow-up attempts by phone.

2.3.3 Challenges in Data Collection

Our experience in the first weeks of administering the 2018 survey informed changes to the original data collection protocol. First, we found that attempting initial contact by phone was almost impossible. The 2018 protocol intended to reduce breakoffs, but it had limited success. The enhanced DMI contact data were not as widely available as expected from our early communications with the vendor. Only about

25 percent of sample cases had a contact name attached that could be a human resources staffer, and often it was an incorrect respondent; for example, a recruiter rather than a benefits administrator. It was difficult to get a respondent on the phone to verify and screen. We instead reached gatekeepers such as receptionists and directory operators. We still experienced breakoffs due to reaching the incorrect respondent on the initial contact call. It was in fact hard to get a person on the phone at all; instead we reached automated menus and voicemail messages.

Another challenge was in the quality of the sample we received. We received wrong phone numbers, phone numbers not in service, and worksite listings with no phone number at all. Our telephone interviewers reported that when they called a firm and the key informant listed was someone who had not worked there for many years or was in a completely wrong department, it made the survey request seem less legitimate. It gave the appearance of a sales call using a dated, purchased list.

Finally, we were conservative in releasing the sample cases into the field because we had not conducted a pilot survey to test the full data collection protocol, due to limitations in time and budget. We conserved the sample while we monitored production. Eventually, we determined that we needed to change the data collection protocol.

2.3.4 Adjusted Data Collection Protocol

To address the challenge posed by attempting initial contact by phone, we implemented two changes. We changed initial contact to emailing or mailing the informational packet, followed by calling. We also revised the phone screening survey. To change the initial contact mode, we added screening questions to the web survey. We emailed the informational packet when an email address was available for a key informant and mailed it if we had only a mailing address (also known as “mail-push-to-web”). We revised the cover letter to include the web link to the survey and language asking the recipient to forward the packet to the benefits administrator. We changed how we sent this mailing from USPS Priority Mail to Federal Express, for added urgency/legitimacy. The emailed or mailed informational packet pre-notified the recipient that we would be following up by phone.

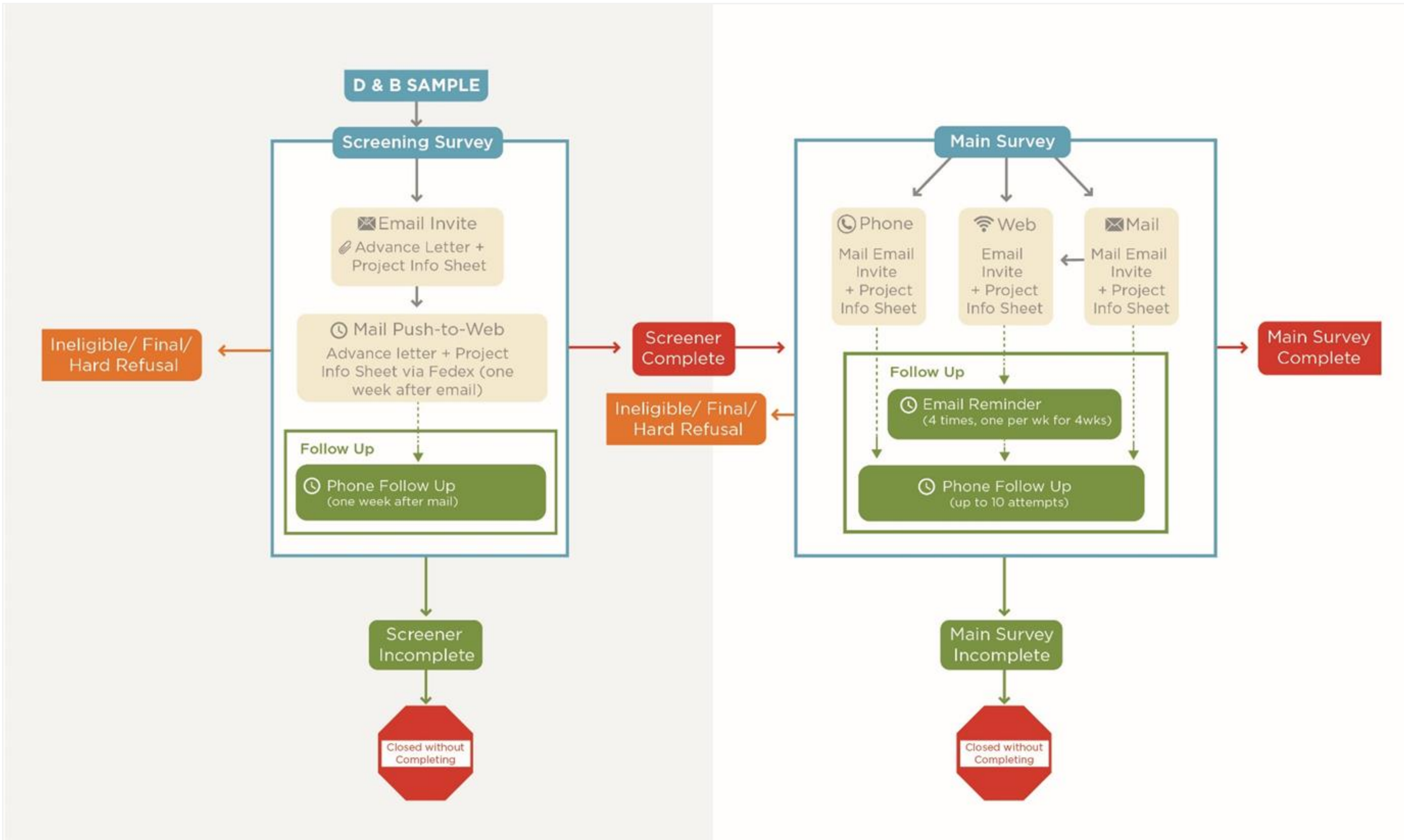
We called non-respondents to the web survey invitation in the informational packet and respondents that started the survey but did not complete. We attempted to conduct the screening interview by phone, identify the correct respondent, and/or administer the survey by phone. If the sampled worksite was a branch location, the respondent may have been located at a different location (e.g., company headquarters). Once a respondent was identified with an email address, we sent four email reminders to non-respondents as well as continued to follow up by phone to encourage completion of the survey, using the 10-call design.

Exhibit 2.6 provides a graphical representation of the adjusted data collection protocol for the Employer Survey.

We made additional changes to the data collection protocol to address the challenges we encountered. We revised the phone screening survey so we did not waste call attempts with trying to get the survey respondent on the phone. In the revised protocol, the survey allowed a key informant to complete the verification and screening. This person could also provide the respondent contact information, so we could then send the respondent the informational packet and follow up by phone.

Other revisions to the phone screening survey addressed the challenge of the poor quality sample. We no longer asked the vendor for the key informant name listed in the sample. More generally, we asked for the person who could best answer questions about medical and family leave policies and benefits. The interviewer could confirm the key informant name, if it seemed appropriate. See Appendix B for example screen shots of the web survey.

Exhibit 2.6. Contact protocol for the 2018 Worksite Survey



Source: 2018 Worksite Survey

We sought out LinkedIn for cases that had wrong phone numbers or phone numbers that were not in service, as a potential source for human resources professionals. Specifically, we worked with LinkedIn to place an ad object that said: “If you are interested in participating in this survey, click here to provide your contact information and we will send you the survey invitation.” Once someone provided their information, we had to check that the employer *and that worksite* were in our sample. If so, we would send the survey invitation and the employer would go into the data collection cycle. If not, we thanked them for their interest and didn’t pursue them further. Of about 800 sampled worksites, we obtained only 12 leads. Of them, 10 were viable, but none resulted in completed surveys.

For those worksites that came from Dun & Bradstreet with no phone number, we performed manual internet searches (e.g., just by using Google) to attempt to identify a publically available phone number. In this manner, we identified a phone number for 76 percent of worksites that we manually searched over the course of the field period.

To address the challenge of a conservative sample release, we adjusted the non-response follow-up schedule by making the phone follow-up concurrent with the email reminders. Originally the phone follow-up came *after* email reminders. We were less conservative about releasing the remaining sample; we started releasing more sample and at a more rapid pace. Once we adjusted the data collection protocol, we also purchased additional sample to complete the targeted number of interviews faster. We released the additional sample aggressively, as well. Because it was such a high volume, we partnered with a mail vendor to help mail the informational packets and with a call center to help with the telephone interviewing. The mail vendor performed address validation on sample addresses using the USPS’s National Change of Address database and Coding Accuracy Support System software, prior to mailing the informational packets.

On average it took respondents approximately 35 minutes to complete the Worksite Survey online and approximately 29 minutes to complete it by phone.[31] We administered the survey in English only.

2.4. Response Rates

The response rate for the 2018 Worksite Survey was computed in three steps. In the first step, we calculated a response rate for the screening interview. In the second step, we calculated a response rate for the extended interview. In the third step, we combined the two response rates to produce the overall survey response rate. These steps are described in detail below.

2.4.1 Response Rate for the Screener

Exhibit 2.7 displays the final dispositions for the screening stage. In total, Dun & Bradstreet sampled 60,664 worksites from the DMI frame, and 9,409 of these employers completed the screening process. The set of sample units completing the screener (AAPOR code 1.100 in **Exhibit 2.7** below) included those employers determined to be eligible for the survey (*n*=6,376) as well as those determined to be ineligible (*n*=3,033; e.g., no longer in business, a public institution).

[31] The timing of the web survey is a very general estimate. It is typical for web surveys to have outlying measurements of timing when the survey is left open on a web browser while the respondent performed other tasks on their computer. Survey lengths that were seemingly outliers were excluded from the computation of the average completion times, because we cannot conclude the reason for extremely lengthy surveys. This included surveys that were two and a half hours or longer, which is five times longer than the average survey duration.

Exhibit 2.7. Final dispositions for the screener and extended (main) interview

Disposition	AAPOR Code	Screener	Main Interview
Interview (Category 1)			
Complete	1.100	9,409	2,206
Eligible, Non Interview (Category 2)			
Refusal	2.110	11,376	3,669
Breakoff	2.120	176	0
Respondent never available	2.210	845	31
Telephone answering device	2.220	9,221	55
Other	2.300	0	1
Deceased respondent	2.310	0	1
Target respondent physically or mentally unable/incompetent	2.320	36	2
Language barrier	2.330	214	3
Unknown Eligibility, Non Interview (Category 3)			
Always busy	3.120	1,373	
No answer	3.130	2,183	
Call blocking	3.150	452	
No screener completed	3.210	9,464	
Not Eligible (Category 4)			
Fax/data line	4.200	572	
Non-working/disconnect	4.300	5,566	
Number changed (phone)	4.410	4,486	
Cell phone	4.420	23	
Other – not dialed	4.900	5,268	
No eligible respondent (zero workers)	4.700		408
Total		**60,664**	**6,376**

Source: 2018 Worksite Survey

Exhibit 2.8 reports the screener response rates. The first row reports the rates based on the AAPOR RR(3) formula as applied to worksite surveys. The second row reports the rates based on the formula used in the 2000 Establishment Survey Report. The latter formula yields a slightly higher response rate because it assumes that only 10 percent of the non-locatable sample units are eligible, as compared to the 66 percent estimated as eligible (*e*) under the AAPOR RR(3) formula.

Exhibit 2.8. Unweighted and weighted 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))*	23.4%	36.8%	8.6%	18.2%	38.4%	7.0%
2000 Establishment Response Rate Formula Formula: *C/(C+RB+LP+NA+NM+MC+(.10(NL+NW)))*	25.2%	36.9%	9.3%	19.0%	38.2%	7.3%

Source: 2018 Worksite Survey
Key (top): AAPOR = American Association for Public Opinion Research; I = completed interview; P = partial interview; R = refusal; NC = non-contact; O = other non-response; UH = unknown if household, UO = other unknown eligibility; NE = not eligible.
Key (bottom): C = completed interview; RB = final refusal; LP = language problem; NA = ring-no-answer; NM = answering machine; MC = maximum calls; NL = non-locatable; NW = non-working.
* 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 23.4 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 18.2 percent. The weighted response rates incorporate the screener base weights. That the unweighted rate is higher than the weighted rate reflects the fact that smaller worksites (which were weighted up to account for their lower selection probabilities) were less likely to respond to the screener than were some of the larger worksites. This pattern was also documented in the 2000 and 2012 Worksite Surveys.

2.4.2 Response Rate for the Extended Interview

The unweighted and weighted extended interview response rates also are presented in **Exhibit 2.8**. The unweighted AAPOR RR(3) for the extended interview is 36.8 percent and the weighted RR(3) is 38.4 percent. This pattern reflects the fact that smaller worksites (which had larger weights) were more likely than larger worksites to complete this 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 8.6 percent, and the weighted response rate is 7.0 percent (again **Exhibit 2.8**).

2.5. Analysis of Non-Response of the Worksite Survey

The response rate achieved in the 2018 Worksite Survey was noticeably lower than that achieved in the 2012 survey. The weighted AAPOR RR(3) for the 2018 survey was 7.0 percent, compared to 20.9 percent for the 2012 survey. However, the drop-off from the 2012 survey to the 2018 survey, a 28 percentage point decline, was not as steep as it was from 2000 to 2012. The weighted AAPOR RR(3) for the 2000 survey was 65.0 percent, compared to 20.9 percent for the 2012 survey, or a 44 percentage point decline.

One potential reason for the recent decline is that we did not conduct a separate pre-screening of the sample as we did in the 2012 survey. In 2012, we placed an initial call to each worksite to verify the firm existed and attempt to identify a survey respondent; we then used the data collected in this verification call to target the respondent for the 2012 Worksite Survey. In 2018, we believed we could combine the resources required for this separate verification call directly into the initial contact call to complete the survey. Also accuracy and completeness of the enhanced DMI contact data from Dun & Bradstreet did not meet the assumptions we developed based on our early communications with the vendor.

Another factor that we suspect led to a lower response rate in 2018 was overall and continuing decline in response rates, even for worksite surveys (IGEN, 1998; Petroni et al., 2004). In the 2018 survey, 26 percent of worksites were ineligible, mostly due to non-working phone numbers, compared to only 10 percent ineligible in 2012. In 2018, we could not determine eligibility for 22 percent due to non-contact dispositions such as “No answer” and “Always busy,” compared to 12 percent unknown eligibility in 2012. Even for those that were eligible, in 2018 we could not get back in touch with 36 percent of worksites we initially reached, compared to only 11 percent in 2012. After confirming eligibility, either the respondent refused to complete the survey or we reached an answering device or we were unable to get the respondent back on the phone to attempt to complete the survey. There were far fewer cases with an “out of business” disposition this time, only 2 percent, compared to 13 percent in 2012. This could have been a function of being unable to make contact with the worksite at all or of economic conditions (where the DMI information was legitimately out of date in 2012 because of the economy-wide recession versus the growth period as of 2018).

Neither of these hypothesized mechanisms of non-response in 2018 (lack of sample verification or inability to reach worksites) would necessarily be expected to systematically bias estimates from the Worksite Survey. That said, the lower response rate obtained in the Worksite Survey does suggest the potential for non-response bias to undermine survey estimates. The analysis below provides an empirical investigation of the potential risk posed by non-response bias. The approaches used to evaluate non-response in the Worksite Survey are a comparison of easier-to-reach versus harder-to-reach worksites and response propensity modeling.

2.5.1 Comparison of Easier-to-Reach versus Harder-to-Reach Worksites

In this analysis, worksites that were harder to reach in the Worksite Survey are compared to those that were easier to reach. The more difficult cases serve as proxies for the worksites 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 non-response bias with respect to the estimate at hand. As discussed in Chapter 1 for the Employee Survey (Subsection 1.5.3 Comparison of Easier-to-Reach versus Harder-to-Reach Respondents), 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 non-response bias.

Despite its limitations, analyzing level of effort is a standard approach to evaluate non-response bias (Halbesleben & Whitman, 2013; Maitland et al., 2017; Montaquila & Olson, 2012; McFarlane et al., 2007). In this analysis, the easy-to-reach versus hard-to-reach dimension was defined as the total number of the calls to the worksite to obtain the completed interview. Unlike in the Employee Survey, we attempted no refusal conversion with cases that refused to participate in the Worksite Survey. The number of call attempts made to Worksite Survey respondents ranged from zero to 12. There were 506 (25.4 percent) respondents with zero call attempts that represent instances when the extended interview was completed via the web before the start of the CATI non-response follow-up phase.

As shown in **Exhibit 2.9,** the mean number of attempts for the total responding sample was 2.44. The exhibit also presents the mean number of call attempts for responding worksites grouped by worksite size (number of employees), FMLA coverage status, whether any of the workforce was unionized, workforce gender ratio (percentage of female employees), and industry (NAICS code). We performed two types of statistical testing with these group means. First, we conducted bivariate tests (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.[32]

Exhibit 2.9. Mean number of attempts by worksite 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?
Worksite Size[a]		Yes	NA
9 or fewer employees	2.60		
10-249 employees	2.28		
250+ employees	2.72		
FMLA Coverage Status		Yes	Yes
FMLA covered	2.36		
Not FMLA covered	2.56		

[32] The log transformation reduced the skewness in the attempts distribution from 1.24 to 0.08, where zero represents no skewness.

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?
Workforce Unionization		Yes	Yes
Any employees unionized	2.95		
No employees unionized	2.37		
Percentage Female Workforce		Yes	No
0	2.28		
1-24.9	2.26		
25-49.9	2.25		
50-74.9	2.63		
75+	2.62		
Industry Type (NAICS code)[a]		Yes	NA
Manufacturing	2.04		
Retail	2.31		
Services	2.72		
Other	2.64		
Overall	**2.44**		

Source: 2018 Worksite Survey
[a] 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 2.9**. All of the variables are significant in these bivariate analyses (no control for multiple testing was performed, so the Type I error is inflated). In general, more attempts were required for the smallest and largest worksites (fewer than 10 employees and 250+ employees), worksites not covered by FMLA, worksites with a unionized workforce, worksites with more female employees, and worksites in the Services sector. Though these patterns are informative about the nature of non-response in the 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 non-response bias. Specifically, the weighting protocol included raking ratio adjustment to DOL's Quarterly Census of Employment and Wages (QCEW) population controls for region and size by industry.

The key question with respect to the risk from non-response 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 2.9** 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 the gender ratio of the workforce and number of call attempts disappears when controlling for size and industry. This result is not surprising given that the workforce gender distribution varies by industry.

The relationship between workforce unionization and the number of call attempts does remain significant ($p = .0006$) when controlling for size and industry. This suggests that the weighted survey estimates may underrepresent worksites with unionized workforces. When controlling for size and industry, the association between FMLA coverage status and the number of call attempts remains marginally significant ($p = .034$), suggesting that the weighted survey estimates may also somewhat underrepresent worksites not covered by FMLA. On balance, this level of effort analysis indicates that non-response did vary across key worksite subgroups, but there is evidence that the raking ratio adjustment to QCEW control totals likely reduced the potential for non-response 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 non-response bias in the Worksite Survey is a response propensity analysis that identifies factors associated with survey response. Many worksite 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:

- ***Screening contact response propensity***. This refers to contact with the worksite for the screening interview.
- ***Interview cooperation response propensity***. This refers to cooperation of the worksite for 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 include worksite size (measured as number of employees), industry (NAICS code), state, Census region, and various recodes and/or interactions of these factors.

Exhibit 2.10 below presents the estimated logistic regression parameters for the model predicting screening contact response propensity as a function of sampling frame variables for industry, size, and region. Specifically the model included terms for:

- *sector*, a 19-level categorical variable that provides more specific information than the industry grouping used in the stratification process;
- *size*, a categorized version of size of the worksite based on number of employees;
- *paid leave state indicator*, included as it was a stratification variable;
- *Census region*;
- *paid leave state indicator by industry interaction*, where "industry" used in this term is the four-level industry grouping used as a stratification variable;
- *paid leave state indicator by size interaction*; and
- *industry by size interaction*.

The main effect of *industry* (four levels) is also included in this table for completeness. Note that any row in this table with a missing beta indicates the term was not included in the model and/or it served as the reference level.

Exhibit 2.10 shows the unweighted response propensity, the sample size, the estimated logistic model parameter and its standard error (the Wald test statistic), and the *p*-value of the test that the coefficient is equal to zero.

Exhibit 2.10. Results from the screening contact propensity logistic model

Variable	Response Propensity[a]	Sample Size	Logistic Model Parameter	Parameter Standard Error	PR > ChiSq
Intercept	67.8***	31,277	1.0257	0.0936	.0000
Industry					
Manufacturing	60.4	6,071			
Retail	71.7	6,932			
Services	62.4	8,300			
Other	74.2	9,974			
Sector					
11. Agriculture, Forestry, Fishing and Hunting	60.9	448	0.4640	0.3324	.1627
21. Mining, Quarrying, and Oil and Gas Extraction	61.7	141	0.5472	0.3652	.1340
22. Utilities	64.2	204	0.2131	0.3758	.5706
23. Construction	58.8	2,095	0.3753	0.3228	.2450
31-33. Manufacturing	61.4*	3,389	0.5296	0.3218	.0999
42. Wholesale Trade	67.6	1,714	0.4380	0.3483	.2086
44-45. Retail Trade	75.4*	3,910	0.8343	0.3478	.0165
48-49. Transportation and Warehousing	66.2	1,102	0.3685	0.3521	.2953
51. Information	58.2	887	0.3817	0.2837	.1784
52. Finance and Insurance	67.5**	1,519	0.7670	0.2803	.0062
53. Real Estate and Rental and Leasing	63.1*	897	0.6519	0.2810	.0203
54. Professional, Scientific, and Technical Services	61.5*	2,640	0.5927	0.2768	.0323
55. Management of Companies and Enterprises	65.3*	150	0.5986	0.3248	.0653
56. Administrative and Support and Waste Management and Remediation Services	61.6*	2,238	0.5513	0.2759	.0457
61. Educational Services	80.7***	3,076	0.6531	0.0779	.0000
62. Health Care and Social Assistance	73.7***	3,005	0.3011	0.0691	.0000
71. Arts, Entertainment, and Recreation	57.5**	464	−0.3867	0.1079	.0003
72. Accommodation and Food Services	76.7***	1,558	0.4831	0.0806	.0000
81. Other Services (except Public Administration)	65.5	1,840			
Employment Size					
1-49 employees	65.6***	11,813	−0.5291	0.0801	.0000
50-249 employees	69.2***	8,328	−0.3431	0.0823	.0000
250-999 employees	62.4***	5,897	−0.8437	0.0865	.0000
1,000+ employees	76.8	5,239			
Paid Leave State Status					
FMLA paid leave states (CA, NJ, RI)	68.7***	10,678	0.7115	0.0908	.0000
Other states	67.4	20,599			
Region					
Northeast Region	67.7	6,340	0.0252	0.0375	.5024
Midwest Region	69.2**	5,363	0.1446	0.0462	.0018
South Region	67.0	8,787	0.0514	0.0419	.2197
West Region	67.8	10,787			

Variable	Response Propensity[a]	Sample Size	Logistic Model Parameter	Parameter Standard Error	PR > ChiSq
Paid Leave States by Industry					
Paid leave states, Manufacturing	59.4***	1,898	−1.6603	0.3361	.0000
Paid leave states, Retail	72.2*	2,434	−0.7929	0.3671	.0308
Paid leave states, Services	61.8**	2,902	−0.8272	0.2886	.0041
Paid leave states, Other Industry	77.1	3,444			
Other states, Manufacturing	60.9***	4,173	−1.3593	0.3309	.0000
Other states, Retail	71.4*	4,498	−0.6381	0.3634	.0791
Other states, Services	62.7*	5,398	−0.5975	0.2837	.0352
Other states, Other Industry	72.6	6,530			
Paid Leave States by Size					
Paid leave states, 1-49 employees	65.4***	4,016	−0.4989	0.0867	.0000
Paid leave states, 50-249 employees	69.9***	2,915	−0.4461	0.0918	.0000
Paid leave states, 250-999 employees	61.6***	2,004	−0.5330	0.0958	.0000
Paid leave states, 1,000+ employees	82.3	1,743			
Other states, 1-49 employees	65.7	7,797			
Other states, 50-249 employees	68.9	5,413			
Other states, 250-999 employees	62.8	3,893			
Other states, 1,000+ employees	74.1	3,496			
Industry by Size					
Manufacturing, 1-49 employees	59.9***	2,240	0.8203	0.1167	.0000
Manufacturing, 50-249 employees	66.2***	1,690	0.8252	0.1161	.0000
Manufacturing, 250-999 employees	54.6***	1,217	0.8475	0.1209	.0000
Manufacturing, 1,000+ employees	58.9	924			
Retail, 1-49 employees	69.8*	2,886	0.2747	0.1335	.0397
Retail, 50-249 employees	70.0	2,097	0.0783	0.1366	.5667
Retail, 250-999 employees	75.9***	1,376	0.9215	0.1447	.0000
Retail, 1,000+ employees	77.0	573			
Services, 1-49 employees	61.9	3,025	−0.0417	0.1045	.6898
Services, 50-249 employees	59.0**	1,878	−0.3642	0.1088	.0008
Services, 250-999 employees	52.3	1,701	−0.1125	0.1119	.3149
Services, 1,000+ employees	77.1	1,696			
Other industry, 1-49 employees	68.8	3,662			
Other industry, 50-249 employees	77.8	2,663			
Other industry, 250-999 employees	67.2	1,603			
Other industry, 1,000+ employees	84.6	2,046			

Source: 2018 Worksite Survey
Model Diagnostics: Area under ROC curve (c) = 0.6269. -2 Log Likelihood = 37,943.1. Hosmer and Lemeshow Goodness-of-Fit Chi-Squared = 15.9807 (8 d.f.). Probability > Hosmer and Lemeshow Chi-Squared = 0.0427.
[a] Response propensity is the unweighted, unadjusted response rate in this column.
*** p<.0001 ** p<.01 * p<.05

These results suggest that many of the sectors are significant predictors of response propensity, even with the presence of other variables in the model. Sector 71 (Arts, Entertainment, and Recreation) has the lowest response rate, as manifested by the only negative model coefficient versus the reference category

(Sector 81. Other Services). Response propensity for this group is 57.5 percent, compared to the overall response propensity of 67.8 percent. At the other end, Sector 44-45 (Retail Trade) had the largest positive coefficient and a response rate of 75.4 percent. The sector with the largest unconditional response rate was Sector 61 (Educational Services), with an observed response rate of 80.7 percent.

Employment size was a significant predictor of non-response. That all model coefficients are negative means that they had response rates lower than that of the reference cell (1,000+ employees), which had the highest response rate, at 76.8 percent (compared to the lowest response rate of 62.4 percent of worksites with 250-999 employees).

The paid leave states indicator was significant, as well as most of the interaction terms it participates in.

Exhibit 2.11 shows results from fitting a logistic model to interview cooperation. The model independent variables considered are equivalent to those noted in **Exhibit 2.10**. The sample size is quite a bit smaller at 5,875, as this model is applicable only to the worksites that had completed the screener. The overall cooperation rate of 37.4 percent is also smaller than the screening contact rate of 67.8 percent. The main interview asks detailed questions on leave allowances, usage, and implementation and is a bigger burden for the respondent, so seeing a lower response rate is not surprising.

Interestingly, Sector 61 (Educational Services) had the largest screening contact rate in **Exhibit 2.10** but the smallest interview cooperation rate at 27.8 percent in **Exhibit 2.11**.

Employment size again proved to be a significant predictor, but the pattern of relative response rate switched. **Exhibit 2.10** shows how the largest employers had the highest screening contact rate; **Exhibit 2.11** shows this same group had the lowest interview cooperation rate, at just 10.7 percent. One explanation for this pattern is that the reporting burden may have been less for smaller worksites than for larger ones. Small worksites may have fewer benefits and policies to report, resulting in a faster and less challenging interview relative to larger establishments that have complex leave benefits and policies or more numerous employees who would have taken their leaves or both, all of which had to be summarized by the human resources representative for the interview.

Effects of paid leave states indicator and region proved not to be significant. Neither was the paid leave state indicator by employment size interaction term. The other two interactions terms, paid leave by industry and industry by size, did show significance for several of their levels.

It is worth noting that the model fit statistics, such as the Hosmer-Lemeshow chi-square tests, suggest the screening contact propensity model is not fitting the screener data very well, whereas the model does seem to fit the interview cooperation rate data. Arguably the most interesting finding from this analysis is that many times response propensity patterns observed at screening seem to go in the opposite direction observed at the interview stage of data collection.

All of these variables considered in these models were also used and/or considered in the weight adjustment process, so we cannot simply conclude that associations seen with this analysis suggest that the final estimates from the Worksite Survey suffer from non-response bias to any extent. One limitation of this analysis is that it does not measure the potential for residual non-response bias after having made the weight adjustments. This response propensity analysis would have been more informative if there had been more relevant information available for the non-responding and responding worksites 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 non-response. This lack of relevant information from sampling frames and other sources is generally what motivates survey data collection in the first place.

Exhibit 2.11. Results from the interview cooperation response propensity logistic model

Variable	Response Propensity[a]	Sample Size	Logistic Model Parameter	Parameter Standard Error	PR > ChiSq
Intercept	37.4***	5,875	−1.6928	0.2187	.0000
Industry					
Manufacturing	44.4	1,316			
Retail	32.9	1,315			
Services	36.8	1,517			
Other	36.1	1,727			
Sector					
11. Agriculture, Forestry, Fishing and Hunting	57.6**	66	1.7120	0.4231	.0001
21. Mining, Quarrying, and Oil and Gas Extraction	47.1*	34	1.1191	0.4983	.0247
22. Utilities	46.2**	39	1.6734	0.5018	.0009
23. Construction	41.1*	360	0.7918	0.3557	.0260
31-33. Manufacturing	44.8**	858	1.3879	0.3465	.0001
42. Wholesale Trade	40.5**	400	1.0513	0.3715	.0046
44-45. Retail Trade	26.8	642	0.3379	0.3710	.3625
48-49. Transportation and Warehousing	33.6*	232	0.8420	0.3912	.0314
51. Information	32.8***	134	1.5384	0.3566	.0000
52. Finance and Insurance	39.2***	309	2.0507	0.3252	.0000
53. Real Estate and Rental and Leasing	36.9**	168	1.2638	0.3260	.0001
54. Professional, Scientific, and Technical Services	41.8***	500	1.6746	0.3067	.0000
55. Management of Companies and Enterprises	41.7***	36	2.2668	0.4894	.0000
56. Administrative and Support and Waste Management and Remediation Services	34.7***	401	1.5051	0.3008	.0000
61. Educational Services	27.8	133	−0.0992	0.2397	.6791
62. Health Care and Social Assistance	35.9	852	0.2380	0.1491	.1105
71. Arts, Entertainment, and Recreation	49.3*	75	0.6432	0.2714	.0178
72. Accommodation and Food Services	30.9	343	−0.1492	0.1748	.3933
81. Other Services (except Public Administration)	36.2	293			
Employment Size					
1-49 employees	50.2***	2,017	1.3319	0.1821	.0000
50-249 employees	42.6***	1,829	1.4810	0.1890	.0000
250-999 employees	28.4**	1,065	0.6956	0.2174	.0014
1,000+ employees	10.7	964			
Paid Leave State Status					
FMLA paid leave states (CA, NJ, RI)	32.8	1,892	−0.2779	0.2481	.2627
Other states	39.6	3,983			
Region					
Northeast region	37.9	1,186	0.0711	0.0878	.4180
Midwest region	41.5	1,129	0.0982	0.1029	.3402
South region	37.5*	1,599	−0.1681	0.0967	.0821
West region	34.7	1,961			

Variable	Response Propensity[a]	Sample Size	Logistic Model Parameter	Parameter Standard Error	PR > ChiSq
Paid Leave States by Industry					
Paid leave states, manufacturing	40.5***	338	−1.9730	0.4781	.0000
Paid leave states, retail	25.1**	455	−1.6443	0.5152	.0014
Paid leave states, services	34.8***	540	−2.7751	0.4001	.0000
Paid leave states, other industry	32.6	559			
Other states, manufacturing	45.7***	978	−2.0355	0.4645	.0000
Other states, retail	37.0**	860	−1.3348	0.5030	.0080
Other states, services	38.0***	977	−2.7664	0.3834	.0000
Other states, other industry	37.8	1,168			
Paid Leave States by Size					
Paid leave states, 1-49 employees	47.1	646	0.1087	0.2557	.6707
Paid leave states, 50-249 employees	37.8	574	0.0354	0.2598	.8916
Paid leave states, 250-999 employees	21.0	333	−0.1845	0.2857	.5186
Paid leave states, 1,000+ employees	8.8	339			
Other states, 1-49 employees	51.7	1,371			
Other states, 50-249 employees	44.9	1,255			
Other states, 250-999 employees	31.7	732			
Other states, 1,000+ employees	11.7	625			
Industry by Size					
Manufacturing, 1-49 employees	62.2***	362	1.8522	0.3667	.0000
Manufacturing, 50-249 employees	47.7**	512	0.9718	0.3601	.0070
Manufacturing, 250-999 employees	35.8**	288	1.2308	0.3847	.0014
Manufacturing, 1,000+ employees	7.8	154			
Retail, 1-49 employees	42.7*	494	0.8863	0.3886	.0226
Retail, 50-249 employees	32.7	471	0.3195	0.3929	.4162
Retail, 250-999 employees	24.6*	232	0.7094	0.4228	.0934
Retail, 1,000+ employees	8.5	118			
Services, 1-49 employees	58.9***	489	1.9822	0.3077	.0000
Services, 50-249 employees	44.0**	405	1.2052	0.3127	.0001
Services, 250-999 employees	25.3**	292	1.1453	0.3402	.0008
Services, 1,000+ employees	5.7	331			
Other industry, 1-49 employees	43.0	672			
Other industry, 50-249 employees	46.3	441			
Other industry, 250-999 employees	26.9	253			
Other industry, 1,000+ employees	17.2	361			

Source: 2018 Worksite Survey
Model Diagnostics: Area under ROC curve (c) = .7101. -2 Log Likelihood = 6,929.4. Hosmer and Lemeshow Goodness-of-Fit Chi-Squared = 3.3860 (8 d.f.). Probability > Hosmer and Lemeshow Chi-Squared = 0.9079.
[a] Response propensity is the unweighted, unadjusted response rate in this column.
*** p<.0001 ** p<.01 * p<.05

2.5.3 Summary of Non-Response Analysis for the Worksite Survey

The analysis of the 2018 Worksite Survey found limited evidence that non-response poses a threat to the survey estimates. A limitation of this analysis is the small set of demographic variables available. In the level of effort analysis, two bivariate associations between level of effort and survey outcomes were significant when accounting for the survey weighting variables. The significant association between workforce unionization and the number of call attempts suggests that worksites with unionized workforces were harder to reach, and thus potentially underrepresented. The marginally significant association between FMLA coverage status and the number of call attempts suggests relative difficulty of interviewing the worksites covered by FMLA. However, the raking ratio adjustment to QCEW control totals likely reduce the potential for non-response bias in a number of the survey estimates.

The findings in the response propensity analysis suggest many industry sectors are significant predictors of response propensity as well as worksite size. Worksite size is positively associated with contact but negatively associated with cooperation. Several of the response propensity patterns observed at screening go in the opposite direction at cooperation. Though interesting, these findings do not necessarily represent a threat from non-response bias because differential non-response across these groups was accounted for in the weighting in a within-stratum non-response adjustment as well as raking ratio estimation.

2.6. Weighting

The weights for the 2018 Worksite Survey were designed to adjust for several key factors: differential probabilities of selection across sampling strata, differential non-response across sampling strata, and any residual difference between the distribution of the weighted respondents and the target population (post-stratification). The following section summarizes how these weight adjustment factors were computed.

2.6.1 Base Weight for Probability of Selection

As discussed in Section 2.1, the Worksite Survey employed a stratified simple random design, with strata defined by paid leave states/non-paid leave states, industry grouping, and worksite size. Given this design, the base weight for all of the sample units in a given stratum was equal to the stratum population size on the DMI file sampling frame divided by the number of sample units in the stratum.

The final sample weight for Worksite Survey respondents is composed of a product of several factors. This base weight is the first factor in the product.

2.6.2 Non-Response Adjustment and Post-Stratification

We made several adjustments to the base sampling weight to account for different types of non-response and to account for differences between the sample frame and the target population of interest. This latter adjustment (i.e., adjusting the sample weights so that the weighted sample reflects the target population of interest) is referred to as a *post-stratification* adjustment to the sample weights. The post-stratification adjustment was the last adjustment we made to the sample weights, applied so that the weighted sample would sum to control totals obtained from the 2018 QCEW.

We created both the non-response adjustments and the post-stratification adjustments using a calibration technique that involves fitting a generalized exponential model (GEM; see for example, Folsom & Singh, 2000; Folsom & Witt, 1994; Witt, 2009). This method offers numerous advantages to performing weight adjustments including these:

- The GEM model parameters are estimated using calibration equations. This means if a solution is found when estimating the model parameters, the adjusted weights that come from the GEM model will equal the correct control totals exactly. This important and desirable feature is not achieved when we use a simple logistic model to derive a weight adjustment (for example) because the score functions that are solved with a logistic model are not calibration equations.
- The GEM calibration model can be used to compute a non-response or a post-stratification (raking) adjustment. Separate models are estimated to get the two adjustments.

- We can use more variables in the adjustment process than what can be used with a standard, ratio-weighting-class type of adjustment. The use of a greater number of variables can reduce the non-response and coverage biases associated with the final estimates.
- Because adjustments are created using a modeling approach, we can test for and include the statistically significant predictors for each adjustment. This is particularly appealing for a non-response adjustment. Including covariates that are not significant predictors of response propensity can cause an undesirable increase in the variability of the adjusted weights.
- We do not need to include higher-order interactions of variables in the adjustment, as we would need to do with a standard ratio adjustment. For example, we may find that the main effects of geography and NAICS grouping to be a significant predictor of response propensity at some stage in the weighting process, but not the interaction of geography and NAICS.
- The GEM model allows users to bind the magnitude of the weights adjustment. This essentially gives this methodology a built-in weight trimming feature.
- Both categorical and continuous variables can be included in the GEM model. This is particularly advantageous for the Worksite Survey, because it allows us to create weights appropriate for the worksite-level analysis. Moreover, by including terms that reflect the total employees, we can simultaneously create a sample weight appropriate for stakeholders interested in the total number of employees in the target population that have various characteristics collected in the Worksite Survey.
- This calibration, model-based approach to doing weight adjustment is available in SUDAAN (Research Triangle Institute, 2012) as one its SAS callable procedures.

For the Worksite Survey, we created multiplicative adjustments to the base weight sequentially as follows:

#1 Base Sampling Weight

#2 Screener Non-Status Adjustment. This weight adjustment accounted for those employers that did not respond and the eligibility of which for this study is unknown. This adjustment was created using the GEM approach. The model included the following variables:

- *NAICS sector* (SECTOR);
- *total employees* (EMPTOT);
- *paid leave states indicator* (GEOG);
- *NAICS design strata* (NAICS);
- *employee size strata* (SIZE);
- and various interactions of the above variables:

  GEOG*NAICS
  GEOG*SIZE
  NAICS*SIZE
  GEOG*NAICS*SIZE
  EMPTOT*GEOG
  EMPTOT*NAICS
  EMPTOT*SIZE
  EMPTOT*GEOG*NAICS
  EMPTOT*GEOG*SIZE
  EMPTOT*NAICS*SIZE
  EMPTOT*GEOG*NAICS*SIZE

#3 **Screener Non-Response Adjustment.** This weight adjustment accounted for those employers that were eligible but did not respond to the screening process. This adjustment was created using the GEM approach. The model included the following variables:

- *Census division* (DIVISION);
- *NAICS sector* (SECTOR);
- *indicator for whether the employer had an email address on the frame or not* (GOTEMAIL);[33]
- *total employees* (EMPTOT);
- *paid leave states indicator* (GEOG);
- *NAICS design strata* (NAICS);
- *employee size strata* (SIZE);
- and various interactions of the above variables:

    GEOG*NAICS
    GEOG*SIZE
    NAICS*SIZE
    GEOG*NAICS*SIZE
    EMPTOT*GEOG
    EMPTOT*NAICS
    EMPTOT*SIZE
    EMPTOT*GEOG*NAICS
    EMPTOT*GEOG*SIZE
    EMPTOT*NAICS*SIZE
    EMPTOT*GEOG*NAICS*SIZE

#4 **Interview Non-Response Adjustment.** This weight adjustment accounted for those employers that did not complete the interview. This adjustment was created using the GEM approach. The model included the following variables:

- *total employees* (EMPTOT)
- *paid leave states indicator* (GEOG)
- *NAICS design strata* (NAICS)
- *employee size strata* (SIZE);
- and various interactions of the above variables:

    GEOG*NAICS
    GEOG*SIZE
    NAICS*SIZE
    EMPTOT*GEOG
    EMPTOT*NAICS
    EMPTOT*SIZE
    EMPTOT*NAICS*SIZE
    EMPTOT*GEOG*SIZE
    EMPTOT*GEOG*NAICS

33 This was included in the model because we speculated it might be highly correlated with response propensity. And it proved to be a significant predictor. At this stage in the weighting process, the weighted response rate for those worksites on the frame with an email address was 31.0 percent, compared to 21.7 percent for those that did not have an email address.

#5 **Weight Trimming**. A small amount of weight trimming was applied to further reduce the effects of unequal weighting. A large amount of unequal weighting can cause a decrease in the precision of estimates. Weight trimming was done by sampling design strata.

#6 **Final Post-Stratification Adjustment**. This weight adjustment was applied so that weighted sums from the Worksite Survey would match DOL's 2018 Quarterly Census estimates. This adjustment was created using the GEM approach. The model included the following variables:

- *total employees* (EMPTOT);
- *paid leave states indicator* (GEOG);
- *NAICS design strata* (NAICS);
- *employee size strata* (SIZE);
- and various interactions of the above variables:

  GEOG*NAICS
  GEOG*SIZE
  NAICS*SIZE
  EMPTOT*GEOG
  EMPTOT *NAICS
  EMPTOT*SIZE
  EMPTOT*NAICS*SIZE
  EMPTOT *GEOG*SIZE
  EMPTOT*GEOG*NAICS

The final worksite sample weight for the Worksite Survey is defined as the product of the above six factors. And the final sample weight to use for an employee-level analysis is defined as the product of the final worksite weight and the worksite's total number of employees (i.e., EMPTOT). On the final analytic file, these two weights are ESTAB_WT and EMP_WT. Note that because EMPTOT was used in the weight adjustments noted above, both the final weighted worksite sum and the final weighted employee sum will ultimately match control totals from DOL's 2018 Quarterly Census.

2.7. Variance Estimation

To account for the complex design of the 2018 Worksite Survey, we used a re-sampling technique (specifically the *bootstrap* method) to create replicate weights for this study. This involves repeatedly taking independent, random subsamples of the original selected sample and creating a weight for each subsample using the same methodology as was used to create the weights for the final study. The bootstrap method for variance estimation is discussed by Kolenikov (2010).

A total of 250 bootstrap replicates were created. The replicate weights are given by variables

- BootStrap_Estab_Wt1 − BootStrap_Estab_Wt250 (for worksite-level analyses); and
- BootStrap_Emp_Wt1 − BootStrap_Emp_Wt250 (for employee-level analyses).

The replicate weight can be used with SAS's standard survey procedures. One way to get bootstrap variance estimates from the SAS is as follows (for example):

```
Proc SurveyMeans VarMethod=JACKKNIFE;
Weight     ESTAB_WT;
RepWeights BOOTSTRAP_ESTAB_WT1-BOOTSTRAP_ESTAB_WT250 / JKCOEFS=0.004;
                                                       /*JKCOEFS is 1/250*/
```

An equivalent and somewhat more compact specification exploits the analogy between balanced repeated replication and the bootstrap:

```
Proc SurveyMeans VarMethod=BRR;
     WEIGHT   EMP_WT;
     REPWEIGHTS BootStrap_Emp_Wt1-BootStrap_Emp_Wt250;
```

We computed statistical significance tests presented in the report 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 worksites). 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), and y_i be the employee-level characteristic of interest (e.g., percentage of unionized employees). In this example, $z_i = y_i\, e_i$. The population percentage of unionized employees is then

$$\theta = \frac{\sum_{i\in U} z_i}{\sum_{i\in U} e_i} = \frac{\mathrm{T}[z]}{\mathrm{T}[e]} \tag{1}$$

$$= \frac{\sum_{i\in U} e_i y_i}{\sum_{i\in U} e_i} = \frac{\mathrm{T}[ey]}{\mathrm{T}[e]} \tag{2}$$

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

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

$$= \frac{\sum_{i\in S} w_i e_i y_i}{\sum_{i\in S} w_i e_i} = \frac{\mathrm{t}[ey]}{\mathrm{t}[e]} \tag{4}$$

where $\mathrm{t}[z]$ 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

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

where $\mathrm{v}[\cdot]$ 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 $\mathrm{t}[z]$, $\mathrm{t}[e]$ can be used to obtain

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

(Korn & Graubard, 1999, formula 2.4.7). 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; that is, a worksite-level variable (e.g., the total number 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 than as the ratio $\mathrm{t}[ey]/\mathrm{t}[e]$. 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). Design effects based on these expansion weights are incorrect, however, as the comparison design is the one with one employee per worksite (rather than an actual cluster sample of all employees, as obtained in the administrative records of the worksite).

As noted above, a separate sample weight is provided for those interested in doing employee-level analyses using the Worksite Survey data. The variable containing this weight is called EMP_WT. And the corresponding replicate weights to use for variance estimation are BOOTSTRAP_EMP_WT1 − BOOTSTRAP_EMP_WT250.

The continuous variable EMPTOT, which is the total number of employees in each worksite, was included in the weight adjustment process discussed in Section 2.6/ Weighting above. Because of the calibration equations used to estimate model parameters, one desirable by-product of the weighting methodology we used is that the final sum of the respondents' final worksite weight will sum to the total number of worksites in the target population. The sum of the respondents' final worksite weight × EMPTOT will sum to the total number of employees in the target population. EMP_WT is therefore the product of ESTAB_WT and EMPTOT.

References

American Association for Public Opinion Research (AAPOR). (2016). *Standard definitions.* Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

American Association for Public Opinion Research (AAPOR). (2017). *The future of U.S. general population telephone survey research.* Task Force Report. Retrieved from https://www.aapor.org/Education-Resources/Reports/The-Future-Of-U-S-General-Population-Telephone-Sur.aspx

American Association for Public Opinion Research (AAPOR). (2019). *Report of the AAPOR Task Force on Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys.* Task Force Report. Retrieved from https://www.aapor.org/getattachment/Education-Resources/Reports/Report-of-the-Task-Force-on-Transitions-from-Telephone-Surveys-FULL-REPORT-FINAL.pdf.aspx?utm_source=AAPOR-Informz&utm_medium=email&utm_campaign=default

Barron, M., LeClere, F., Montgomery, R., Greby, S., & Kennedy, E. D. (2015, February). The accuracy of small area sampling of wireless telephone numbers. *Survey Practice, 8* (2). Retrieved from http://www.surveypractice.org/index.php/SurveyPractice/article/view/281

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57* (1), 289-300. Retrieved from http://www.jstor.org/stable/2346101

Bethlehem, J. G. (2002). Weighting non-response adjustments based on auxiliary information. In R. Groves, D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey non-response* (pp. 275–287). New York: Wiley.

Blumberg, S. J., & Luke, J. V. (2015, December). *Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2015.* National Center for Health Statistics Technical Report. Washington, DC: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Division of Health Interview Statistics. https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201512.pdf

BRFSS combined landline and cell phone weighted response rates by state. (2018). Retrieved from https://www.cdc.gov/brfss/annual_data/2018/pdf/2018-response-rates-table-508.pdf

Brick, J., & Tourangeau, R. (2017). Responsive survey designs for reducing nonresponse bias. *Journal of Official Statistics, 33*, 735-752. Retrieved from doi:10.1515/JOS-2017-0034

Chang, T., & Kott, P. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika*, *95*, 555-571. Retrieved from https://www.jstor.org/stable/20441486

Daley, K., Kennedy, C., Schalk, M., Pacer, J., Ackermann, A., Pozniak, A., & Klerman, J. A. (2013). *Family and Medical Leave in 2012: Methodology report*. Cambridge, MA: Abt Associates.

de Leeuw, E. D., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, A. D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse* (pp. 41-54). New York: Wiley.

Dillman, D., Smyth, J., & Christian, L. (2014). *Internet, phone, mail and mixed-mode surveys: The tailored design method* (4th ed.). New York: Wiley.

Dutwin, D., & Lavrakas, P. (2016). Trends in telephone outcomes, 2008–2015. *Survey Practice, 9*(3). Retrieved from https://doi.org/10.29115/SP-2016-0017

Dutwin, D., Sherr, S., Goyle, A., Langdale, K., & Best, J. (2018). *California Health Interview Survey. CHIS Methodology Series, Report 4, Response rates*. Retrieved from https://healthpolicy.ucla.edu/chis/design/Documents/CHIS_2017_MethodologyReport4_ResponseRates.pdf

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, non-response, and poststratification. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 598-603). Alexandria, VA: American Statistical Association.

Folsom, R. E., & Witt, M. (1994). Testing a new attrition non-response adjustment method for SIPP. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 428-433). Alexandria, VA: American Statistical Association.

Groves, R. M. (2006). Non-response rates and non-response bias in household surveys. *Public Opinion Quarterly, 70,* 646-675.

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

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-salience theory of survey participation: Description and an illustration. *Public Opinion Quarterly, 64,* 299-308.

Halbesleben, J. R. B., & Whitman, M. V. (2013). Evaluating survey quality in health services research: A decision framework for assessing nonresponse bias. *Health Services Research, 48,* 913-930.

Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of internet panels to conduct surveys. *Behavior Research Methods, 47,* 685-690. Retrieved from https://dx.doi.org/10.3758%2Fs13428-015-0617-9

Haziza, D., Lesage, E. (2016). A Discussion of Weighting Procedures for Unit Nonresponse. *Journal of Official Statistics*, 32, 129-145. Retrieved from https://doi.org/10.1515/jos-2016-0006

IGEN (Interagency Group on Establishment Survey Nonresponse). (1998). Establishment Nonresponse: Revisiting the Issues and Looking to the Future. *Statistical Policy Working Paper 28* (pp. 181-227). Washington, DC: Office of Management and Budget.

Kalton, G., & Anderson, D. (1986). Sampling rare populations. *Journal of the Royal Statistical Society, 149,* 65-82. doi:10.2307/2981886

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

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing non-response in a national telephone survey. *Public Opinion Quarterly, 64,* 125-150.

Kolenikov, S. (2010). Resampling variance estimation for complex survey data. *The Stata Journal, 10* (2), 165-199.

Korn, E. L., & Graubard, B. I. (1999). *Analysis of health surveys*. New York: Wiley.

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

Lohr, S. (2009). *Sampling: Design and analysis* (2nd ed.). Pacific Grove, CA: Duxbury Press.

Lundstrom, S., & Sarndal, C. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics, 15,* 305-327. Retrieved from https://pdfs.semanticscholar.org/c9c8/5f1261db5b94a7593bfdcf72932933d293d0.pdf

Maitland, A., Lin, A., Cantor, D., Jones, M., Moser, R. P., Hesse, B. W., Davis, T., & Blake, K. D. (2017). A nonresponse bias analysis of the Health Information National Trends Survey (HINTS). *Journal of Health Communication*, *22*(7), 545–553.

McFarlane, E., Olmsted, M. G., Murphy, J., & Hill, C. A. (2007). Nonresponse bias in a mail survey of physicians. *Evaluation & the Health Professions*, *30*(2), 170-185.

McGeeney, K. (2015). Appending a prepaid phone flag to the cellphone sample. *Survey Practice*, *8* (4). Retrieved from http://www.surveypractice.org/index.php/SurveyPractice/article/view/311

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R .J. A. Little (Eds.), *Survey nonresponse* (pp. 243-258*)*. New York: Wiley.

Montaquila, J., Brick, J., Hagedorn, M., Kennedy, C., & Keeter, S. (2008). Aspects of non-response bias in RDD telephone surveys. In J. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. Lavrakas, M. Link, & R. Sangster (Eds.), *Telephone survey methodology* (pp. 561-586). New York: Wiley.

Montaquila, J. M., & Olson, K. M. (2012, April). *Practical tools for nonresponse bias studies.* SRMS/AAPOR Webinar. Retrieved from https://higherlogicdownload.s3.amazonaws.com/AMSTAT/20d2b15c-9cc4-4c39-807c-088d6a8b6228/UploadedImages/WebinarFiles/NRBiasWebinarApril2012.pdf

Petroni, R., Sigman, R., Willimack, D., Cohen, S., & Tucker, C. (2004). Response rates and nonresponse in BLS and Census Bureau Establishment Surveys. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 4159-4166). Alexandria, VA: American Statistical Association.

The Pew Charitable Trusts. (2015, March). *Survey of American Family Finances: Detailed methodology*. Retrieved from https://www.pewtrusts.org/~/media/Assets/2015/03/FSM-Poll-Results-Methodology_ARTFINAL_v2.pdf

Pew Research Center. (2012, May 15). *Assessing the representativeness of public opinion surveys.* Retrieved from https://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys

McGeeney, K. & Kennedy, C. (2015, November 18). *Advances in telephone sampling.* Retrieved from http://pewrsr.ch/1Of5EqB

McGeeney, K. & Kennedy, C. (2016, October 24). *Cellphone activity flags: A trade-off between efficiency and coverage.* Retrieved from http://www.pewresearch.org/2016/10/24/cellphone-activity-flags/

Kennedy, C. & Hartig, H. (2019, February 2). *Response rates in telephone surveys have resumed their decline.* Retrieved from https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/

Putnam, R. D. (1995). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

Rao, J. N. K., Wu, C. F. J., & Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology, 18* (2), 209-217.

Research Triangle Institute. (2012). *SUDAAN language manual* (Vols. 1 and 2, Release 11). Research Triangle Park, NC: Author.

Ruther, N., & Sokolowski, J. (2020). *2019 Community Health Survey response rate memo.* Rockville, MD: Abt Associates.

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology, 35* (1), 101-113.

Singer, E. (2003). Exploring the meaning of consent: Participation in research and beliefs about risks and benefits. *Journal of Official Statistics, 19,* 273-285.

Smyth, J. D., Dillman, D., & Christian, L. M. (2009). *Context effects in Internet surveys: New issues and evidence.* Retrieved from https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199561803.001.0001/oxfordhb-9780199561803-e-027

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55,* 775-801.

Tourangeau, R. (2017). Presidential address: Paradoxes of nonresponse. *Public Opinion Quarterly, 81,* 803-814. Retrieved from doi:10.1093/poq/nfx031

Williams, D., & Brick, J. M. (2018). Trends in U.S. face-to-face household survey non-response and level of effort. *Journal of Survey Statistics and Methodology, 6* (2), 186-211. Retrieved from https://academic.oup.com/jssam/article/6/2/186/3930833

Witt, M. (2009). Overview of software that will produce sample weight adjustments. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 3009-3023). Alexandria, VA: American Statistical Association.

Yan, T., Curtin, R., & Jans, M. (2010). Trends in income nonrespones over two decades. *Journal of Official Statistics, 26,* 145-164.