
**Supporting
Statement for OMB
Clearance Request
Part B**

**Family and Medical
Leave Act, Wave 4
Surveys**

OMB Control Number 1290-
ONEW

July 2017

Table of Contents

Part B: Collection of Information Involving Statistical Methods.....1

- B.1. Respondent Universe and Sampling Methods.....1
 - B.1.1 Employee Survey.....1
 - B.1.2 Employer Survey.....8
- B.2. Analysis Plans.....12
 - B.2.1 Employee Survey.....12
 - B.2.2 Employer Survey.....17
 - B.2.3 Who Will Collect the Information and How It Will Be Done.....19
 - B.2.4 Use of Periodic Data Collection Cycles to Reduce Burden.....20
- B.3. Methods to Maximize Response Rates and Deal with Nonresponse.....20
 - B.3.1 Employee Survey.....20
 - B.3.2 Employer Survey.....26
- B.4. Tests Procedures.....28
 - B.4.1 Cognitive Testing of the Employee Survey.....28
 - B.4.2 User-Testing of the Employer Survey.....30
- B.5. Individuals Consulted on Statistical Aspects of the Design.....30
- References.....32

List of Attachments

Attachment A: Employee Survey

Attachment B: Employer Survey

Attachment C: Advance Letter for the Employer Survey

Attachment D: 60-Day Federal Register Notice

Attachment E: Summary of Public Comments and DOL Response

Attachment F: Privacy Statement

Attachment G: Weighting and Estimation Procedures

Part B: Collection of Information Involving Statistical Methods

As required under the Paperwork Reduction Act (PRA), the Chief Evaluation Office (CEO) at the U.S. Department of Labor (DOL) is seeking approval from the Office of Management and Budget (OMB) for a study of the federal Family and Medical Leave Act (FMLA). Consistent with its commitment to tracking and analyzing user feedback, DOL periodically conducts surveys of employees and employers who are covered by the provisions of the FMLA. Surveys were conducted in 1996 (DOL 1996), 2000 (Cantor et al. 2001), and again in 2012 (Klerman et al. 2013). DOL has contracted with Abt Associates Inc. (in partnership with Abt SRBI¹) to conduct a fourth “wave” of Employee and Employer Surveys.

This submission seeks clearance to administer two surveys:

- A survey of employees that primarily examines leave usage patterns and unmet need for leave (Employee Survey); and
- A survey of employers that examines their leave policies, as well as their experiences with and perceptions of the FMLA (Employer Survey).

The Wave 4 Employee and Employer Surveys will update and expand upon the knowledge gained from prior studies in the current policy landscape. Part B of the Supporting Statement considers the issues pertaining to Collection of Information Employing Statistical Methods.

B.1. Respondent Universe and Sampling Methods

Section B.1 discusses the respondent universe and sampling methods to be used in conducting the Wave 4 FMLA Employee and Employer Surveys.

B.1.1 Employee Survey

For the Wave 4 Employee Survey, this section discusses the universe of potential respondents, the sampling frame, respondent selection process, and expected response rates.

Potential Respondent Universe

Employees age 18 or older who live in the United States, have a telephone (landline or cell), and have been employed for pay (private or public sector) in the 12 months prior to the interview will constitute the respondent universe from which the sample for the Wave 4 Employee Survey will be taken.

According to the 2016 National Health Interview Survey, 97.1 percent of U.S. adults live in a household with landline or cellular telephone service. According to the March 2010 Current Population Survey (CPS), 59.7 percent of the entire civilian non-institutional adult population is employed (excluding self-employed). The estimated total size of the eligible respondent universe is thus 130.0 million adults.² Exhibit B.1 lists this total universe size, as well as Employer Survey sample sizes, both overall and by strata.

¹ Effective April 1, 2017, Abt SRBI will be consolidated into the existing Abt Associates Inc. government segment. Abt SRBI will cease to function as an independent entity and will become a part of the parent company. Abt is in the process of executing this transition, including integrating current Abt SRBI employees into the parent company.

² This calculation assumes no covariance between the proportion of the adult population with access to a phone, and the proportion of employed adults.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Exhibit B.1. Respondent Universe and Sample Sizes for the Wave 4 FMLA Employee Survey

	Universe Size (persons covered by data collection, '000s)	Landline Sample	Cellular Sample	Total Sample Size
All	130,013	578	3,422	4,000
Paid leave states ¹	22,924	127	760	887
Non-paid leave states	107,089	451	2,662	3,113

¹For the purposes of these calculations, we include the following paid leave states: California, New Jersey, Rhode Island, and Washington. (We are aware that the implementation of the Washington paid leave policy is stalled; if it has not been implemented by the time of the survey, we will adjust the survey accordingly.)

Sampling Frame

Respondents will be sampled through a dual frame, landline and cellular random digit dialing (RDD) telephone design. We project that 578 interviews (approximately 14 percent) will be completed with respondents sampled through the landline frame, and 3,422 interviews (approximately 86 percent) will be completed with respondents sampled through the cellular frame, for a total of 4,000 interviews.³

Numbers for the landline sample will be 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 will be drawn through systematic sampling from 1,000 blocks dedicated to cellular service according to the Telcordia database.

With both the landline and cellphone frames we will oversample states with paid family leave policies (“paid leave states”) at a 25 percent rate. On the landline frame, geographic targeting is achieved 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 blocks are found. On the cellphone frame, geographic targeting is achieved by pegging the frame entries (cellphone numbers) to “rate centers,” which are somewhat arbitrary yet well-defined telecom service areas. These types of geographic data are routinely available from sample vendors such as MSG or SSI. 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; Pierannunzi et al. 2013) reduce efficiency of oversampling, which we account 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 has been employed by The contractor in dozens of statewide, citywide, and other contained-geography surveys.⁴

³ Using the data on phone use trends from the National Health Interview Survey (Blumberg and Luke 2015), we estimate that in 2018, some 3.6 percent of the eligible population will have only a landline, 39.0 percent will be dual users (i.e., use both a landline phone and a cellphone), and 57.4 percent will be cellphone only. We also project that a completed landline interview will be approximately 35 percent more expensive than a cellphone interview due to the deterioration of the landline frame as the general population migrates to cellphones (Martonik et al. 2015). This higher incidence of workers in the cellphone frame leads to comparatively lower screening costs. Given these projections and the implied cost per complete, we will conduct 86 percent of the interviews in the cell RDD sample and 14 percent of the interviews in the landline RDD.

⁴ The contractor’s team has used these methods in several recent projects, including the FMLA Wave 3 Surveys (2012) and Worker Knowledge projects (2016) for the Department of Labor, the National Intimate Partner Violence Study (2015) for the Centers for Disease Control, and a number of Behavioral Risk Factor Surveillance Survey data collections for several states.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Consistent with the basic design, the calculations for the proposed sampling design included here assume an oversample of the following paid leave states: California, New Jersey, Washington, and Rhode Island.⁵ Combined, these states represent 17.6 percent of the total U.S. labor force.⁶ For the cellphone frame, the expected statewide sample sizes listed in Exhibit B.2 are given using undercoverage and overcoverage rates reported in Pierannunzi et al. (2013).⁷ Higher rates of frame deficiencies reduce efficiency of oversampling strategies.

Exhibit B.2. Oversampling of Paid Leave States in the Cellular Frame

Paid Leave States	% of Labor Force	Cell Frame Under-Coverage	Cell Frame Over-Coverage	Oversampling Rate	Expected Sample Size	Effective Oversampling Rate
CA	12.1%	18.0%	2.3%	25.0%	531.2	28%
NJ	3.0%	7.4%	11.5%	25.0%	119.2	19%
WA	2.3%	20.9%	6.2%	25.0%	96.2	25%
RI	0.4%	3.8%	17.4%	25.0%	13.3	13%
Other states	82.4%	7.6%	7.6%		2662.0	-6%

The sampling frame for the cellular sample also has two additional key features: (1) a 33 percent prepaid phone oversample (using prepaid cellphone flags, implemented via subsampling 75 percent of the non-prepaid phones), and (2) a 30 percent subsampling rate of inactive numbers (using “activity flags”).

- (1) The oversample of prepaid cellphones will provide an oversample of low-wage workers, a group that is traditionally difficult to reach and survey, and also a group that is of special interest to DOL. The low-income population is known to use prepaid cellphones at a rate higher than the general population (among prepaid phone users, 53 percent of respondents had annual household incomes below \$30,000, versus 24 percent among non-prepaid phone users; McGeeney 2015). Traditionally, oversampling to reach low-income populations has been achieved through geographic targeting of landline numbers in low-income Census tracts. Because cellphone numbers are not easily associated with geographies, the prepaid flag is a good option. We estimate that a 33 percent prepaid phone oversample will increase from 22.3 percent to 26.5 percent the expected sample representation of individuals living in households with incomes under \$30,000.
- (2) The subsampling of inactive numbers—rather than removing them from the sample entirely—will avoid coverage loss for the 8 percent of the population with inactive numbers (Pew Research Center 2015). A standard component of cellphone RDD design is the use of “activity flags” (Pew Research Center 2016), which indicate whether a given number has had any recent activity such as calls or text messages. Given that cellphone numbers flagged as active are approximately five times more productive than inactive numbers, the latter are often removed from the sample.⁸ We have determined that the subsampling rate that provides the optimal balance between the higher cost of data collection

⁵ We are aware that the implementation of the Washington paid leave policy is stalled; if it has not been implemented by the time of the survey, we will adjust the survey accordingly.

⁶ Abt Associates calculation, based on American Community Survey (ACS) 2014 Public Use Microdata Sample (PUMS) data.

⁷ These counts reflect the 75 percent undersampling of non-prepaid numbers and the 30 percent subsampling of inactive numbers (both discussed below), as well as the 25 percent oversampling of paid leave states.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

(higher subsampling rates) and higher design effects due to subsampling (lower rates) is 30 percent. For oversampling low-wage workers using prepaid cellphone activity flags, note that prepaid phones are always flagged active, so this rate is applied on top of the 75 percent subsampling rate of non-prepaid phones.

Once the cell sample is drawn, in this sampling stage we will first append the prepaid flag, then oversample/subsample as described above. Given these oversampling and subsampling rates on the cell RDD, for the 3,422 estimated cellphone interviews, we project 543 interviews on prepaid cellphones, 2,814 on non-prepaid cellphones flagged as active, and 65 interviews on cellphones flagged as inactive.⁹

Respondent Selection

The respondent selection process will oversample leave-takers (employees who took leave for a covered family or medical reason during the screening period) and leave-needers (employees who needed to take leave for a covered reason, but did not). As described below, the necessary information is collected early in the interview, which comprises a set of screening questions structured to identify sufficient subsamples of leave-takers and leave-needers. Given the survey patterns observed in the Wave 3 FMLA surveys conducted in 2012, and the respondent selection rates discussed below, we project completing 1,778 interviews with leave-takers, 422 interviews with leave-needers, and 1,800 interviews with “employed-only” respondents (those not having taken or needed leave).

The process of identifying a prospective survey respondent will vary depending on the phone frame—cellphone or landline—from which it is drawn.

Cellphone Calls

For the cellphone calls, the interviewer will ascertain that they are talking with an adult age 18 or older who has been employed (excluding self-employed) during the last 12 months. (As noted below in the discussion of expected response rates, our experience in the FMLA Wave 3 Employee Survey demonstrated that a within-household selection procedure on the cellphone sample is not cost-effective, thus we will treat cellphone calls as individual surveys, and landline calls as household surveys.) The interviewer will then determine if the individual has taken, or needed to take, family or medical leave during the reference period. The respondent will then be classified into one of three family or medical leave groups: leave-needer, leave-taker, or employed-only.

To generate sufficient sample sizes of leave-needers and leave-takers, these two groups will be selected at a higher rate than the employed-only group because their incidence rates are significantly lower. Past studies suggest that about 80 percent of U.S. workers belong to the employed-only group. Completing all of these cases would yield over 6,000 interviews, but only approximately 1,800 completed surveys of

⁸ In a study conducted by Abt SRBI based on 2014 data (used in Pew 2016), the cell 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.

⁹ This allocation balances three imperatives: minimizing survey costs, maximizing the precision of the survey estimates, and achieving the goal of boosting the sample of low-wage workers. Unequal representation of demographic groups would require larger differences in weights, increasing the standard errors; in the proposed allocation, the largest weights will occur in the landline-only group, which is the smallest in both the population and the sample (see discussion on weights in Section B.2.1). The landline and cellular samples will be combined using an appropriate statistical weighting protocol to ensure the landline-only, dual-use, and cell-only groups are represented in proportion to their incidence in the target population.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

employed-only respondents are needed for the analysis. Thus employed-only individuals reached on a cellphone will be subsampled at a rate of approximately 30 percent in order to achieve the desired composition of the final sample. This rate can be adjusted during the survey process if we discover that its productivity with respect to the proportion of leave-needers, leave-takers, and employed-only respondents differs from our initial projections. If a respondent is not selected for the extended interview, the call will be terminated and assigned final disposition as a completed screener.

Landline Calls

For the landline calls, as in Wave 3, in the survey screener interviewers will determine whether the household contains at least one person 18 years of age or older who has been employed (excluding self-employed) during the last 12 months. Each eligible adult will be classified into one of the three leave groups: leave-needer, leave-taker, or employed-only. One eligible person will be selected from the screened household for the extended interview, using the following process.

Respondent selection will be conducted in three stages. Stage 1 determines from which leave group represented in the household the respondent will be selected. For households where multiple groups are represented, the leave-needer and leave-taker groups will be selected at a higher rate than the employed-only group. Stage 2 subsamples households in which the employed-only group was selected at Stage 1. Stage 3 selects a random adult from the leave group identified in Stage 1 as the extended interview respondent. The details of this algorithm are provided below. The primary goal of this selection process is to provide a study sample with sufficient leave-takers and leave-needers, two relatively rare groups, while still providing some coverage of employed-only respondents living within households with at least one leave-taker or leave-needer.

STAGE 1: Select a leave group

- If all adults are of one leave group, that group is selected. Skip to STAGE 2.
- If the household has a leave-needer and a leave-taker, select the leave-needer group with 90 percent probability and the leave-taker group with 10 percent probability.
- If the household has a leave-needer and an employed-only adult, select the leave-needer group with 90 percent probability and the employed-only group with 10 percent probability.
- If the household has a leave-taker and an employed-only adult, select the leave-taker group with 90 percent probability and the employed-only group with 10 percent probability.
- If the household has a leave-needer, leave-taker and employed-only adult, select the leave-needer group with 80 percent probability, the leave-taker group with 10 percent probability, and the employed-only group with 10 percent probability.

STAGE 2: Subsample 20 percent of the households where the employed-only group was selected¹⁰

- If the leave-needer group or leave-taker group was selected in Stage 1, skip to Stage 3.

¹⁰ The Stage 2 subsampling rate is based on our experience conducting the Wave 3 FMLA Employee Survey in 2012. In Wave 3, 6,894 employed-only adults were identified in the screener, but only 1,301 were interviewed (18.9 percent). We plan to implement a similar subsampling approach. As noted above, subsampling rates can be adjusted throughout the survey should the productivity of this frame with respect to the leave-needers, leave-takers and employed-only individuals be different from our initial projections.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

- If the employed-only group was selected in Stage 1, select the household for the extended interview with 20 percent probability. If the household is not selected for the extended interview, terminate the call and assign final disposition as a completed screener.

STAGE 3: Select a household member from the selected leave group

- Select a random person (with equal probability of selection) from the leave group selected in Stage 1.

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

For both the cell and landline frames we will closely monitor the yield of extended interviews in each of the three leave groups throughout the field period. Although the group selection rates will be pre-determined before each replicate is released, depending on the yields observed in the early sample replicates, we may modify the group selection rates for later replicates. There are two reasons for this. First, the population incidences of the leave groups may have changed since 2012 when Wave 3 was conducted. (For the landline frame, this makes it difficult to anticipate the optimal group selection rates for households containing adults in multiple leave groups.) Second, the performance of the group selection rates will also depend on the distribution of respondents by leave group. The rates may therefore differ from one replicate to the next, but will be identical for all households (on the landline) or individuals (on the cellphone) within a given replicate. We will document which selection rates are implemented for each replicate. This information will be incorporated into the survey weights (see Section B.2.1) so that computations of the probabilities of selection are accurate for each replicate.

Response Rates

Based on the factors discussed below, our projected overall response rate for the Wave 4 Employee Survey is 15 percent, approximately equal to the rate for the Wave 3 Employee Survey. The overall survey response rate is computed as the product of the *screener* and *extended interview* rates. The extended interview response rates represent the proportion of interviews that were completed among those eligible and selected for the extended interview. Based on the AAPOR RR(3) formula, the screener response rate for Wave 3 was 23.6 percent, and the extended interview response rate was 63.9 percent. The AAPOR RR(3) overall survey response rate for the Wave 3 Employee Survey was therefore 15.1 percent ($0.151=0.236*0.639$).

Despite secular long-term and ongoing trends towards lower response rates (e.g., Tourangeau 2004; Groves 2006; Pew Research Center 2012; National Research Council 2013), we expect to achieve comparable, if not higher, response rates in Wave 4. Our projected response rate for the Wave 4 screener, which is equal to the 23.6 percent response rate in Wave 3, is at the high end of the response rates for similar surveys reported in Exhibit B.3.¹¹ (The response rates in Exhibit B.3 are drawn from RDD surveys with similar sample designs and sponsorship conducted since 2005, which is roughly when survey designers first began supplementing landline RDD samples with cell RDD samples.) Our projected Wave

¹¹ Similarly, our projected 15.1 percent overall survey response rate is based on our experience in Wave 3.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

4 response rate nonetheless seems plausible given our experience in Wave 3 and the changes aimed at increasing response rates (see below, as well as Section B.3.1).

Exhibit B.3. Published Response Rates for RDD Surveys Conducted 2005-2016

Survey	Response Rate (AAPOR 3)
Landline RDD	
2012 California Health Interview Survey (CHIS: adult)	17%
2012 Pew Internet and American Life Study	30%
2015 Behavioral Risk Factor Surveillance System (BRFSS) national median	48%
Cell RDD	
2012 California Health Interview Survey (CHIS: adult)	11%
2012 Pew Internet and American Life Study	20%
2014 National Immunization Survey ¹	33%
2015 Behavioral Risk Factor Surveillance System (BRFSS) national median	47%

¹The NIS is excluded from the landline comparison, because the sample design does not include household sampling in the landline portion of the study but interviews the adult who self-identifies as the most knowledgeable about household immunization information.

Source: CDC BRFSS 2015 Summary Data Quality Report

In the Wave 3 Employee Survey we treated both the landline and cellphone interviews as a household survey, rostering all adults in the household. Yet we found the within-household selection procedure to be especially difficult in the cellphone sample. Similar to the landline process described above for Wave 4, one eligible adult from the screener respondent's household was selected at random for the extended interview. When the process did *not* select the screener respondent—the person who received the call—this led to a “handoff” to the person selected for the extended interview. These handoffs were more difficult in the cellphone sample than in the landline sample because landlines typically are a household-level device, whereas cellphones typically are a personal device. Furthermore, approximately a third of cellphone respondents are typically reached while they are away from their home (AAPOR 2010), making it less likely that the recipient of the handoff is present.

In Wave 3, the overall extended interview response rate across both frames was 33 percent among the cases requiring a handoff, versus 84 percent among the cases in which the screener and extended interview respondent were the same person. Furthermore, the negative effect of a handoff was statistically stronger in the cell frame than the landline frame.¹² Given these rates, and recent patterns in phone usage, household composition, and labor force participation, for the Wave 4 survey we estimate that only 2.9 percent of cellphone interviews with a handoff would be likely to result in a completed interview.¹³ We

¹² We used a logistic regression to model response to the extended interview conditional upon completion of the screener and found that the strongest predictor of response to the extended interview was the handoff. The sampling frame was also associated with extended interview response. When the model for response included an interaction term for frame (cellphone versus landline) with handoff, the interaction of sampling frame and hand-off flag was statistically significant ($p < 0.01$). Indeed, while hand-offs decreased response propensity in both samples, the effect was stronger in the cell sample than the landline sample.

¹³ Abt calculations, using the 2014 ACS PUMS microdata.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

have therefore eliminated this handoff on the cellphone frame, and expect this to improve response rates for the extended interview.¹⁴

However, this positive effect on response rates may be counteracted by the continuing decline in cellphone RDD response rates (Pew Research Center 2012). Most survey researchers attribute the decline in survey response rates to societal factors (Tourangeau 2004). These factors include the general decline in civic engagement (Putnam 1995; see also Groves et al. 2000), increased concern about privacy and confidentiality (Singer et al. 1993), rising hostility toward telemarketers, and the possibility of identity theft. In addition, shifts in the demographic composition of the U.S. population are likely compounding nonresponse. Some of the fastest growing segments of the population (e.g., Hispanics) are known to have generally lower response rates to surveys relative to other Americans.

B.1.2 Employer Survey

For the Wave 4 Employer Survey, this section discusses the universe of potential respondents, the sampling frame, respondent selection process, and expected response rates.

Potential Respondent Universe

The potential respondent universe for the Wave 4 FMLA Employer Survey consists of all private-sector business establishments, excluding self-employed without employees, and government and quasi-government units (federal, state, and local governments, public educational institutions, and post offices). As in the previous FMLA Employer Surveys, a *worksites* is defined as the “single physical location [or address] where business is conducted or where services or industrial operations are performed” (DOL 1996). Data will be collected and analyzed with respect to this worksite, even if the employer has other worksites.

Sampling Frame

As in the Wave 3 Employee Survey, the sampling frame will be created from the Dun and Bradstreet Dun’s Market Identifiers (DMI) file, which provides all essential frame information (e.g., worksite size, industry via North American Industry Classification System [NAICS] code, location, and contact information) for over 15 million private business establishments. The DMI database also includes information that can be used to identify and remove out-of-scope worksites. The DMI file is considered the most comprehensive commercially available business list.

We will use a stratified sampling procedure to select the sample of 2,000 establishments for the Wave 4 Employer Survey. As in Wave 3, sampling strata will be defined in part by the cross-classification of worksite size and NAICS grouping. See Exhibit B.4 for a detailed description of the four worksite size classes and how NAICS codes will be combined to create the four industry groups.¹⁵ We will also oversample paid leave states with parameters identical to those in the Employee Survey (as currently

¹⁴ Overall, the gain in incidence from rostering cellphone interviews is not worthwhile given the additional problems it would create: (1) higher data collection costs due to additional interview length and making the qualifying callbacks to try to reach the selected respondent, (2) increased public burden, and (3) a likely increase in dropout and nonresponse. As a result, the changes in selection probabilities and additional nonresponse would have to be compensated in weighting, increasing the design effects/reducing effective sample sizes and power (see Section B.2.1). A reduction in response rate would also negatively affect the face validity of the study. There is no published evidence that conducting random adult selection for shared cellphones reduces bias in weighted estimates, relative to our recommended approach of selecting the person who answers the phone.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

planned, California, New Jersey, Rhode Island, and Washington will be oversampled so that the base rate of selection is 25 percent higher for worksites in these states). We will implement a fully interacted three-way stratification across these dimensions. This design yields 32 sampling strata (4 size categories times 4 NAICS groups times 2 geographies, paid versus unpaid leave states).

Exhibit B.4. Stratification Variables and Categories for the Wave 4 Employer Survey

Group	Description
Size: small	1–49 employees
Size: medium	50–249 employees
Size: large	250–999 employees
Size: very large	1000+ employees
NAICS Group 1 (Manufacturing)	NAICS codes: Agriculture, Forestry, Fishing and Hunting (11); Mining, Quarrying, and Oil and Gas Extraction (21); Construction (23); Manufacturing (31-33)
NAICS Group 2 (Retail)	NAICS codes: Utilities (22); Wholesale Trade (42); Retail Trade (44-45); Transportation and Warehousing (48-49)
NAICS Group 3 (Services)	NAICS codes: 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 (All Other Industries)	NAICS codes: Educational Services (61); Health Care and Social Assistance (62); Arts, Entertainment, and Recreation (71); Accommodation and Food Services (72); Other Services (81)
Location: FMLA areas	Paid leave states ¹
Location: other	All other states

¹Paid leave states include California, New Jersey, Washington and Rhode Island. The final allocation will align with the Employee Survey.

To arrive at the sample size for each stratum, we first allocate the sample to employment size classes

proportional to the square root (\sqrt{X}) of the aggregate number of employees working for establishments in the class X , as a compromise allocation for computing both per-worksite and per-employee estimates from the Wave 4 Employer Survey. Given that many tabulations from the Employer Survey will be weighted to a “per employee” basis, this is the appropriate approach. This allocation method allows worksites with a large number of employees to be selected at a higher rate than worksites with fewer employees. This will ensure that enough large worksites will be available for the analysis.

Given this approach, Exhibit B.5 provides the detailed cross-classified sample sizes for each of the 32 strata. Across worksite size groups, this design should yield 668 small establishments, 565 medium establishments, 405 large establishments, and 362 very large establishments. Across industry groups, it should yield 389 manufacturing worksites, 449 retail worksites, 559 services worksites, and 603 worksites in all other industries. Lastly, this design should yield 721 worksites in paid leave states, and 1,279 worksites in non-paid leave states.

¹⁵ While these size classes, as defined in the DMI file, will be used for sampling, all establishments will be asked the exact number of employees during the survey.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Exhibit B.5. Detailed Sampling Strata for the Wave 4 Employer Survey

Geography	NAICS Group	Employment Size	# Worksites in Universe ¹	Projected # Worksites in Sample
Paid leave states	1	1 to 49	160,773	43
Paid leave states	1	50 to 249	8,750	40
Paid leave states	1	250 to 999	1,102	29
Paid leave states	1	1,000+	117	19
Paid leave states	2	1 to 49	267,767	60
Paid leave states	2	50 to 249	16,076	54
Paid leave states	2	250 to 999	1,606	33
Paid leave states	2	1,000+	125	21
Paid leave states	3	1 to 49	405,403	61
Paid leave states	3	50 to 249	14,749	52
Paid leave states	3	250 to 999	2,676	46
Paid leave states	3	1,000+	529	50
Paid leave states	4	1 to 49	422,138	77
Paid leave states	4	50 to 249	19,932	57
Paid leave states	4	250 to 999	1,551	35
Paid leave states	4	1,000+	478	44
Non-paid leave states	1	1 to 49	784,022	77
Non-paid leave states	1	50 to 249	47,495	75
Non-paid leave states	1	250 to 999	8,065	63
Non-paid leave states	1	1,000+	916	43
Non-paid leave states	2	1 to 49	1,353,171	109
Non-paid leave states	2	50 to 249	73,238	93
Non-paid leave states	2	250 to 999	8,068	52
Non-paid leave states	2	1,000+	437	27
Non-paid leave states	3	1 to 49	1,807,063	104
Non-paid leave states	3	50 to 249	65,710	88
Non-paid leave states	3	250 to 999	12,646	80
Non-paid leave states	3	1,000+	2,336	78
Non-paid leave states	4	1 to 49	1,940,985	137
Non-paid leave states	4	50 to 249	107,092	106
Non-paid leave states	4	250 to 999	9,129	67
Non-paid leave states	4	1,000+	2,405	80

See Exhibit B.4 for the worksite size and industry group descriptions.

¹Sources: Geography Area Series: County Business Patterns by Employment Size Class more information; 2014 Business Patterns; retrieved from <http://factfinder.census.gov/bkmk/table/1.0/en/BP/2014/00A3//naics-ALL-L2> on 06/14/2016.

To demonstrate the advantage of the $\sqrt{X_{ij}}$ 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.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

In the FMLA Wave 4 Employer Survey, however, many key variables are related to employment size (e.g., percentage of employees covered by FMLA).¹⁶ 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 of an estimate of such a variable (i.e., increase the precision), the strata of large worksites should be sampled with a higher probability than the probability used for the strata of smaller worksites. This way, there are more large worksites in the sample, and the influence of large worksites on the variance of the estimate can be reduced.

Respondent Selection

The first component of the Wave 4 Employer Survey will identify eligible establishments, and collect the name of a “key informant” within the establishment with whom the survey will be conducted. After selecting worksites from the DMI using the process described above, sampled establishments meeting one or more of the following three criteria will be treated as ineligible for the Employer Survey: 1) those where telephone recruitment efforts cannot confirm that the establishment is open/in business during the field period; 2) those where the initially contacted person reports that it is not a private-sector business; and 3) those where the establishment owner is self-employed without employees.

Given the detailed nature of the survey, which includes some questions that may require reference to company administrative records, it will also be necessary during this first step to identify the human resources director or the person responsible for the company’s benefits plan (the “key informant”). Since 2012 the DMI has added additional contact information that can aid in identifying upfront the “key informant” within a worksite’s human resources department. This added information will allow an interviewer to more precisely target and pursue a prospective respondent.¹⁷ As discussed below in Section B.3.2, we expect this to increase response rates in the Wave 4 Employer Survey.

Response Rates

Since the protocol for the Wave 4 Employer Survey largely mirrors that of the Wave 3 survey, our projected response rate for the Wave 4 survey is equal to the Wave 3 survey’s rate of 20.9 percent. While all else equal, the secular trends in response rates are likely to make this figure lower, we expect modest improvements in response rates due to the newly available information on HR contacts within the DMI database. Every effort, within the specifications of the study, will be made to exceed this expectation.

B.2. Analysis Plans

Section B.2 discusses the plans for analysis using the data collected in the Wave 4 FMLA Employee and Employer Surveys (Sections B.2.1 and B.2.2, respectively). This section ends with a brief discussion of

¹⁶ The percentage of employees covered by FMLA is the ratio of the number of employees covered by FMLA to the total number of employees, with the former strongly related to the employment size. See the Abt Methodology Report for the Wave 3 FMLA Surveys (Daley et al. 2013). The report also points out the difficulties in variance estimation and design effect calculations associated with the dual goals of reporting per-worksites and per-employee estimates.

¹⁷ In recent work developing marketing materials for the On-Site Consultation Program for the Occupational Safety & Health Administration (OSHA), Abt developed expertise in navigating the DMI to identify the most appropriate employer contact to receive the marketing literature.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

survey administration (Section B.2.3) and the use of period data collection cycles for the FMLA surveys (Section B.2.4).

B.2.1 Employee Survey

For the Wave 4 Employee Survey, this section discusses the analysis plan, sampling weights and expected design effects, and the expected degree of accuracy.

Plans for Analysis

Consistent with the overall intent and design of the study, our analysis will rely primarily on descriptive statistics. (The cross-sectional nature of survey data makes them ill-suited for assessing causality.) Our descriptive analysis will entail, for example, estimating means of continuous variables, distributions of binary and multinomial variables, and selected cross-tabulations. In the Employee Survey, some key variables include knowledge and awareness of FMLA (e.g., “Have you ever heard about the federal Family and Medical Leave Act?”), prevalence of leave-taking and reasons for and length of leaves taken, and unmet need for leave.

Where feasible, we will compare the Wave 4 Employee Survey findings to those findings from the Wave 3 survey. This will not be possible in instances where question content or the reference period has changed. The changes in content reflect evolving policy interests and priorities while the change in reference period (from 18 to 12 months) adopt a cognitively easier timeframe for respondents, and a consistent timeframe used in the Wave 3 and Wave 4 Employer Surveys.¹⁸

For questions asked of leave-takers in the opening of Section A of the Employee Survey, Exhibit B.6 lists which questions will be comparable across the Wave 3 and Wave 4 surveys. This change in reference period has no effect on the comparability of the remaining questions asked only of leave-takers (the remainder of Section A), questions asked only of leave-needers (Section B), or questions on demographics (Section D) and employment (Section E).

¹⁸ Although the Wave 4 Employee Survey will oversample respondents who took or needed to take leave in the last 12 months, the screener will include a question about leave-taking in the prior 6 months (i.e., 12 to 18 months ago). We will therefore be able to compare the incidence of leave-taking in the last 18 months to the similar statistics collected in Wave 3 (2012), Wave 2 (2000) and Wave 1 (1995). The previous three survey waves oversampled respondents who took or needed to take leave in approximately the prior 18 months (Wave 1 used a 17 to 19 month reference period, and Wave 2 used an 18 to 21 month reference period).

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Exhibit B.6. Comparability of Wave 4 Employee Survey Section A Questions with Wave 3

Survey Question (as it appears in Wave 4)		Comparable with Wave 3? Yes (Y), No (N), or Not Applicable (N/A)		
Question Number	Question Wording	Leaves Overall	Most Recent Leave in Last 12 Months	Longest Leave in Last 12 Months (≥ 3 weeks)
A1	Can you please confirm that in the last 12 months, that is, since [INSERT 12 MONTH PERIOD], you have taken leave from work for ANY of the following reasons: <ul style="list-style-type: none"> • to care for a newborn, newly adopted or new foster child; • for your own serious health condition or to care for someone else's serious health condition; • for pregnancy-related reasons; or • to care for a military service member, or for reasons related to the deployment of a military service member? 	Y		
A3	Are you currently on this type of leave from work?	Y		
A4	We are interested in the number of times you took leave from work for A SINGLE reason or condition (yours, or that of the person you cared for), and this is regardless of whether you took time off all at once or in separate blocks of time. So, for how many TOTAL reasons or conditions did you take leave from work since [INSERT 12 MONTH PERIOD]?	Y		
A5	What was the main reason you took this type of leave from work most recent/longest leave?	N/A	Y	N
A6	What is that person's relationship to you?	N/A	Y	N/A
A8	What was the age of your care recipient?	N/A	Y	N
A10	What was the nature of the health condition for which you took this leave?	N/A	Y	N
A13	For this leave, in what month and year did you start taking time off?	N/A	Y	N
A14	Did you take this time off continuously—that is, all in a row without returning to work—or did you take leave on separate occasions?	N/A	Y	N
A15	How many separate blocks of time did you take off from work during this leave?	N/A	Y	N
A16	In what month and year did the last block of time for this leave begin?	N/A	Y	N
A17	And in what month and year did this leave end?	N/A	Y	N
A18	To review: You've taken leave for [X]/you are not able to tell us when it began/and you are not able to tell us when it ended. Is that correct?	N/A	Y	N
A19	Great, so how much time in TOTAL did you take off from work [so far] for the reason you mentioned [including all blocks of time]?	N/A	Y	N
A19b	In the last 12 months, did anyone else in your household take leave for the same reason you mentioned?	N/A	Y	N/A

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Survey Question (as it appears in Wave 4)		Comparable with Wave 3? Yes (Y), No (N), or Not Applicable (N/A)		
Question Number	Question Wording	Leaves Overall	Most Recent Leave in Last 12 Months	Longest Leave in Last 12 Months (≥ 3 weeks)
A19c	What is this person's relationship to you?	N/A	Y	N/A

For respondents who took more than one leave in the prior 12 months, the Wave 4 survey asks an abbreviated series of questions about the longest leave. In addition, in both Wave 4 and Wave 3 the survey focuses the detailed questions (e.g., A5 and beyond) on a focal leave, rather than all leaves overall.

We will also report Wave 4 results separately by key subgroups to identify differences between groups. For the Employee Survey we will analyze key results by (1) paid leave policies (workers living in the group of states with paid leave versus those living in other states); (2) income (low-wage workers versus those not); (3) FMLA eligibility (eligible workers versus non-eligible); and (4) demographics (e.g., gender, race, ethnicity, marital status, and parental status). We will report all subgroups, but we will discuss in the text only those comparisons for which we can confidently reject a hypothesis of equality of outcomes using appropriate statistical tests (e.g., chi-square, *t*-test).

Following our practice for Wave 3, we expect to do the analysis using SAS PROC SURVEY procedures to estimate level differences in key outcomes across these subgroups (e.g., FMLA eligible and ineligible employees), reporting confidence intervals and testing for statistical significance.¹⁹ Our reporting of the results will focus primarily on those differences that are statistically significant. It is important to note, however, that statistically significant differences across groups—such as differences in leave-taking between employees who live in paid leave states versus other states—are not evidence of a causal effect of FMLA or access to pay under FMLA. For instance, states that adopted paid leave may have had higher leave-taking rates even in the absence of (i.e., before) the legislation.

We will also supplement these simple weighted comparisons with regression-adjusted comparisons for selected outcomes, continuing to use survey design estimation as necessary. This will provide results that are “closer to causal,” by reporting the level differences that remain, after controlling for some observed systematic differences in characteristics that may themselves be correlated with the outcome of interest.²⁰ In our discussion of these results, however, we will take care to use language that clearly notes that no observed differences, either regression-adjusted or otherwise, should be interpreted as identifying the causal effect of FMLA alone.

¹⁹ As discussed in Attachment D, we will replace the jackknife variance estimation method used in Wave 3 with the bootstrap replicate method to compute standard errors.

²⁰ For example, in Wave 3 we observed higher rates of leave-taking among employees eligible for FMLA than among those not eligible (16% versus 10%; Klerman et al. 2013). Although some of this difference may be due to FMLA’s causal effect, some of this difference likely is due to the factors that affect eligibility (e.g., employer size, job tenure, hours worked). It is thus likely that at least some of the observed difference would have remained even in the absence of FMLA. Including regression-adjusted comparisons will allow us to control for such systematic differences in characteristics. The remaining differences will therefore provide a “closer to causal” measure of FMLA’s impact.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Weights

We will develop survey weights to adjust for differential probabilities of selection, integrate the landline and cell samples, and adjust for nonresponse. For the Employee Survey, we will account for the overlap in the landline and cell RDD sample frames (i.e., many people have both types of phones), as well as for oversampling of low-wage workers and workers living in paid leave states. The weights will be computed with the following components:

- A base weight reflecting the overall probability of selection of the household (landline) or individual (cellphone), including the oversampling rates;
- Nonresponse adjustment within the frame using the frame characteristics (such as geographic location, appended demographics, or activity flags);
- An adjustment for the number of telephones in the household (landline) or used by the individual (cellphone);
- The probability of selection of the person within the household (landline only);
- A sample frame integration adjustment (Lohr 2009); and
- A calibration adjustment (Kolenikov 2014; Nadimpalli et al. 2004) to the distribution of the U.S. non-institutional population on age, gender, race/ethnicity, education, and labor force status, based on figures from the Current Population Survey and American Community Survey.

The calibration adjustment will expand the population estimates from the survey to reflect all persons in the country, including those without telephones. It also will adjust for differential nonresponse across demographic groups in the population. Attachment G provides more details on weight construction.

The multi-frame sampling, within-household selection of eligible adults (for the landline interviews), and subsampling of workers who did not need nor took FMLA leave, as well as unit nonresponse throughout all stages of sampling and data collection, inevitably lead to unequal weights. Those unequal weights, in turn, lead to an increase in the variance of survey estimators relative to a simple random sample design. (It should be noted, however, that a simple random sample design is an exclusively hypothetical construct, as no list of all workers in the country exists nor can be created short of an effort comparable to the decennial census.) The ratio of variances in the actual survey to that of a simple random sample, known as the “design effect,” must be incorporated into the process of designing the survey and sample.²¹

In the following design, sample size, and power calculations, we use the design effects observed for Wave 3 (see Exhibit B.7).²² The major component of variation in Wave 3 was the subsampling of the employees who did not take or need FMLA leave (employed-only); followed by the unequal weighting due to design (e.g., dual frame RDD and frame integration), followed by nonresponse adjustments. By using state-of-the-art sampling schemes and carefully optimizing the design parameters, we expect that design effects will be comparable or lower for Wave 4.

²¹ As noted above, following our practice in Wave 3 we expect to use SAS PROC SURVEY, or the equivalent command in Stata, to estimate standard errors and conduct tests of statistical significance.

²² Design effects are commonly approximated as $1 + CV^2$, where CV is the coefficient of variation of weights (CV=standard deviation/mean).

Exhibit B.7. Sample Sizes and Design Effects in Wave 3 Employee Survey

Demographic Group	Leave-Needers And Leave-Takers (Interest)		Employed-Only (Other)	
	Nominal Sample Size	DEFF	Nominal Sample Size	DEFF
Non-Hispanic White	1,118	1.638	958	1.964
Non-Hispanic Black	188	1.635	124	1.851
Non-Hispanic Other	108	2.095	92	1.735
Hispanic	137	1.818	127	1.716
Income <\$30,000	230	1.735	180	1.774
State of CA (frame)	122	2.436	119	1.867
Total	1,551	1.783	1,301	1.932

Degree of Accuracy

The Wave 4 Employee Survey will have a sample size of at least 4,000 completed telephone interviews at the national level. As discussed above, the Wave 4 Employee Survey will oversample respondents (i) living in paid leave states, and (ii) low-wage workers, so that appropriate inferences may be made for these two key subpopulations relative to other workers. This section reports estimates of the minimum detectable differences (MDD) across these subgroups in the proportion of respondents who took leave in the last 12 months,²³ given the projected sample sizes of each subgroup.

The MDD in the proportion of respondents who took leave (p_1-p_2 , for comparing subgroups 1 and 2), is related to the (effective) subgroup sample sizes (n_1 and n_2) in the following way (Fleiss et al. 2003, with continuity corrections):

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

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

$$r = \frac{n_2}{n_1}; \bar{p} = \frac{p_1 + r p_2}{1+r}; \bar{q} = 1 - \bar{p}; q_1 = 1 - p_1; q_2 = 1 - p_2$$

where:

n_1, n_2 are the (effective) sample sizes (number of completed surveys corrected for the expected design effects²⁴);

²³ Question S-11 of the Wave 4 Employee Survey Instrument (see Attachment A to this submission) asks all respondents whether they have taken leave from work in the last 12 months for a family or medical reason (e.g., to care for a newborn, for their own serious medical condition, or to care for a close family member with a serious medical condition).

²⁴ The effective sample size is equal to the nominal sample size divided by the design effect (see for example Exhibit B.8).

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

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

$Z_{1-\alpha/2}, Z_{1-\beta}$ are the percentiles of the standard normal distribution.

In Wave 3 the baseline incidence (proportion) of leave-taking in the last 12 months was 13.2 percent (Klerman et al. 2013). Further, as discussed in Section B.1.1, the overall Employee Survey sample of 4,000 respondents is projected to include:

- 887 respondents living in paid leave states, and 3,113 respondents living in all other states, and
- 970 low-wage respondents, and 2,539 non-low-wage respondents (and 491 with item nonresponse on wage).

Applying the projected overall design effect of 1.85 to these projected sample sizes to obtain effective sample sizes, and assuming conventional values for α and β ,²⁵ we estimate the detectable difference, $p_2 - p_1$, of 5.34 percentage points for workers living in paid leave states versus unpaid leave states, and 5.44 percentage points for low-wage workers versus all others (see Exhibit B.8).

Exhibit B.8. Calculating Estimated Detectable Differences

Analysis	Nominal Sample Size	Design Effect	Effective Sample Size	Base Incidence of Leave-Taking	Estimated Detectable Difference
Paid leave states	887	1.851	479.5	18.54%	5.34%
Non-paid leave states	3,113	1.851	1682.0	13.2%	
Low-wage workers	970	1.774	546.9	18.64%	5.44%
Higher wage workers	2,539	2.039	1245.5	13.2%	

B.2.2 Employer Survey

For the Wave 4 Employer Survey, this section discusses the analysis plan, sampling weights and expected design effects, and the expected degree of accuracy.

Plans for Analysis

As with the analysis of the Wave 4 Employee Survey data, our analysis will rely primarily on descriptive statistics, for example, estimating means of continuous variables, distributions of binary and multinomial variables, and selected cross-tabulations. We will also report findings separately by key subgroups to identify differences between groups. For the Employer Survey, we will analyze key results separately by FMLA coverage (covered worksites versus uncovered sites), and will estimate level differences using design specifications such as survey weights and strata.

Applying the survey weights (discussed below) and other necessary components of survey design-based estimation, we will use SAS or Stata to estimate level differences in key outcomes across these subgroups (e.g., FMLA covered and uncovered worksites), reporting confidence intervals and testing for statistical significance. We will also supplement these simple weighted comparisons with regression-adjusted comparisons for selected outcomes. (See the discussion in Section B.2.1 for more general detail on our analysis approach.)

²⁵ For these calculations we assume a type I error rate $\alpha = 5\%$, so that significance level is 95%, and a type II error rate $\beta = 20\%$, so that power is 80%.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Weights

For the Employer Survey, we will compute weights for each responding worksite that adjust for the differential probabilities of selection as well as nonresponse. Specifically, the weights will be computed with the following adjustments:

- A base weight as the inverse probability of selection of the worksite;
- An adjustment for nonresponse to the screener and extended interview; and
- A calibration adjustment to Bureau of Labor Statistics population controls for the number of worksites and employment size in each of the 32 strata defined by size, industry, and location (paid versus non-paid states).

In addition, as in Wave 3, we will build separate weights to allow us to report results weighted by worksites, and by the number of employees employed at these worksites. Worksite weights allow us to construct estimates for the population of worksites (e.g., “x percent of worksites are covered by FMLA”). Employee weights give larger worksites more weight and allow us to construct estimates for the population of employees in the worksites (e.g., “y percent of employees are in covered worksites”). For more detail see the discussion in Attachment D.

As with the Employee Survey, simple random sample assumptions would understate the true variance of the Wave 4 Employer Survey results. We plan to compute the standard errors using replicate variance estimation methods (Kolenikov 2010) available in SAS or Stata, which take into account the sampling design used for the survey and the sampling weights. Attachment D provides additional details on weighting and replicate weights.

As in Section B.2.1, the following design, sample size, and power calculations build on the results observed in Wave 3 (the sample sizes in the Wave 4 and Wave 3 Employer Surveys are quite similar, 2,000 and 1,800, respectively). Exhibit B.9 reports the power analysis for selected subgroups. Design effects are computed for each specific cell, accounting for the differential sampling probabilities and base weights, as well as for a design effect of 1.3 due to nonresponse adjustments and post-stratification. The proposed sampling design will result in a design effect of 3.19 on variables that are *unrelated* to the stratification variables—worksite size and industry. However, we expect that most variables of substantive interest will be associated either with industry or with the establishment size (e.g., through the FMLA coverage rule), or with geography (i.e., paid leave states). For these variables, the design will provide lower design effects. In our simulations, the design effect for the estimate of total employment is 0.14, and the design effect for the percent female employees is 1.11 (percent female is strongly related to industry).

Exhibit B.9. Projected Subgroup Sample Sizes and Effect Sizes, Wave 4 Employer Survey

Group 1	Group 2	Group 1		Group 2		Base p1	Alternative p2	Effect Size
		Nominal Size	Effective Size	Nominal Size	Effective Size			
Non-paid leave states	Paid leave states	1,279	355	721	198	20%	29.30%	9.30%
NAICS group 1	NAICS group 2	389	91	449	126	20%	38.22%	18.22%

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Group 1	Group 2	Group 1		Group 2		Base p1	Alternative p2	Effect Size
		Nominal Size	Effective Size	Nominal Size	Effective Size			
50 to 249	250+	565	365	767	492	20%	28.49%	8.49%
250 to 999	1000+	405	263	362	229	20%	31.47%	11.47%
1 to 49	50 to 249	668	434	565	365	20%	28.77%	8.77%
Generic	Generic		1,000		100	20%	33.05%	13.05%
Generic	Generic		200		100	20%	35.87%	15.87%
Generic	Generic		500		200	20%	30.37%	10.37%

Degree of Accuracy

The Wave 4 Employer Survey will have a sample size of at least 2,000 completed worksite interviews at the national level. As discussed above, this survey will oversample worksites located in paid leave states. In this section we estimate the minimum detectable difference (MDD) in the proportion of FMLA-covered worksites that had at least one employee take leave in the prior 12 months for an FMLA-qualifying reason, given the projected sample sizes of worksites located in paid leave states versus non-paid leave states.²⁶

As discussed in Section B.1.2, the overall Employer Survey sample of 2,000 worksites is projected to include 721 worksites located in states with paid leave, and 1,279 worksites located in all other states. Given the projected design effect of 4.14, the effective sample sizes are 355 and 198, respectively. Using the estimates from Wave 3 for a baseline incidence (p), 41.8 percent of covered worksites had at least one employee take leave in the prior 12 months for an FMLA qualifying reason. Following the equation and assumptions discussed above in Section B.2.1, the estimated minimum detectable difference in the proportion of covered worksites that have at least one employee take leave in the last 12 months is 12.78 percentage points ($p_2 - p_1$), for an incidence in paid leave areas (p_2) of 54.58 percent.

B.2.3 Who Will Collect the Information and How It Will Be Done

Separately for the Employee and Employer Surveys, this section discusses how the Wave 4 FMLA surveys will be administered.

Employee Survey

The contractor will administer the Employee Survey from its centralized call centers. The survey will be conducted by computer assisted telephone interview (CATI) in both English and Spanish.

Consistent with the Telephone Consumer Protection Act (47 U.S.C. 227), all calls to cellphones will be manually dialed. Landline telephone numbers will be dialed using an autodialer. Telephone numbers are dialed until contact is established with a respondent associated with the number, or until the telephone number is determined to be incorrect or out of service. See Section B.3.1 for additional details on how the Employee Survey will be administered in order to maximize response rates.

Employer Survey

The contractor will also administer the Wave 4 Employer Survey. As in Wave 3, this survey will be conducted using a sequential multi-mode (web and CATI) methodology. As described above, to

²⁶ A worksite is covered by the FMLA if it has 50 or more employees, or if, in combination, there are 50 or more employees working at firm-owned worksites within a 75-mile radius.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

determine worksite eligibility, as well as the name and contact information of the person who is most appropriate to complete the survey, the study team will call each worksite and conduct a short telephone screener survey. Where available, we will utilize the name of the HR contact in the company if it is provided in the DMI data. While the call is taking place, the interviewer will offer to email the key informant about the study. The email will also contain a secure link to take the respondent directly to his/her own landing page to log into the survey. Interviewers will offer to complete the survey with respondents on the phone or walk them through how to access the web-based survey on their own. See Section B.3.2 for additional details on how the Employer Survey will be administered in order to maximize response rates.

B.2.4 Use of Periodic Data Collection Cycles to Reduce Burden

The Wave 4 FMLA Employee and Employer Surveys are a one-time data collection effort and will not require periodic data collection cycles.

B.3. Methods to Maximize Response Rates and Deal with Nonresponse

Separately for the Wave 4 Employee and Employer Surveys, the following section discusses methods to maximize survey response rates and to address survey nonresponse.

B.3.1 Employee Survey

For the Wave 4 Employee Survey, the following section first discusses proposed methods to maximize survey response rates, and then describes multiple methods to evaluate survey nonresponse.

Methods to Maximize Response Rates

Several recruitment strategies will be used to increase the response rate to the Wave 4 Employee Survey:

- As in Wave 3, interviewers will attempt to reach respondents on landlines a maximum of 15 times, and on cellphones a maximum of 8 times. We set different calling rules for landline and cellphones, because what may seem a moderate number of calls to a household's landline can seem excessive to an individual's cellphone. For landlines, more calls will be permitted if contact is made with an eligible household but the interviewer is asked to call back later.
- For participants completing the questionnaire on a cellphone, a \$15 incentive will be issued to cover any extenuating expenses associated with per minute carrier charges. The National Research Council (2013) notes that "Holbrook et al. (2008) analyzed 114 RDD surveys between 1996 and 2005 and found, after controlling for other variables, that incentives were significantly associated with higher response rates, with the effect due mainly to a reduction in refusals (with no change in contact rates)." According to the projections by Mercer et al. (2015), a promised \$15 incentive boosts response rates by 6 percent.
- Interviews will be conducted during various times of the day and seven days a week to increase the likelihood of finding the respondent at home. We will analyze production data to determine the best days and times to contact respondents and avoid refusals. To increase the probability of completing an interview, we will establish a differential call rule requiring that call attempts be initiated at different times of the day and days of the week.
- Respondents will be provided with the option of scheduling the interview at the time that is convenient for them (i.e., outside the regular calling hours).

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

- Using pre-translated instruments, the interviewing will be conducted in English or Spanish, depending on the respondent’s preference.²⁷
- We will also use the following refusal conversion approach:
 - For soft-refusals, “interview converters” who have extensive training in telephone interviewing and converting nonresponders will be used to increase the response rate. (We will exclude verbally abusive respondents and those who explicitly say they do not want to participate.)
 - We will code refusals by type and by when they occurred in the survey, which will help us to understand why refusals occurred in order to adjust our protocols.
 - After the first refusal, we will send a refusal conversion letter to individuals or households with an address available through reverse lookup, reminding them about the validity of the study and the importance of their participation.
 - We will track refusals in real time, so interviewers who are generating refusals can be identified and retrained.
- Lastly, as discussed in Section B.1.1, for the Wave 4 Employee Survey we will structure the cellphone interviews as an individual survey, rather than a household survey. By eliminating the within-household selection procedure, and thus the need for a “handoff” from the original respondent (who answered the screening section) to the selected interviewee, we expect to increase response rates for the extended survey.

Our proven approach to optimally allocate the field interviewing resources is to analyze all available information and incorporate new sources of information—such as geographic incidence by rate center for cellphone numbers, or activity status. By computing multiple predictive scenarios using different sample stratification options, we can determine the best solution for reaching the target number of interviews from each stratum while minimizing design effects and potential bias. This is accomplished by releasing a set of replicates, analyzing the results, and re-stratifying the sample if needed.

Methods for Coping with Nonresponse

The pattern of declining response rates is a serious concern, although several studies indicate that lower response rates do not necessarily imply greater error (Curtin et al. 2005; Groves 2006; Keeter et al. 2000; Pew Research Center 2012). However, it is still the responsibility of the researcher to understand and if possible to quantify the risk of error from nonresponse bias.

Beyond designing the data collection procedures to minimize nonresponse (see above), and weighting for nonresponse (see Section B.2.1), we plan to evaluate nonresponse in the Employee Survey in four conventional, but imperfect ways: (1) a nonresponse follow-up survey (NRFU); (2) a comparison of easy-to-reach versus harder-to-reach respondents; (3) fitting response propensity models; and (4) comparing survey estimates with external benchmarks. If, despite our best efforts, our response rates fall below OMB’s required threshold, we will use these methods to identify any evidence of bias generated by low response rates, and make that information available to future users of the data.

²⁷ In the Wave 3 Survey, 0.3 percent of all numbers dialed were recorded as non-interviews because the respondent spoke neither English nor Spanish.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

In order to minimize the possibility of nonresponse bias, however, we will build on our experience fielding the Wave 3 Employee Survey to focus our efforts on increasing response rates in those groups that are hard to reach or interview. For instance, in the Wave 3 level of effort analysis, we found that leave-takers and leave-needers were more likely to complete the interview on the first few attempts than employed-only respondents. This may be because respondents with some leave experience find the survey more relevant to them. During Wave 4 we will therefore closely watch the performance of low-participating strata, such as employed-only respondents, to monitor progress. If we find evidence of lower response rates, or that respondents in these strata look systematically different from other respondents, we will update the introductory script and work with interviewers to make certain that we are not excluding certain subgroups.

Nonresponse Follow-Up (NRFU) Survey

The NRFU will collect information on employees who fail to respond to the survey, to provide insight into whether they differ from respondents on the characteristics of interest. Specifically, interviewers will call back a subsample (n=1,200) of households that declined the original survey and attempt to recruit an eligible employee to complete a shortened interview with a \$40 incentive.²⁸ In addition, all landline sample cases that can be matched to an address through reverse lookup will receive a letter encouraging them to cooperate with the interview.²⁹ We expect to complete approximately 500 NRFU interviews. This will provide a sufficient case base for meaningful nonresponse analysis.

For the NRFU, we will sample both non-contact nonresponding households and non-cooperative nonresponding households. This way, we can evaluate whether employment and leave characteristics differ between these two groups and whether they differ from the responding sample.

We will compare the employment and leave characteristics of Employee Survey respondents with the characteristics of NRFU respondents using paired t-tests. This analysis will provide insights about the direction and magnitude of possible nonresponse bias. We will investigate whether any differences remain after controlling for major weighting cells (e.g., within race and education groupings). If weighting variables eliminate any statistical differences, this suggests that the weighting adjustments discussed in Section B.2.1 will reduce nonresponse bias in the final survey estimates. If, however, the differences persist after controlling for weighting variables, then this would be evidence that the weighting may be less effective in addressing nonresponse bias.

Level of Difficulty Analysis

Our second proposed nonresponse analysis is a level-of-difficulty comparison (Curtin et al. 2005; Keeter et al. 2000; Lin and Schaeffer 1995). This analysis will compare the leave-related characteristics of respondents who were easy to reach with characteristics of respondents who were harder to reach. The level of difficulty in reaching a respondent will be defined in terms of the number of call attempts required to complete the interview and whether the case was a converted refusal. In some studies, this is

²⁸ Incentives are a common feature in NRFU surveys because, by definition, the NRFU sample did not cooperate with the original survey, and so a major change in the recruitment protocol is required to elicit cooperation in the NRFU. Large monetary incentives (e.g., \$20 to \$50) are a common element of NRFU designs for household surveys (Zimowski et al. 1997). For example, Peytchev et al. (2009) documented how a \$20 incentive was used in a successful NRFU to the National Intimate Partner and Sexual Violence Survey for the Centers for Disease Control and Prevention.

²⁹ Reverse lookup generally results in addresses for 5 to 20 percent of the sample. Landlines tend to be associated with physical addresses more commonly than are cellphones.

described as an analysis of “early versus late” respondents, though we propose to also explicitly incorporate refusal behavior.

If the employment and leave-related characteristics of the harder-to-reach cases are not significantly different from characteristics of the easy-to-reach cases, this would suggest that survey estimates may not be substantially undermined by nonresponse bias. The harder-to-reach cases serve as proxies for the nonrespondents who never complete the interview. If the harder-to-reach respondents do not differ from the easy-to-reach ones, then presumably the sample members never reached would also not differ from those interviewed.³⁰ Support for this “continuum of resistance” model is inconsistent (Lin and Schaeffer 1995; Montaquila et al. 2008), but it can still be a useful framework for assessing the relationship between level of effort and nonresponse bias.

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

Estimating Response Propensity Models

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

In order for a response propensity model to be informative, the researcher must know the values for respondents and nonrespondents on one or more predictors of survey response. In RDD surveys, propensity models are often quite limited because little information is generally known for the nonrespondents. For the Employee Survey, we therefore propose to fit a response propensity model predicting the probability of completing the extended interview, conditional on having completed the screener.³¹

By focusing on screened respondents, we can include richer independent variables in the model, including the selected respondent’s age, gender, employment status, and leave status. In addition, the model will include an indicator for sampling frame (landline RDD or cellular RDD), an indicator for whether or not

³⁰ As noted above, in the level of difficulty analysis for Wave 3 (see Daley et al. 2013), when comparing respondents who completed the interview on the first few attempts to those who completed after four or more attempts, the analysis found that the “early respondents” were more likely to be leave-takers (41.3% versus 35.6%) or leave-needers (16.7% versus 14.5%; combined 58.0% versus 40.1%, chi-square $p < 0.001$). One post-hoc explanation is that leave-takers and leave-needers may have felt that the survey was more relevant to them, and may have therefore been more eager to participate than those who with no leave experience.

³¹ As we noted in in Section B.1.1, the response rate to the extended interview in Wave 3 was lower for those that required a handoff (33% among the cases requiring a hand-off versus 84% percent among the cases in which the screener and extended interview respondent were the same person). Therefore, in the landline frame, we expect a moderate number of households to complete the screener, but to be lost in an attempt to interview the randomly chosen respondent.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

the respondent ever refused the interview, and a log-transformed variable for the number of call attempts made to the respondent.

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

The response propensity model will help us to identify the most powerful predictors of response when all available predictors are tested simultaneously. If employment-related or leave-related variables show a significant association with response to the extended interview (after controlling for other factors), this would be evidence of possible nonresponse bias. If, however, the employment- and leave-related predictors do not have a significant effect, this suggests that the screener nonresponse adjustment described in Section B.2 will be effective in reducing nonresponse bias.

Similarly, comparisons of the respondent characteristics across the five response propensity classes will also provide insight on which types of screened respondents were most likely to complete the extended interview and which types were less likely to do so. In Wave 3, response propensity analysis identified hand-off, especially on the cell frame, as the main determinant for nonresponse to the extended interview. As discussed above, based on this finding we have decided to treat the cellphone frame as an individual survey for Wave 4—thus requiring no handoff.

Comparisons with External Benchmarks

The final analysis we will conduct for nonresponse is a comparison of survey estimates with national benchmarks. One limitation of the techniques discussed above is that they analyze only a subset of all nonrespondents to the survey. The NRFU analysis relies on the NRFU participants as proxies for all nonrespondents; the level-of-difficulty analysis relies on the harder-to-reach respondents as proxies for all nonrespondents; the response propensity model captures only variation between respondents and nonrespondents to the screened extended interview.

One approach for evaluating the total level of nonresponse bias in a survey is to compare the weighted survey estimates with external estimates based on a “gold standard” survey. The gold standard survey should feature a more rigorous protocol (e.g., area-probability sampling with in-person interviewing) and a higher response rate than the target survey (in this case the Wave 4 Employee Survey). Critically, the gold standard survey and the target survey must feature highly similar target populations and one or more questions administered in a highly similar manner. Estimates based on these questions can be compared. By virtue of the gold standard survey’s more rigorous design, its estimates are assumed to contain less nonresponse bias than those from the target survey.

Following this approach, we will compare weighted estimates from the Wave 4 Employee Survey with those from the Current Population Survey (CPS). The outcomes of interest are the differences between these two sets of estimates.³² Examples of possible analytic variables administered in both surveys are

³² Comparisons with external benchmarks of the Wave 3 data found that the respondents in the FMLA survey were more likely to be unionized (14.5% versus 11.4%) and less likely to be currently employed (88.4% versus 95.1%), although the definitions of the labor force status differed slightly between the FMLA survey and CPS,

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

marital status, employment status, and employer type (government, private company, non-profit, or self-employment).. In this analysis, nonresponse must be treated in the aggregate (i.e., in terms of responding to the survey, not partial interviews or item nonresponse). Differences in mode of administration may confound the comparison; however, there are no gold standard RDD surveys, and CPS is commonly used as a benchmark for RDD surveys. Additionally, in spite of our attempts to match wording, population coverage, and ordering, where appropriate, there may be some confounds due to differences where matches are not practical. In light of these considerations, we will interpret these results cautiously.

Nonresponse Analysis Considerations

Where possible, we will treat non-contact and non-cooperation as two distinct outcomes. Non-contact (never reaching a respondent) and non-cooperation (reaching a respondent who, for example, asks to be called back later) are generally considered to reflect two different dimensions on which sample respondents can be placed (Stinchcombe et al. 1981; Goyder 1987; Groves and Couper 1998). As noted by Stoop (2005), decomposing nonresponse into these two different dimensions can be analytically useful in several ways:

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

Though each of these analyses methods discussed above relies on imperfect assumptions, all are standard techniques for assessing potential nonresponse error. No single nonresponse analysis for this study can be definitive because the true probability of responding cannot be known. That said, using several different methodologies (nonresponse follow-up analysis, easy-to-reach versus hard-to-reach comparisons, response propensity models, and comparisons of estimates with external benchmarks) is likely to be insightful about the level of risk to survey estimates from nonresponse bias. This information may also be helpful in modifying nonresponse weighting adjustments to reduce bias to the extent possible.

Missing Data

We discuss here our approach to missing covariate data for the regression adjustment process described in Section B.2. For those factors included as controls, a dummy variable adjustment approach will be used to address item-non-response in the survey responses. This strategy sets missing cases to a constant and adds “missing data flags” to the impact analysis model. As detailed by Puma et al. (2009), the dummy variable adjustment approach involves the following three steps:

as the former included some individuals that would be considered out of the labor force by CPS.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

1. For each covariate X with missing data, create a new variable Z that is set equal to X for all cases where X is non-missing, and set to a constant value for those cases where X is missing.
2. Create a new “missing data flag” variable D , which is set equal to one for cases where X is missing and set equal to zero for cases where X is not missing.
3. In the regression model use Z and D (not X) as covariates.

B.3.2 Employer Survey

For the Wave 4 Employer Survey, the following section first discusses proposed methods to maximize survey response rates, and then describes multiple methods to evaluate survey nonresponse.

Methods to Maximize Response Rates

In order to achieve our estimated response rate, as discussed in Section B.2.3 we will use multiple modes of survey administration by implementing both telephone and web-based modes. We have developed a strategy that will maximize internet interviews, which will minimize respondent burden while increasing response rates. We anticipate that the following three components of the survey administration process for the Wave 4 Employer Survey will increase response rates.

First, as discussed in Section B.1.2, the screening interview must confirm that the establishment contacted is the worksite listed on the sampling frame, that the worksite is in the private sector, and that the worksite has at least one employee. However, in Wave 3 for larger firms we experienced challenges completing the extended survey because the initial screener respondent often could not answer questions about the company’s benefit plan, requiring a handoff to another party, and often leading to breakoffs. With the aid of the enhanced DMI contact data described above, for the Wave 4 survey we will therefore attempt to avoid this handoff by directly contacting the most knowledgeable respondent for the screening interview.

Second, in the Wave 3 data collection, after a firm was verified and a “key informant” was identified, an informational packet was mailed explaining the survey and the information necessary to complete the questionnaire. In Wave 4, whenever possible (and with the consent of the identified respondent) we plan to collapse these two phases. While the call is taking place, the interviewer will offer to email the key informant about the study. The email will also contain a secure link to take the respondent directly to his/her own landing page to log into the survey. Interviewers will offer to complete the survey with respondents on the phone or walk them through how to access the web-based survey on their own. It seems likely that this more efficient and shorter process between initial screening and interviewing will improve overall response rates.

Third, we will continue attempting to contact those prospective respondents that we are unable to reach and interview immediately. Specifically, we will send the survey materials (via priority mail), with a cover letter printed on DOL letterhead, so the recipient can clearly distinguish the survey materials from junk mail.³³ All nonrespondents to this mailing will then be contacted by our professional interviewers, who will attempt to complete the interview over the phone. We will send up to four email reminders when

³³ As feasible, these materials will be addressed to a key HR contact at the worksite (if the DMI database lists this contact, or if the initial screener respondent provides contact information for a key informant in HR).

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

an email address has been provided, and the calling protocol for the CATI follow-up effort will be 10 calls per phone number available for the informant.

Methods for Coping with Nonresponse

Given that the anticipated response rate for the Wave 4 Employer Survey is under 70 percent, we will conduct an extensive nonresponse analysis per Office of Management and Budget (2006) Guidance on Agency Survey and Statistical Information Collections. We will conduct an empirical investigation of the potential risk posed by nonresponse bias. The approaches used to evaluate nonresponse in the Employer Survey are a comparison of easier-to-reach versus harder-to-reach worksites and response propensity modeling.³⁴

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

As in the Wave 3 Employer Survey, worksites that are more difficult to interview in the Wave 4 Employer Survey will be compared with those that are easier to interview, where difficulty is defined as a function of the number and types of contact attempts. The more difficult cases serve as proxies for the worksites that never complete the extended interview. As discussed above in Section B.3.1, if the harder-to-reach cases do not differ from the easier-to-reach ones, then presumably the sample members never reached also do not differ from those interviewed. If observed differences disappear after controlling for weighting variables, then that would suggest that the weighting protocol has minimized the risk of nonresponse bias with respect to the estimate at hand.³⁵ As noted above, support for this “continuum of resistance” model is inconsistent (Lin and Schaeffer 1995; Montaquila et al. 2008), but it can still be a useful framework for assessing the relationship between level of effort and nonresponse bias.

Response Propensity Models for Contact and Cooperation

The second approach for evaluating the potential for nonresponse bias in the Wave 4 Employer Survey is a response propensity analysis that identifies factors associated with survey response. Many worksite characteristics can influence the propensity to respond to the survey. 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 will consider two different outcomes: contact with the worksite, and cooperation with the extended interview conditional upon contact. Given the rich sampling frame, the following variables will be known for both the contacted and non-contacted cases: worksite size (measured as number of employees), industry (NAICS code), and Census region.³⁶ We plan to present two separate logistic

³⁴ For the Wave 4 Employer Survey the nonresponse analysis does not include a comparison of weighted estimates with external benchmarks because the three estimates available from the key benchmark study (the Quarterly Census of Employment and Wages) will be used in the Wave 4 weighting protocol, and therefore cannot be used for a benchmark comparison analysis.

³⁵ The Wave 3 level of difficulty analysis found that establishments with 10 to 249 employees required the fewest attempts (4.11), compared to those with 1 to 9 employees (4.99) and 250 or more employees (6.19). We also found that FMLA covered worksites required 0.5 more attempts to complete than uncovered worksites, and that worksites with a larger proportion of female employees required more attempts than those with fewer female employees (establishments that were 50% to 74.9% female required 5.19 attempts on average, while those with 1% to 24.9% required 3.87). In addition, manufacturing worksites required the fewest attempts (3.87 on average), while those in services required the most attempts (5.23). After controlling for weighting variables (including size and industry), however, only the relationship with percent female employment remained significant.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

regression models in this analysis because most readers find those results easiest to interpret. The contact model will be based on all establishments in the released replicates.

The cooperation model will be based only on establishments for which we have a completed screener. By focusing on screened establishments, we can include richer explanatory variables in the model. (In addition to the variables used in the contact model, we will include an indicator of whether the establishment maintains records of FMLA leave, and whether FMLA requests are processed internally or outsourced.) In a well-specified model, responding and nonresponding establishments will be equivalent on the characteristics of interest within each response class, and likelihood of survey participation will vary across the classes. The cooperation propensity model will help us to identify the most powerful predictors of Employer Survey cooperation when all available predictors are tested simultaneously.³⁷

Missing Data

See the discussion in Section B.3.1 for our approach for addressing missing data.

B.4. Tests Procedures

Section B.4 discusses cognitive testing for the Wave 4 Employee Survey and user-testing for the Wave 4 Employer Survey.

B.4.1 Cognitive Testing of the Employee Survey

Prior to finalizing the questionnaire, we cognitively tested the draft instrument. The goals of the cognitive testing were twofold: (1) to learn how respondents perform the tasks necessary to complete the questionnaire; and (2) to determine what difficulties respondents encounter and the likely cause of those difficulties, and to devise potential solutions.

The cognitive interview research protocol on the Employee Survey followed guidelines recommended in Statistical Policy Directive No. 2: Standards and Guidelines for Statistical Surveys, section on Cognitive Interviewing³⁸. The team developed a sampling approach to purposively select respondents with a range of characteristics that we hypothesized were important for the research questions, namely leave taking experience, gender education and occupation. We developed a recruitment plan with planned advertisements and flyers. Our protocol included structured probes in the survey to test the several aspects of the response process. We asked respondents to use their own words to explain the meaning of questions; to think aloud as they answered questions and we asked debriefing questions at the end of the interview. Also, we recorded spontaneous questions asked by respondents. To analyze the data we reviewed each case and compared responses to probes and spontaneous questions across cases. To the extent possible given the small sample size, we analyzed data for patterns across the categories of respondents we recruited, however, our sample size was limited by OMB requirements to limit the sample to under ten respondents.

³⁶ For the Wave 4 Employee Survey, the response propensity modeling does not attempt to model respondent contact because essentially no useful information is available for the non-contacted cases.

³⁷ In Wave 3, response propensity modeling showed that larger worksites, retail worksites, and worksites in the Midwest Census region were easier to establish contact with. However, the probability of cooperation (completing the main interview) was found to be declining with company size.

³⁸ See guidelines here: https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/directive2/final_addendum_to_stat_policy_dir_2.pdf

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Because the Wave 3 Employee Survey was already field tested, in Wave 4 we focused 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?

We conducted nine cognitive interviews. We recruited a purposive sample of respondents for the cognitive testing, setting quotas for respondents across a number of categories that we anticipated would be important for response. We included respondents working across a range of occupations. To detect variation in response quality (comprehension, retrieval recall, and response), we also included respondents across age, gender, and race/ethnicity.

These interviews took place in person in the contractor's Chicago office and at the homes of recruited volunteers. For interviews conducted in the office we used an interviewer-administered CATI questionnaire in five of nine interviews. In the home interviews we used a paper-and-pencil questionnaire. Interviewers introduced the study and the tasks associated with cognitive interviewing (e.g., "thinking aloud"). Interviewees were instructed to tell the interviewer what they were thinking as they answered questions. Interviewer probes included specific, scripted queries 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 (i.e., "What were you thinking?") if the respondent seemed to have difficulty answering.

Cognitive interviews were led by two professional staff or interviewers. Interviewers administered the questionnaire and took notes to observe any nonverbal cues, adding additional probes and follow-up questions as necessary. The interviewer pairs debriefed the larger study team on the questionnaire testing and suggested edits.

B.4.2 User-Testing of the Employer Survey

The Employer 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 web instrument. For this instrument we prioritized user testing over cognitive testing because interviewing respondents are well versed in the concepts and language used in the instrument. In addition, in both survey modes we provide detailed references and definitions as needed.

We conducted user-testing with five individuals. We recruited a convenience sample of respondents for the user-testing. In recruiting a sample of prospective respondents we sought diversity with respect to the type of company, position of respondent and whether the company processed FMLA request internally or outsourced to another company.

Respondents completed an online version of the questionnaire and then took part in a debriefing interview, by phone. The debrief interview focused on the following topics:

- Overall ease of completing the questionnaire;

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

- Retrieval of administrative data for answering the survey;
- Including comprehension of terms and topic; and
- Scope of the survey topics and applicability for the respondents' firms.

Debriefing interviews were led by one or two professional interviewers. The interviewer debriefed the larger study team on the questionnaire testing and suggested edits. User-testing revealed that respondents would benefit from clearer instructions and additional technical definitions, and led to the clarification of one question. We updated the Employer Survey based on these findings.

B.5. Individuals Consulted on Statistical Aspects of the Design

The U.S. Department of Labor (DOL) has contracted with Abt Associates Inc. (in partnership with Abt SRBI) to conduct the FMLA Wave 4 surveys. The contractor is responsible for the study design, instrument development, data collection, analysis, and reporting. The individuals listed in Exhibit B.10 developed the survey instruments and will have primary responsibility for the data collection and analysis.

Exhibit B.10: Contact Information for Contractor

Name	Organization	Telephone Number	Role in Study
Mr. Glen Schneider	Abt Associates	(617) 349-2471	Principal Investigator
Ms. Radha Roy	Abt Associates	(301) 347-5722	Project Director
Mr. Jacob Klerman	Abt Associates	(617) 520-2613	Project Quality Advisor
Dr. Jane Herr	Abt Associates	(617) 520-3042	Director of Analysis

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

Dr. Jolene Smyth University of Nebraska-Lincoln (402) 472-0662

Inquiries regarding the study's planned analysis should be directed to:

Mr. Glen Schneider Principal Investigator, Abt Associates (617) 349-2471
Dr. Christina Yancey Senior Evaluation Specialist, Chief Evaluation Office, DOL (202) 693-5910

References

- AAPOR (American Association for Public Opinion Research). 2010. *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Cell Phone Numbers. Final Report of the AAPOR Cell Phone Task Force*. Oakbrook Terrace, IL: American Association for Public Opinion Research.
- Barron, Martin, Felicia LeClere, Robert Montgomery, Staci Greby, and Erin D. Kennedy. 2015, February. "The Accuracy of Small Area Sampling of Wireless Telephone Numbers." *Survey Practice* 8 (2). <http://www.surveypractice.org/index.php/SurveyPractice/article/view/281>.
- Blumberg, Stephen J., and Julian V. Luke. 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>.
- Cantor, David, Jane Waldfogel, Jeffrey Kerwin, Mareena McKinley Wright, Kerry Levin, John Rauch, Tracey Hagerty, and Martha Stapleton Kudela. 2001. *Balancing the Needs of Families and Employers: Family and Medical Leave Surveys, 2000 Update*. Rockville, MD: Westat.
- Curtin, Richard, Stanley Presser, and Eleanor Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69: 87–98.
- Daley, Kelly, Courtney Kennedy, Marci Schalk, Julie Pacer, Allison Ackermann, Alyssa Pozniak, and Jacob Klerman. 2013. *Family and Medical Leave in 2012: Methodology Report*. Cambridge, MA: Abt Associates. <http://www.dol.gov/asp/evaluation/fmla/FMLA-Methodology-Report-Appendices.pdf>.
- DOL (U.S. Department of Labor), Commission on Leave. 1996. *A Workable Balance; Report to Congress on Family and Medical Leave Policies*. Washington, DC: U.S. Department of Labor. <https://www.dol.gov/whd/fmla/1995Report/family.htm>.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd edition. New York: Wiley.
- Goyder, J. 1987. *The Silent Minority: Nonrespondents on Sample Surveys*. Boulder, CO: Westview Press.
- Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 646–675.
- Groves, R., and M. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R., Singer, E., & Corning, A. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64(3), 299–308.
- Holbrook, A., Krosnick, J., and Pfent, A. (2008). The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. In *Advances in Telephone Survey Methodology*, edited by James M. Lepkowski, Clyde Tucker, J. Michael Brick,

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

- Edith D. De Leeuw, Lilli Japac, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 499-528. New York: Wiley.
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64: 125–150.
- Klerman, Jacob, Kelly Daley, and Alyssa Pozniak. 2013. *Family and Medical Leave in 2012: Technical Report*. Cambridge, MA: Abt Associates.
- Kolenikov, Stanislav. 2010. "Resampling Variance Estimation for Complex Survey Data." *The Stata Journal* 10 (2): 165–169. <http://stata-journal.com/article.html?article=st0187>.
- Kolenikov, Stanislav. 2014. "Calibrating Survey Data Using Iterative Proportional Fitting (Raking)". *The Stata Journal* 14 (1): 22–59.
- Kolenikov, S., and Hammer, H. 2015. "Simultaneous Raking of Survey Weights at Multiple Levels." *Survey Methods: Insights from the Field, Special issue: 'Weighting: Practical Issues and "How to" Approach*. <http://surveyinsights.org/?p=5099>
- Lin, I-Fen, and Nora Schaeffer. 1995. "Using Survey Participants to Estimate the Impact of Nonparticipation." *Public Opinion Quarterly* 59: 236–258.
- Little, R. J. A. 1986. "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review* 54: 139–157.
- Lohr, S. 2009. *Sampling: Design and Analysis*. 2nd edition. Pacific Grove, CA: Duxbury Press.
- Martonik, Rachel, Tara Merry, Nicole Lee, Stephen Immerwahr, and Michael Sanderson. 2015. "Managing Efficiency in Telephone Surveys: Insights from Survey Paradata Trends 2010-2015". Paper presented at American Association for Public Opinion Research 71st Annual Conference, Austin, TX.
- McGeeney, Kyley. 2015. "Appending a Prepaid Phone Flag to the Cellphone Sample." *Survey Practice* 8 (4). <http://www.surveypractice.org/index.php/SurveyPractice/article/view/311>.
- Mercer, A., A. Caporaso, D. Cantor, and R. Townsend. 2015. "How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys." *Public Opinion Quarterly* 79 (1): 105–129.
- Montaquila, J., J.M. Brick, M. Hagedorn, C. Kennedy, and S. Keeter. 2008. "Aspects of Nonresponse Bias in RDD Telephone Surveys." In *Advances in Telephone Survey Methodology*, edited by James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. De Leeuw, Lilli Japac, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 561–586. New York: Wiley.
- National Research Council. 2013. *Nonresponse in Social Surveys: A Research Agenda*. R. Tourangeau and T. J. Plewes, editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Nadimpalli, V., D. Judkins, and A. Chu. 2004. "Survey Calibration to CPS Household Statistics." *Proceedings of the Survey Research Methods Section, American Statistical Association*: 4090-4094.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

- Olson, K. 2006. "Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias." *Public Opinion Quarterly* 70: 737–758.
- OMB (Office of Management and Budget). 2006. "Questions and Answers when Designing Surveys for Information Collections."
http://www.cio.noaa.gov/itmanagement/pdfs/OMBSurveyGuidance_0106.pdf
- Pew Research Center. 2012, May 15. "Assessing the Representativeness of Public Opinion Surveys." Pew Research Center website, U.S. Politics & Policy. <http://pewrsr.ch/JFbCj7>.
- Pew Research Center. 2015, November 18. "Advances in Telephone Sampling." Pew Research Center website. <http://pewrsr.ch/1Of5EqB>.
- Pew Research Center. 2016, October 24. "Cellphone Activity Flags." Pew Research Center website. <http://www.pewresearch.org/2016/10/24/cellphone-activity-flags/>.
- Peytchev, A., R. K. Baxter, and L. R. Carley-Baxter. 2009. "Not All Survey Effort is Equal: Reduction of Nonresponse Bias and Nonresponse Error." *Public Opinion Quarterly* 73 (4): 785–806.
- Pierannunzi, C., M. Town, WS. Garvin, L. Balluz, et al. 2013, May. "Movers and Shakers: Discrepancies between Cell Phone Area Codes And Respondent Area Code Locations In RDD Samples." Presentation at 2013 annual conference of the American Association for Public Opinion Research, Boston.
- Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. 2009. *What to Do When Data Are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/ncee/pdf/20090049.pdf>.
- Putnam, Robert D. 1995. "Bowling Alone: America's Declining Social Capital." *Journal of Democracy* 6 (1): 65-78.
- Kolenikov, S. 2010. "Resampling Variance Estimation for Complex Survey Data." *The Stata Journal* 10 (2): 165-199.
- Singer, E., N. A. Mathiowetz, and M. P. Couper. 1993. "The Impact Of Privacy And Confidentiality Concerns On Survey Participation: The Case Of The 1990 U.S. Census." *Public Opinion Quarterly* 57 (4): 465–482.
- Stinchcombe, A. L., C. Jones, and P. Sheatsley. 1981. "Nonresponse Bias for Attitude Questions." *Public Opinion Quarterly* 45 (3): 359–375.
- Stoop, I. A. L. 2005. *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague, The Netherlands: Social and Cultural Planning Office.
https://www.scp.nl/english/Publications/Publications_by_year/Publications_2005/The_Hunt_for_the_Last_Respondent.
- Tourangeau, R. 2004. "Survey Research and Societal Change." *Annual Review of Psychology* 55: 775–801.

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

Zimowski, M., R. Tourangeau, R. Ghadialy, and S. Pedlow. 1997. *Nonresponse in Household Travel Surveys*. Chicago, IL: National Opinion Research Center.

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.570.6554&rep=rep1&type=pdf>