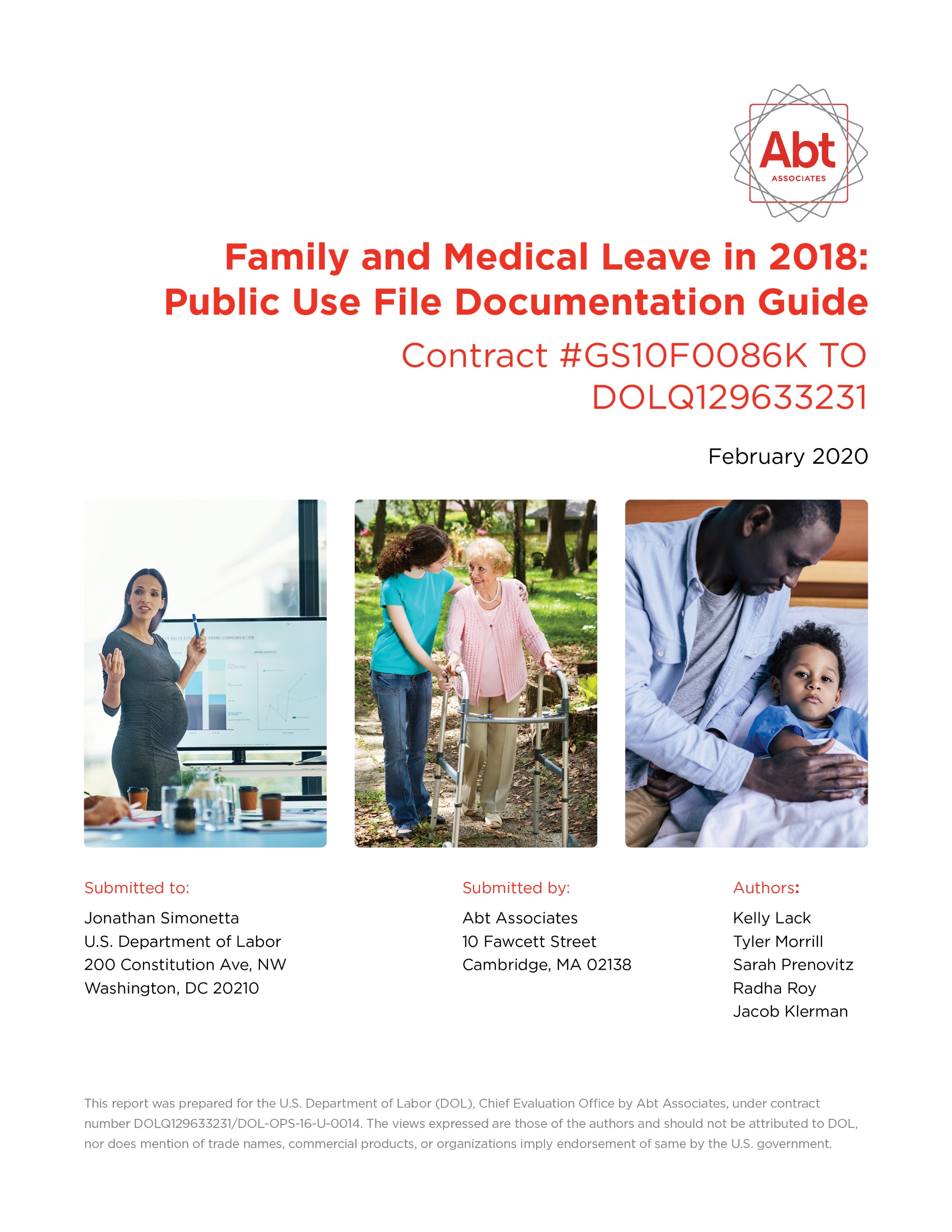
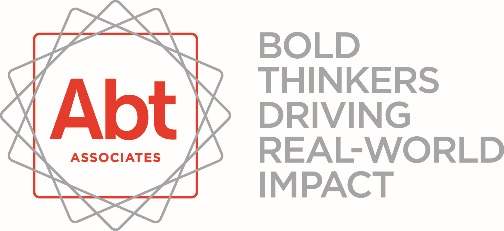
****



Abt Associates | 6130 Executive Boulevard | Rockville, MD 20852

CONTENTS

[Introduction 1](#_Toc31877092)

[Chapter 1: 2018 FMLA Worksite Survey 2](#_Toc31877093)

[Section 1: Datasets and Code 2](#_Toc31877094)

[Data Files 2](#_Toc31877095)

[Data Analysis 2](#_Toc31877096)

Introduction

This documentation guide provides an overview of the public use file (PUF) containing data from the 2018 FMLA Worksite and Employee Surveys. The first chapter discusses the 2018 FMLA Worksite Survey. The second chapter discusses the Employee Survey. Within each chapter, Section 1 presents information on the PUF and appropriate analysis code. Section 2 provides important survey context, including information on response option experiments, identifiers, suppressed or modified variables, and key constructs. Section 3 describes other documentation for the PUF, including the codebook and survey instruments.

Chapter 1: 2018 FMLA Worksite Survey

This chapter provides detailed information about the data files, analysis code, survey context, identifiers and other specific variables, and other related documents for the 2018 FMLA Worksite Survey.

Section 1: Datasets and Code

Data Files

The public use data file for the 2018 FMLA Worksite Survey is provided in both SAS and Stata data formats. The data files are named:

* <SAS format> FMLA\_2018\_Worksite\_PUF.sas7bdat
* <Stata format> FMLA\_2018\_Worksite\_PUF.dta

A SAS format catalog called “fmla\_2018\_worksite\_formats.sas7bcat” and a SAS program called “FMLA\_2018\_Worksite\_Read\_in\_SAS\_Dataset.sas” are also provided. The program reads in the SAS dataset above and assigns formats from the formats catalog to the relevant variables.

Data Analysis

To accommodate the complex sampling design of the 2018 FMLA Worksite Survey, the appropriate survey software and procedures should be used. Two sets of weights were created for analyzing survey data: one set that weights the data by the number of employees working at each worksite (employee weights), and another set that weights the data by the number of worksites (worksite weights). The weights for the 2018 FMLA Worksite Survey were designed to adjust for several key factors: differential probabilities of selection across sampling strata, differential non-response across sampling strata, and any residual difference between the distribution of the weighted respondents and the target population (post-stratification). The worksite weights can be used to draw inferences about the population of employers in worksite-level analyses, whereas the employee weights can be used to provide employee-level estimates in employee-level analyses. Accordingly, the publicly available report “Employee and Worksite Perspectives of the Family and Medical Leave Act: Results from the 2018 Surveys” uses data weighted by the number of employees for reporting results that reflect the experience of workers (e.g., the proportion of employees who work at a worksite that offers paid sick leave), and uses data weighted by the number of worksites for reporting results that reflect the experience of worksites (e.g., how costly it is to manage FMLA). More information about the two sets of weights created for the 2018 FMLA Worksite Survey can be found in the FMLA 2018 Methodology Report.

Below, we present examples of the appropriate SAS® and Stata® syntax to calculate weighted means and frequencies.[[1]](#footnote-1) In addition, for both the SAS and Stata code examples, the code output is provided for the first worksite-weighted frequency and the first employee-weighted frequency for illustrative purposes.

#### SAS Code

##### Applying Worksite Weights

Appropriate code to run the analysis in SAS applying worksite weights follows.

###### Industry frequencies on full sample

PROC SURVEYFREQ data=<libname>.FMLA\_2018\_Worksite\_PUF VARMETHOD = brr;

WEIGHT Estab\_Wt;

REPWEIGHT BootStrap\_Estab\_Wt1 - BootStrap\_Estab\_Wt250;

TABLES Q4\_CAT;

RUN;

*Output:*

| **Data Summary** | |
| --- | --- |
| **Number of Observations** | 2206 |
| **Sum of Weights** | 9446660 |

| **Variance Estimation** | |
| --- | --- |
| **Method** | BRR |
| **Replicate Weights** | FMLA\_2018\_WORKSITE\_PUF |
| **Number of Replicates** | 250 |

| **Organization's main industry** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Q4\_CAT** | **Frequency** | **Weighted Frequency** | **Std Err of Wgt Freq** | **Percent** | **Std Err of Percent** |
| **Manufacturing** | 584 | 1283713 | 4.70488E-6 | 13.5891 | 0.0000 |
| **Retail** | 434 | 1918384 | 0.0000188 | 20.3075 | 0.0000 |
| **Services** | 562 | 2864417 | 5.56912E-6 | 30.3220 | 0.0000 |
| **Other** | 626 | 3380146 | 3.4956E-6 | 35.7814 | 0.0000 |
| **Total** | **2206** | **9446660** | **0.0000223** | **100.0000** |  |

###### Mean percentage of female employees across full sample

PROC SURVEYMEANS data=<libname>.FMLA\_2018\_Worksite\_PUF VARMETHOD = brr;

WEIGHT Estab\_Wt;

REPWEIGHT BootStrap\_Estab\_Wt1 - BootStrap\_Estab\_Wt250;

VAR Q7\_PCT;

RUN;

###### Industry frequencies on subsample of worksites that self-report being covered by FMLA

PROC SURVEYFREQ data=<LIBNAME>.FMLA\_2018\_Worksite\_PUF VARMETHOD = brr;

WEIGHT Estab\_Wt;

REPWEIGHT BootStrap\_Estab\_Wt1 - BootStrap\_Estab\_Wt250;

TABLES Q4\_CAT \* COVERED\_SELF\_REPORTED;

RUN;

###### Mean percentage of female employees by self-reported FMLA coverage status

PROC SURVEYMEANS data=<libname>.FMLA\_2018\_Worksite\_PUF VARMETHOD = brr;

DOMAIN COVERED\_SELF\_REPORTED;

WEIGHT Estab\_Wt;

REPWEIGHT BootStrap\_Estab\_Wt1 - BootStrap\_Estab\_Wt250;

VAR Q7\_PCT;

RUN;

##### Applying Employee Weights

Appropriate code to run the analysis in SAS applying employee weights follows.

###### Industry frequencies on full sample

PROC SURVEYFREQ data=<libname>.FMLA\_2018\_Worksite\_PUF VARMETHOD = brr;

WEIGHT Emp\_Wt;

REPWEIGHT BootStrap\_Emp\_Wt1- BootStrap\_Emp\_Wt250;

TABLES Q4\_CAT;

RUN;

*Output:*

| **Data Summary** | |
| --- | --- |
| **Number of Observations** | 2206 |
| **Sum of Weights** | 114066684 |

| **Variance Estimation** | |
| --- | --- |
| **Method** | BRR |
| **Replicate Weights** | FMLA\_2018\_WORKSITE\_PUF |
| **Number of Replicates** | 250 |

| **Organization's main industry** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Q4\_CAT** | **Frequency** | **Weighted Frequency** | **Std Err of Wgt Freq** | **Percent** | **Std Err of Percent** |
| **Manufacturing** | 584 | 18931469 | 1132779 | 16.5968 | 0.9931 |
| **Retail** | 434 | 25845555 | 781684 | 22.6583 | 0.6853 |
| **Services** | 562 | 28769323 | 4338938 | 25.2215 | 3.8039 |
| **Other** | 626 | 40520337 | 4130910 | 35.5234 | 3.6215 |
| **Total** | **2206** | **114066684** | **0.0002236** | **100.0000** |  |

###### Mean percentage of female employees across full sample

PROC SURVEYMEANS data=<libname>.FMLA\_2018\_Worksite\_PUF

VARMETHOD = brr;

WEIGHT Emp\_Wt;

REPWEIGHT BootStrap\_Emp\_Wt1 - BootStrap\_Emp\_Wt250;

VAR Q7\_PCT;

RUN;

###### Industry frequencies on subsample of worksites that self-report being covered by FMLA

PROC SURVEYFREQ data=<libname>.FMLA\_2018\_Worksite\_PUF VARMETHOD = brr;

WEIGHT Emp\_Wt;

REPWEIGHT BootStrap\_Emp\_Wt1 - BootStrap\_Emp\_Wt250;

TABLES Q4\_CAT \* COVERED\_SELF\_REPORTED;

RUN;

###### Mean percentage of female employees by self-reported FMLA coverage status

PROC SURVEYMEANS data=<libname>.FMLA\_2018\_Worksite\_PUF

VARMETHOD = brr;

DOMAIN COVERED\_SELF\_REPORTED;

WEIGHT Emp\_Wt;

REPWEIGHT BootStrap\_Emp\_Wt1 - BootStrap\_Emp\_Wt250;

VAR Q7\_PCT;

RUN;

#### Stata Code

##### Applying Worksite Weights

Appropriate code to run the analysis applying worksite weights in Stata follows.

Unlike in SAS, where the specification of the sampling design and replicate weights is included in the PROC syntax itself, in Stata the user first declares a survey design and then specifies estimation procedures. Thus, one needs to use the following “settings” for each of the analyses that follow:

svyset [pw=estab\_wt], vce(bootstrap) bsrw(bootstrap\_estab\_wt\*) mse

As noted, each of the following analyses needs to be preceded by the worksite weight “settings” command above.

###### Industry frequencies on full sample

svy: tab q4\_cat

Number of obs = 2,206

Population size = 9,446,660

Replications = 250

|  |  |
| --- | --- |
| **Organization's Main Industry** | **Proportion** |
| Manufacturing | .1359 |
| Retail | .2031 |
| Services | .3032 |
| Other | .3578 |
| **Total** | **1** |

###### Mean percentage of female employees across full sample

svy: mean q7\_pct

###### Industry frequencies on subsample of worksites that self-report being covered by FMLA

svy, subpop(if covered\_self\_reported==1): tab q4\_cat

###### Mean percentage of female employees by self-reported FMLA coverage status

svy: mean q7\_pct, over(covered\_self\_reported)

##### Applying Employee Weights

Appropriate code to run the analysis applying employee weights in Stata follows.

Use the following “settings” for each of the analyses that follow:

svyset [pw=emp\_wt], vce(bootstrap) bsrw(bootbtrap\_emp\_wt\*) mse

As noted, each of the following analyses needs to be preceded by the employee weight “settings” statement above.

###### Industry frequencies on full sample

svy: tab q4\_cat

Number of obs = 2,206

Population size = 114,066,684

Replications = 250

|  |  |
| --- | --- |
| **Organization's Main Industry** | **Proportion** |
| Manufacturing | .166 |
| Retail | .2266 |
| Services | .2522 |
| Other | .3552 |
| **Total** | **1** |

Key: proportion = cell proportion

###### Mean percentage of female employees across full sample

svy: mean q7\_pct

###### Industry frequencies on subsample of worksites that self-report being covered by FMLA

svy, subpop(if covered\_self\_reported ==1): tab q4\_cat

###### Mean percentage of female employees by self-reported FMLA coverage status

svy: mean q7\_pct, over(covered\_self\_reported)

Section 2: Survey Context

This Section provides important survey context and detailed explanations of modifications to variables in the PUF dataset. This Section begins by detailing considerations related to response option experiments and question context. The next subsection describes modifications to survey variables. The final subsection provides detailed definitions of important survey constructs.

Response Option Experiments

A few questions on the 2018 FMLA Worksite Survey were the subject of wording or ordering experiments. In these experiment questions, response options could be presented to a given respondent in one of two slightly different ways. The two sets of response options either had slightly different labels (for example, “Most (half or more)” versus “Half or more”), or they were presented in a slightly different order (for example, “Neither positive nor negative” could either be presented last among the options, or between the options “Somewhat positive” and “Somewhat negative”). The experiments involved questions Q11, Q14a, Q25, Q27, Q52, and Q56. Each respondent was randomly assigned once, to see either the first set of response options for all experiment questions or to see the second set of response options for all experiment questions. The results of these response option experiments are described in the FMLA 2018 Methodology Report.

In the PUF, the *RESPONSE\_OPTION\_VARIANT* variable indicates whether a respondent saw the first set of response options for all experiment questions or saw the second set of response options for all experiment questions. The *RESPONSE\_OPTION\_VARIANT* variableis identical to the *samp1* variable that can be found on the instrument.

For each question in the experiment, all responses to that question are presented in a single variable. For example, the variable Q14A contains the combined responses of both respondents who saw the first set of response options (“All”, “Most (half or more), “Some (less than half)”, or “None”) and respondents who saw the second set of response options (“All,” “Half or more”, “Less than half”, and “None”).

Question Context

***Instrument format:*** The 2018 FMLA Worksite Survey was originally administered via phone and web to knowledgeable respondents from worksites after initial phone contacts with the worksites. Shortly after data collection began, the mode of initial contact transitioned from phone to mail and email, and screening questions were added to the web survey. With this transition, prospective respondents were emailed a choice of survey access including a unique link to their web survey and general survey website and login information. Respondents who could not be reached or did not complete the web survey were then called by interviewers to take the survey via phone using CATI software. The majority of 2018 FMLA Worksite Survey responses (86%) are web-based. More information about the data collection protocol can be found in the publically available FMLA 2018 Methodology Report.

***Skip Logic***: The survey skip logic is the same in the phone interviews and web surveys.

***Questions Added During Data Collection:*** Screening questions were added to the web survey a few weeks after the start of data collection, when the initial contact mode was changed from phone to email and mail. The screening questions were intended to verify the eligibility of the firm and identify the correct respondent at the firm who would be sufficiently knowledgeable to answer the survey.

***Question Numbering:*** The survey question numbering is designed to maximize consistency with the 2012 FMLA Worksite Survey question numbering. In some cases, 2012 FMLA Worksite Survey questions were omitted in the 2018 FMLA Worksite Survey. Consequently, consecutive survey questions in 2018 FMLA Worksite Survey (such as Q35 and Q37) are not always numbered consecutively.

***Variable Naming:*** The “n” prefix is included in variable names for survey questions that are new in the 2018 FMLA Worksite Survey and for survey questions where response options are different from those in the 2012 FMLA Worksite Survey.

***Variable Types:*** All variables in the PUF are numeric.

***Missing Codes***: The codebook and value labels detail which types of missing values are associated with each variable. The universe of missing codes includes the following:

* “.”: A general missing code representing responses of “don’t know” or “refused” to survey questions. In some cases, this code also includes special missing scenarios (e.g., inconsistent answers).
* “.S”: The respondent skipped the question due to their previous responses (valid skip). In a limited number of cases, respondents answered questions they were not supposed to have been shown. These responses are also set to .S.

Suppressed and Modified Variables

For the PUF, modifications were made to some variables in order to protect respondent privacy. Modifications included collapsing response categories, top-coding, and suppressing variables completely.

* With the exception of S8, no screener variables are included in the PUF.
* Q5 (NAICS code) was suppressed to protect respondent privacy. The respondent’s industry is provided in the Q4\_CAT PUF variable.
* All “other, specify” text responses are suppressed to protect respondent privacy.
* The only geographic information on 2018 FMLA Worksite Survey respondents included in the PUF are the respondent’s Census region and an indicator for whether the worksite was located in a paid leave state. Information on paid leave state classification can be found in the “Identifiers and Other Important Constructs” subsection below. All geographic information on respondents other than Census region and paid leave state was suppressed to protect respondent privacy.

In addition, some survey questions with small cell sizes were modified to protect respondent privacy. Variables that were originally continuous and that have been grouped into categories are indicated by the suffixes of “\_CAT” in the variable name.

* **Variables with discrete response options**
* For Q27\_CAT, a new code of 1 was created encompassing categories 1 (“All”) and 2 (“Most (half or more)” for response option variant 1, “Half or more” for response option variant 2).
* **Continuous variables grouped into categories**
* For Q1, a continuous variable was first created based on responses to Q1, Q1.1, and Q1.2. When a respondent provided an exact number of employees in Q1, that value was used for the continuous variable. Alternatively, when a respondent indicated a nonzero number of employees in Q1.1 and then only provided a range in Q1.2 (rather than an exact number in Q1), the midpoint of that range was taken for the continuous variable. A categorical variable, Q1\_CAT, was then created by grouping the values of that continuous variable into categories. Note that the categories in Q1\_CAT are identical to the Q1.2 ranges, so responses to Q1.2 are preserved exactly in Q1\_CAT. The same recoding was also done for Q2, Q2.1, and Q2.2 to create Q2\_CAT, and for Q3 and Q3.1 to create Q3\_CAT.
* To create Q57\_1\_A\_CAT, responses to Q57\_A were converted into days to create a single continuous variable as described in the “Other Variable Changes” subsection below. A categorical variable was then created by grouping the values of that continuous variable in days into categories. The same was done for Q57\_1\_B to create Q57\_1\_B\_CAT, for Q57\_1\_C to create Q57\_1\_C\_CAT, for Q57\_1\_D to create Q57\_1\_D\_CAT, for Q57\_1\_E to create Q57\_1\_E\_CAT, and for Q57\_1\_F to create Q57\_1\_F\_CAT.
* To create Q58\_CAT, values of Q58 that were 4 or higher were grouped into a new category, “4 or more.”
* To create Q59\_CAT, values of Q59 that were 1 or higher were grouped into a new category, “1 or more.” The same was done for Q60 to create Q60\_CAT.

Other Variable Changes

In addition to the above modifications to preserve respondent confidentiality, several other changes were made for other reasons. Specifically:

* **Converting time units into a common metric**

Survey questions that asked respondents to provide a duration of time (e.g. the items in the Q57\_1 series) were converted into units of “days” as follows:

* Converting hours to 1 day = *T* hours/8; i.e., 8 working hours per day.
* Converting weeks to 1 day = *T* weeks × 5; i.e., 5 working days per week.
* Converting months to 1 day = *T* months × 22.5; i.e., 4.5 weeks per month × 5 working days per week.
* **Updating firm size variable**

In some cases, respondents reported total firm sizes in Q1 that were smaller than the worksite size reported in Q2 or the number of employees within 75 miles reported in Q3. In these cases Q1 was updated to be the larger of Q2 and Q3, and a flag (Q1\_CHANGED\_BY\_Q2\_OR\_Q3) was created to document the change.

Identifiers and Other Important Constructs

This subsection discusses the identifier variables in the PUF, as well as other important constructs derived from firm profile data or survey responses.

* **Identifier variables:**
* **worksiteID:** randomly generated unique identifier for each Worksite Survey respondent.
* **interviewerid:** randomly generated unique identifier for each interviewer who administered the phone interview (web surveys did not involve interviewers).
* **survey\_type:** identifier of survey mode (phone survey vs. web-based survey).
* **Important constructs:**
* **paid\_leave\_state:** Denotes whether worksite is in a paid-leave state based on sampling frame data. All worksites in California, New Jersey, and Rhode Island were coded as being in paid leave states, as these states had paid leave laws in effect throughout the 12-month survey recall period for all respondents. New York implemented a paid leave program on January 1, 2018, which was after the survey fielding period had started. Worksites in New York were coded as being in a paid leave state if at least half of the survey recall period occurred after paid leave went into effect, and not otherwise.[[2]](#footnote-2)
* **fmla\_eligible\_pct:** Calculated percentage of employees at worksite estimated to be eligible for FMLA. Estimated eligibility was calculated based on survey responses about the number of employees at the worksite provided in Q2, the number of employees who had worked at the firm for at least one year provided in Q8, and the number of employees with sufficient tenure who had worked at least 1,250 hours during the reference period provided in Q9.
* **covered\_self\_reported:** Indicator for whether the respondent reports that FMLA applies to the worksite in Q17. Respondents who answered “Applies” to Q17 were coded as reporting they were covered by FMLA. Respondents who answered “Does not apply” or “Not sure” in Q17, or who refused to answer the question, were coded as not reporting they were covered by FMLA. This construct was used to conduct subgroup analysis in the survey report.

Section 3: Other PUF and Contextual Files

In addition to the data files discussed in Section 1 of this chapter, the 2018 FMLA Worksite Survey codebook is provided as a part of the PUF download package. The codebook is included in the documentation package as an Excel workbook named “FMLA\_2018\_Worksite\_PUF\_Codebook.xlsx.” The 2018 FMLA Worksite Survey instrument can be found in the appendices of the FMLA 2018 Methodology Report.

Chapter 2: 2018 FMLA Employee Survey

This chapter provides detailed information about the data files, analysis code, survey context, identifiers and other specific variables, and other related documents for the 2018 FMLA Employee Survey.

Section 1: Datasets and Code

Data Files

The public use data file for the 2018 FMLA Employee Survey is provided in both SAS and Stata formats. The files are named:

* <SAS format> **FMLA\_2018\_Employee\_PUF.sas7bdat**
* <Stata format> **FMLA\_2018\_Employee\_PUF.dta**

A SAS format catalog called “fmla\_2018\_employee\_formats.sas7bcat” and a SAS program called “FMLA\_2018\_Employee\_Read\_in\_SAS\_Dataset.sas” are also provided. This program reads in the SAS dataset above and assigns SAS formats from the provided formats catalog to the relevant variables.

Data Analysis

To accommodate the complex sampling design of the 2018 FMLA Employee Survey, the appropriate survey software and procedures should be used. A set of bootstrap replicate weights was created to facilitate correct variance estimation in analysis of 2018 FMLA Employee Survey data. More information about these weights can be found in the FMLA 2018 Methodology Report.

Below, we present examples of the SAS® and Stata® syntax that must be used to properly calculate weighted frequencies. In addition, for both the SAS and Stata code examples, the code output is provided for the first frequencies for illustrative purposes.

#### SAS Code

Appropriate code to run the analysis in SAS follows.

###### Frequencies of FMLA eligibility on full sample

PROC SURVEYFREQ data=<LIBNAME>. FMLA\_2018\_employee\_PUF

VARMETHOD = brr;

WEIGHT combo\_trimmed\_weight;

REPWEIGHT cmb\_bsrw1 - cmb\_bsrw200;

TABLES fmla\_eligible;

RUN;

*Output:*

| **Data Summary** | |
| --- | --- |
| **Number of Observations** | 4470 |
| **Sum of Weights** | 140590310 |

| **Variance Estimation** | |
| --- | --- |
| **Method** | BRR |
| **Replicate Weights** | FMLA\_2018\_EMPLOYEE\_PUF |
| **Number of Replicates** | 200 |

| **Respondent was eligible for FMLA at start of reference period (12 months prior to survey date) based on stated work tenure, weekly hours, and employer size.** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **fmla\_eligible** | **Frequency** | **Weighted Frequency** | **Std Err of Wgt Freq** | **Percent** | **Std Err of Percent** |
| **Not FMLA eligible** | 1775 | 58091904 | 2179192 | 44.2839 | 1.6637 |
| **FMLA Eligible** | 2350 | 73088800 | 2372930 | 55.7161 | 1.6637 |
| **Total** | 4125 | 131180703 | 1300132 | 100.0000 |  |
| **Frequency Missing = 345** | | | | | |

###### Frequencies of FMLA eligibility on subsample of males

PROC SURVEYFREQ data=<LIBNAME>. FMLA\_2018\_EMPLOYEE \_PUF

VARMETHOD = brr;

WEIGHT combo\_trimmed\_weight;

REPWEIGHT cmb\_bsrw1 - cmb\_bsrw200;

TABLES fmla\_eligible \* male;

RUN;

#### Stata Code

Appropriate code to run the analysis in Stata follows.

Use the following “settings” for each of the analyses that follow:

svyset [pw=combo\_trimmed\_weight], vce(bootstrap) bsrw(cmb\_bsrw\*) mse

As noted, each of the following analyses needs to be preceded by the “settings” command above.

###### Frequencies of FMLA eligibility on full sample

svy: tab fmla\_eligible

*Output:*

Number of obs = 4,125

Population size = 131,180,703

Replications = 200

|  |  |
| --- | --- |
| **Respondent was eligible for FMLA at start of reference period (12 months prior to survey date) based on stated work tenure, weekly hours, and employer size** | **Proportion** |
| Not FMLA eligible | .4428 |
| FMLA Eligible | .5572 |
| **Total** | **1** |

###### Frequencies of FMLA eligibility on subsample of males

svy, subpop(if male==1) : tab fmla\_eligible

Section 2: Survey Context

This Section provides important survey context and detailed explanations of modifications to variables in the PUF dataset. This Section begins by detailing considerations related to the change in instrument format, survey loops, skip logic, and unique questions. Next, it describes reasons for modifications to survey variables. The final part of the section provides detailed definitions of important survey constructs.

Question Context

***Change in instrument format:*** The original random digit dialing (RDD) survey was transitioned to a web-based format during survey administration. The data from the two surveys was then combined for analysis. The majority of respondents (83%) are web-based. The screener sections between the RDD and web-based survey differ because the web-based survey company’s respondent profiles already contained respondent demographic data. Section D of the web-based survey does not include a question about sexual orientation for the same reason. Respondents were screened for employment in the past 12 months on the individual level for the web survey and the cell phone version of the RDD survey. For the landline version of the RDD survey, respondents were screened for past-year employment at the household level, with attempts made to find an eligible respondent in the household. Moreover, the RDD survey provided prompt questions for interviewers to ask respondents, with the interviewer updating the respondent’s answer if he indicated a different answer upon confirmation. By contrast, the confirmation questions on the web survey were programmed as separate questions. Context on the change from an RDD to web-based survey, as well as information on data collection, response rates, and weight development can be found in the publically available FMLA 2018 Methodology Report.

***Skip Logic***: The survey skip logic is the same in the RDD and web surveys. In one case involving questions about how employees covered wages while on leave and leave-related financial hardship (questions A53 and A54), the skip logic for the web survey was modified slightly to better reflect the intent of the original skip logic. Consequently, a small number of RDD respondents were shown A53/A54 although they would not have had they been administered the web survey. The team reviewed these responses and included them in both the public use dataset and the tabulations published in the official report. In a second case involving the survey question about employers requiring multiple doctor visits to obtain initial medical certification (A30), a correction was made to the RDD skip logic when the web survey was developed. Consequently, fourteen RDD Respondents should have answered question A30 but did not and are coded as missing.

***Questions Added During Phone Data Collection:*** The E0K variable series was added shortly after RDD phone interviews began to better handle and clarify inconsistent respondent answers. These variables were present for the entire duration of the web survey.

***Longest/Most Recent Leave Loops***: For employees taking leave, the survey asked about up to two leaves in questions A5 through A19; that is, there were up to two “loops” for each of these questions for each respondent. The loops are indicated with the suffix or pre-suffix[[3]](#footnote-3) “\_mr” (most recent) or “\_long” (longest). When the most recent leave represents the respondent’s longest leave, the “\_mr” variable represents both the longest and most recent leave and the “long” variables are skipped over (skipped missing code = “.S”).

***Variable Types:*** All variables are numeric, except ne15coded. A character type was used for this variable in order to preserve the response values representing NAICS industry code ranges. For example, codes “31-33” represent manufacturing industries.

***Missing Codes***: The Codebook and value labels detail which types of missing scenarios are associated with each variable. The universe of missing codes includes the following:

* “**.**”: A general missing code representing responses of “don’t know” or “refused” to survey questions. In some cases, this code also includes special missing scenarios (e.g., inconsistent answers).
* “.S”: The respondent skipped the question due to their previous responses (valid skip).
* “.U”: For annual earnings questions d3 and d4, this code represents unbelievably high or low values that were deemed likely to be respondent errors. For occupation question e16, this code represents uncodeable responses.
* “.N”: In order to ask as broad a range of questions as possible without putting undue burden on respondents, each respondent was randomly asked only 4 of the subparts of the question asking about which types of leave the respondent believes FMLA covers (question e4a). Response options that were not randomly selected for a given respondent take a value of “.N” in the corresponding PUF dummy variables e4a\_a – e4a\_l. Response options for this question were randomly selected within the following parameters:
* 2 questions from subpart A, B, C, F, J, K, or L;
* 1 question from subpart G, H, or I;
* 1 question from subpart D or E.

Suppressed and Modified Variables

For the PUF, modifications were made to some variables in order to protect respondent privacy. Modifications included collapsing response categories, top-coding, and suppressing variables completely. Variables that have been modified are indicated by the suffixes of “\_CAT” in the variable name.

* Unless otherwise noted below, no screener variables are included in the PUF.
* Questions A13, A16, A17, and A18 were suppressed to protect respondent privacy, with leave length information being provided in the A19 PUF variables.
* No geographic information on 2018 FMLA Employee Survey respondents is included in the PUF to protect respondent privacy. The one exception is an indicator for whether the respondent resided in a paid leave state.

In addition, the following changes were made for other reasons:

* Variable names for survey questions that are new in the W4 PUF or response options that have changed since the W3 survey due to the addition of new response categories have a lowercase “n” prefix in the variable name.
* Response values that were truncated for the purpose of specific report tables (noted in table notes) were coded into the appropriate category in the corresponding PUF variable (e.g., A19, B4) so that these PUF variables represent the complete set of survey responses.

In addition, the following survey questions with discrete response options were modified due to small cell sizes:

* For A5, new codes of 20 and 21 were created encompassing response categories 2 and 4 (code 20), as well as 6, 7, 9, and 10 (code 21) for the most recent leave. For the longest leave, a new code of 20 was created encompassing all categories except 1, 3, 5, and 13.
* For A6, a new code of 6 was created encompassing categories 1, 2, and 4.
* For the A10 longest leave variable, a new code of 6 was developed encompassing categories 4 and 5.
* For A19c, a new code of 7 was created encompassing categories 7, 8, 9, 10, and 11.
* For B6, new codes of 20 and 21 were created encompassing categories 2 and 4 (code 20), as well as 6, 7, 10, and 17 (code 21).
* For E0C, D7, and D8, a new code of 4 was created encompassing categories 4 and 5.
* For D7, a new code of 4 was created encompassing categories 4 through 7, representing 4 through 7 children.
* For D8, a new code of 3 was created encompassing categories 3 through 7, representing 3 through 7 people over age 65.

Survey questions that asked respondents to provide a duration of time were converted into units of “days” as follows:

* Converting hours to 1 day = *T* hours/8; i.e., 8 working hours per day.
* Converting weeks to 1 day = *T* weeks × 5; i.e., 5 working days per week.
* Converting months to 1 day = *T* months × 22.5; i.e., 4.5 weeks per month \* 5 working days per week.

Identifiers, Screeners, and Other Important Constructs

This Section discusses the constructs included in the PUF. There are three classes of constructs: identifier variables, variables created from screener or web panel respondent profile data, and subgroup constructs created from survey variables that were analyzed in the official report.

* Identifier variables:
* empid: randomly generated unique identifier for each 2018 FMLA Employee Survey respondent.
* interviewerid\_rdd: randomly generated unique identifier for each interviewer who administered the RDD survey (web surveys did not involve interviewers).
* survey\_type: identifier of survey mode (RDD phone survey vs. web-based survey).
* Variables created from screener or web panel respondent profile data:
* age\_cat: Created from screener and Section D data in RDD survey, created from web panel respondent profile data for web survey.
* race\_cat: Created from Section D data in RDD survey, created from web panel respondent profile data for web survey.
* gender\_cat: Created from screener data for RDD survey, created from web panel respondent profile data for web survey.
* educ\_cat: Created from screener data.
* leave\_cat: Created from screener data.
* govt\_emp: Created from screener data.
* num\_jobs: Created from screener data.
* Subgroup constructs analyzed in report tables:
* paid\_leave\_state: Whether respondent resides in a paid-leave state. Paid-leave states are California, Rhode Island, and New Jersey. New York implemented paid leave on January 1, 2018, which was during the survey administration period. To account for this, New York respondents who completed the survey on or after July 1, 2018 are considered to be in a paid leave state (because the majority of their 12-month recall period occurred after paid leave went into effect on January 1, 2018).
* low\_wage: Whether respondent earned less than $15/hour at main job at start of reference period (12 months prior to the survey). This variable is defined based on both income data provided in D3b and weekly hours data provided in the e0 questions. It is imputed for respondents without wage data using a probit regression imputation model based on wage and demographic characteristics from the survey. There are 271 cases where respondents’ annual income is under $30,000 in PUF variable D3abc\_CAT, but they are not low wage in the LOW\_WAGE variable. This is due to respondents having a wage greater than $15/hour but working few weekly hours, missing source variable data and being imputed as not low wage, or simply having different response periods between questions (d3b asks about current income if respondent is currently employed and the e0 questions ask about weekly hours 12 months ago). Conversely, there are 3 respondents who were imputed as low wage but have incomes greater than $65,000 in PUF variable D3abc\_CAT. These 3 values were imputed due to missing source variable data.
* fmla\_eligible: Calculated FMLA eligibility at the start of the reference period (12 months prior to the survey date). Eligibility was calculated based on survey responses about employer size, employment tenure, and hours worked, all at the start of the reference period.
* eligibility\_cat: FMLA eligibility variable including reasons for ineligibility for those who are ineligible.
* fmla\_elig\_if15hrwk, fmla\_elig\_if30hrwk, fmla\_elig\_if30emp, fmla\_elig\_if20emp: FMLA eligibility simulation variables representing FMLA eligibility if eligibility criteria were modified (15 hours worked per week, 30 hours worked per week, employer size of 30 employees, employer size of 20 employees).
* a19\_mr\_0\_10d, a19\_mr\_11\_40d, a19\_mr\_41\_60d, and a19\_mr\_61Pd: Length of leave subgroup variables representing length of leave categories (0-10 days, 11-40 days, 41-60 days, 61 or more days). Leave length is doubled for respondents currently on leave (A3) and FMLA-ineligible leave reasons (A5) are set to missing.
* a5\_ownIllness, a5\_newchild, a5\_chc A5\_adulthealthlt65, a5\_adulthealthgt65, a5\_nonfmlaperson: Most recent leave reason subgroup variables (A5).
* Other constructs:
* d3abc\_cat: Annualized own current earnings based on D3a, D3b, D3c, and D3d. Weekly hours reported in e0b and e0g were used to annualize values for respondents who reported income as an hourly wage. For the 72 respondents who were currently employed, not employed 12 months ago, and reported their income as an hourly wage, e0b and e0g were skipped and thus their income could not be annualized. These 72 respondents take values of “Don’t Know/Refused/Not able to be annualized” in the D3abc\_CAT PUF variable.
* nd4\_cat: Family earnings derived by creating categories from both continuous D4 responses and the midpoint of responses to D4a through D4j. Question D4 asks about earnings in the last 12 months whereas D3 asks about current earnings if the respondent is currently employed, so responses to these questions are not always consistent.

Section 3: Other PUF and Contextual Files

In addition to the data files discussed in Section 1 of this chapter, the 2018 FMLA Employee Survey codebook is provided as a part of the PUF download package. The codebook is included in the documentation package as an Excel workbook named “FMLA\_2018\_Employee\_PUF\_Codebook.xlsx.” The 2018 FMLA Employee Survey instruments can be found in the appendices of the FMLA 2018 Methodology Report.

1. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.Stata and all other Stata Corp LP product or service names are registered trademarks or trademarks of StataCorp LP in the USA and other countries. ® indicates USA registration. [↑](#footnote-ref-1)
2. More information about paid leave state classification can be found in the FMLA 2018 Supplemental Results Appendix. [↑](#footnote-ref-2)
3. That is, the text immediately preceding the suffix (e.g., A19\_MR\_CAT). [↑](#footnote-ref-3)