Balancing the Needs of Families and Employers:

Family and Medical Leave Surveys

Methodology Report
Draft 1
October 5, 2001

Westat
1650 Research
Rockville, MD 20850

This report was prepared under contract MS-23F-8144H from the Department of Labor.

Chapter 1 Overview

The 2000 Surveys of Employees and Establishments, conducted by Westat for the Department of Labor (DOL), aimed to provide estimates of the use of leave under the Family and Medical Leave Act since January 1, 1999, and to examine the Act's effects on U.S. employees and businesses. The surveys were designed to follow up on surveys conducted for the Commission on Family and Medical Leave by the University of Michigan's Survey Research Center (1995 Survey of Employees), and by Westat (1995 Survey of Establishments). This report summarizes the methods used by Westat to conduct the 2000 Surveys of Employees and the 2000 Survey of Establishments. The results from these surveys are reported in *Balancing the Needs of Families and Employers: Family & Medical Leave Surveys, 2000 Update.* (referred to as *Balancing the Needs*), published by Westat in 2001.

Chapter 2 describes methods used for the 2000 Survey of Employees, while Chapter 3 describes the methods for the 2000 Survey of Establishments.

Chapter 2

2000 Survey of Employees

The 2000 Survey of Employees was a telephone survey designed to sample U.S. residents who had been employed at any time since January 1, 1999. Telephone numbers were randomly generated using a list-assisted procedure. Once a household was contacted, the interviewer identified potential respondents who had been employed since January 1, 1999. Three unique samples of respondents were identified and interviewed: (1) those who had taken leave from work for a family or medical reason; (2) those who had needed but not taken this type of leave; and (3) those who were employed but had neither taken or needed leave during the period covered by the survey. A total of 2,558 interviews were completed. The final weighted response rate for the survey was 58.3%.

In this chapter, we describe the methods used to conduct and analyze the survey, including the following:

- Survey instrument development;
- Interviewer training;
- Sample design;
- Data collection;
- Response rate calculations;
- Weighting;
- Variance estimation;
- Defining key study measures; and
- Comparing 1995 and 2000 data.

2.1 Survey Instrument Development

The instrument development process included the following steps:

- Revising the 1995 screener and extended interview instruments;
- Conducting cognitive interviews to test the revised instruments; and
- Programming the extended instrument as a computer-assisted telephone interview (CATI).

These steps are described in greater detail below.

2.1.1 Instrument Revisions

Two instruments were developed for the 2000 Survey of Employees: the screener instrument and the extended instrument. The screener instrument was used to identify and obtain contact information for potential respondents that matched study criteria (i.e., leave-takers, leave-needers, and employed only since January 1, 1999). The extended interview instrument was used to collect the study data.

The general approach taken to develop instruments was to build on the 1995 instruments. First, obsolete items were deleted. Although some items were revised, emphasis was placed on maximizing comparability with the data from the 1995 survey. New items were added as needed to address emerging issues in family and medical leave. The new items included, for example, obtaining reasons for leaves beyond the single longest leave, more in-depth questions about paid leave, and items regarding the effects of leave on the leave-taker and his/her family.

The extended interview instrument for the 2000 Survey of Employees consisted of four major sections:

- Leave experiences, asked specifically of leave-takers;
- Reasons for needing leave and for not taking leave, asked of leave-needers;
- Employment status and their opinions about family and medical leave; and
- Demographic information on respondents.

The final version of both the screening and extended instruments provided in Appendix D of *Balancing the Needs*.

2.1.2 Cognitive Testing

In order to gain insights into potential problems with survey items, project staff conducted cognitive interviews with a draft version of the extended interview instrument. Interviews were conducted with eight Westat employees who had taken leave in the past 18 months, including four for their own health conditions, two for maternity-related reasons, one for a child's health condition, and one for a spouse's health condition. Length of leaves ranged from 3 weeks to several months. The most important issues stemming from these interviews are discussed in the paragraphs that follow.

When asked how many leaves they had taken (survey item A2), respondents whose leave was intermittent were unsure of how to count it. For example, one person had taken a week of leave per month for four months, followed by two months away from work, and then alternating between work and leave for another two months. Since this was considered a key question from the 1995 survey, we were reluctant to recommend any changes. Nevertheless, we learned that we would need to train interviewers how to handle this situation when presented by a confused respondent. Specifically, interviewers were instructed that intermittent leave for the same condition/reason should be counted as one leave (although multiple episodes for the same condition, such as having influenza on two different occasions, should be considered multiple leaves).

When asked to state the reason for their leave (survey item A3), some respondents answered in such a way that could be coded as more than one reason. The most common example was when a woman takes time off for maternity-related disability but also remains at home for some time after birth to care for her newborn. It was decided that the list of reasons in the survey should include one for this particular situation, in order to minimize interviewer confusion.

When asked how long their leave had lasted (survey item A3d), respondents frequently had difficulty answering, especially if they had taken intermittent leave. They were not sure whether to count only the time away from work, or the entire period of time spanned by the leave. It was decided that answers to A3d should be probed to ensure that we obtain leave lengths restricted to time away from work. This resulted in the addition of items A3e and A3f (it turned out that the vast majority of leave-takers in the survey did restrict their answer to time away from work when answering A3d).

The version of survey item A12 (regarding the impact leave had on the leave-taker or his/her family) that was pretested included "negative effect" as a potential response option. The cognitive interview respondents tended to be confused by this option, since they could not imagine how taking leave might have a negative effect on such things as physical health and emotional well-being. Consequently, the response options were limited to "positive effect" and "no effect."

In the draft interview used for the cognitive interviews, items in Section C (such as C3 and C6) specified "the federal Family and Medical Leave Act of 1993" (as was done in the 1995 survey). Respondents were often confused by the reference to 1993. Even those who

knew of the law did not associate it with a particular year, since they did not know when the law was enacted. The reference to 1993 may well have been important for the 1995 survey, but we felt the phrase may actually do more harm than good if used in the 2000 survey (e.g., it might imply that there are multiple versions of FMLA). Thus, the reference to 1993 was deleted from all items regarding FMLA.

2.1.3 CATI Programming

Once the instruments were finalized, they were programmed into Westat's computer-assisted telephone interviewing (CATI) system. This system includes standard software for managing random digit dial samples, respondent selection, and questionnaire management. During programming, the CATI system was tested at multiple levels to ensure quality control. These tests included checking skip patterns, timing survey administration, and checking the coordination of the scheduler management system with the questionnaire management system.

2.2 Interviewer Training

Training sessions (lasting approximately 8 hours) were conducted to prepare interviewers for administration of the survey. Project staff prepared training materials to familiarize interviewers with all aspects of the task. These materials included an interviewer's manual that described the background and purpose of the project, provided answers to commonly asked questions, and presented question-by-question specifications. The survey instrument was presented in this manual in the form of CATI screens, so that from the beginning of their training, trainees saw exactly what appears in the live interview.

A total of 51 interviewers were trained for the project. The training sessions simulated the actual conditions of the interview and required the active participation of all interviewers trained for the project. The sessions included interactive lectures and dyadic role-playing. Interactive lectures were used to present the basic objectives of the questionnaire. The trainer led interviewers through the instrument by calling on each of them to perform the role of the interviewer, while the trainer provided information from a script. Interviewers recorded responses on their computer terminal. In dyadic role-playing exercises, interviewers were paired. Within each dyad, one trainee performed the role of the interviewer while the other acted as the respondent. During training, interviewers were monitored by training and

supervising staff. Problems and remedies related to administering the questionnaire were discussed at the end of each session. Interviewers were not assigned real interviews until they had successfully completed training activities.

2.3 Sample Design

The survey was planned to include a sample of individuals aged 18 or older in U.S. households who were employed at any time between January 1, 1999 and the time of the survey (between 18 and 20 months, depending on when the interview occurred). A sample of approximately 24,500 telephone numbers was drawn from the universe of all known U.S. households with telephones, using list-assisted Random Digit Dial (RDD) methodology. In this approach, the sampling frame consisted of all 100-banks[1] of telephone numbers in the 50 states and Washington, DC with at least one listed residential telephone number. 100-banks without any listed telephone numbers are not sampled, resulting in less than 4 percent of households with telephone numbers being excluded from the frame. Thus, the sample frame represented virtually all employees, both the public and private sector, that had a telephone.

The frame of telephone numbers was sorted by 10 Census Divisions and a metropolitan/non-metropolitan area split within each region. Within each of the 10 regional metro categories, telephone exchanges were ordered from those serving the largest metropolitan statistical area/primary metropolitan statistical area to those serving the smallest. Statistical areas were further sorted by county. Within the 10 non-metro categories, states were ordered geographically. Within each state, non-metro counties were further ordered geographically. A systematic sample was then selected from the sorted frame of telephone numbers.

Two methods were used to reduce the cost associated with identifying business and nonworking telephone numbers. First, all sampled telephone numbers were electronically matched to a file of Yellow Pages lists of business numbers. Telephone numbers identified as business numbers were excluded from further calls. Second, Westat used an automated procedure to dial all sampled telephone numbers in order to identify nonworking and business numbers. Each telephone number was allowed to ring up to two times. In many cases, a tritone message, the distinctive three-tone sound heard when a nonworking number is reached was immediately detected. The telephone call was then discontinued and the number excluded

[1] A 100-bank is a set of 100 telephone numbers with the same first eight digits.

from further dialing. A Westat TRC staff member responded to answered phone calls to determine whether the number was a business number. Purging business and nonworking telephone numbers by these two methods introduced only a small amount of error, but reduced costs significantly.

For purposes of the survey, the population was divided into four groups: (1) leave-takers, (2) leave-needers, (3) employed-only, and (4) not employed. For each telephone number, an interviewer attempted to screen for eligibility by determining whether the household contained at least one person 18 years of age or older who had been employed since January 1, 1999. For all persons within a household meeting these criteria, the interviewer attempted to determine if they had taken (or needed without taking) family or medical leave since January 1, 1999. All persons identified as having taken or needed leave were eligible for the extended interview. Thus, more than one person per household could be interviewed. Those not having taken or needed leave (i.e., those who were employed only) were sub-sampled for the extended interview.

2.4 Data Collection

As described above, data for the Survey of Employees were collected by interviewers specially trained for the project using a Computer Assisted Telephone Interviewing (CATI) system. Interviewing began on July 15, 2000 and continued for approximately 10 weeks. A total of 2,558 interviews were completed: 1,229 with persons who took leave (for reasons covered by FMLA) since January 1, 1999; 203 with persons who needed leave (for a covered reason) but did not take it; and 1,126 with persons that were employed only and had not needed any family or medical leave.

2.5 Response Rate Calculations

The response rate for the Survey of Employees was computed in three steps. In the first step, a response rate was calculated for the screening interview, which identified eligible respondents in the household. In the second step, a response rate was calculated for the extended interview, which collected the data from the selected household respondent. In the third step, the two response rates were combined to produce the overall survey response rate. These steps are described in greater detail below.

2.5.1 Screening Interview Response Rate

The final outcomes of call attempts (at the screener level) were as follows:

	Number of Cases
Completed screeners, respondent selected	2,525
Completed screeners, no one selected	3,411
Complete screeners, no one eligible	1,509
Language problem	402
Maximum number of calls[2]	564
No answer	1,462
Answering machine	268
Non-residential	5,154
Non-working number	16,078
Refusals	2,013
Other non-response	41

The weighted response rate for the screener ranged from 67.5 percent to 69.3 percent. The range reflects different assumptions made about the eligibility of those telephone numbers where no respondent ever answered the telephone. These different assumptions are described in greater detail in Section 2.7. The lower screener weighted response rate (67.5%) was calculated using the following formula:

$$C/(C+R+.27NA+.6M + LP+MC+ONR)$$

where
- C = complete
- R = refusal
- NA = no answer
- M = answering machine
- LP = language problem
- MC = maximum calls
- ONR = other non-response

This assumed that a residential household existed for 27 percent of those calls where someone never answered the telephone and for 60 percent of those calls where the interviewers only encountered an answering machine. This is the standard formula used by Westat when computing response rates for random digit dial surveys. It is based, in part, on guidelines published by the Council of American Survey Research Organizations (CASRO). It

[2] When a person answered the phone but could not do the interview at that time, the phone number was called back multiple times over a number of time slices, up to a set maximum number of calls. For this study, phone numbers were called back at least ten times.

modifies these guidelines by reducing the number of "no answers" that are classified as eligible, based on research that tracks telephone numbers through the telephone company.

The higher response rate of 69.3 percent was computed using the following formula:

C/(C+R+M + LP+MC+ONR)

This formula excludes the calls where someone did not answer the telephone (NAs above), but includes calls that reached an answering machine. This rate is comparable to the method used by the University of Michigan for the 1995 Survey of Employees.

2.5.2 Extended Interview Response Rate

The final outcomes of call attempts at the extended level (for cases in which a screener was completed and at least one household member was selected for the survey) were as follows:

	Number of Cases
Completes, leave-takers[3]	1,277
Completes, leave-needers	226
Completes, employed-only	1,066
Respondent found to be ineligible	14
Language problem	36
Maximum number of calls	80
Non-residential	6
Non-working number	44
Refusals	240
Other non-response	116

The final weighted response rates for completing the extended interviews represent the proportion of interviews that were completed among those eligible and selected for the study. The rates, calculated separately for each type of respondent, were:

	Extended Interview Response Rate
Overall:	84.2%
Leave-taker	83.9%
Leave-needer	85.3%
Employed only	84.2%

[3] For analysis purposes, leave-takers were restricted to those whose leave had been for a reason covered by the FMLA. Similarly, leave-needers were restricted to those who needed leave for a reason covered by the Act. This explains why these numbers do not match the sample sizes reported in *Balancing the Needs*.

2.5.3 Final Response Rate

The final response rate was computed by multiplying each respective extended interview response rate by the screener response rate. For this step, the higher screener response rate of 69.3 percent was used to maintain comparability with the 1995 survey. The final weighted response rates for each type of respondent were:

	Final
Overall:	58.3%
Leave-taker	58.1%
Leave-needer	59.1%
Employed only	58.3%

2.6 Weighting

Two sets of weights were devised for the analysis of the 2000 Survey of Employees data. One set of weights ("preferred") was designed to produce estimates of the number of leave-takers and leave-needers with the smallest variance and least bias, using industry-standard weighting procedures. Two components of these weights, however, had not been used in the weighting for the 1995 survey, and would quite likely have an effect on the expected value of survey estimates. Since the prime purpose of the 2000 survey was to make comparisons to the 1995 survey, we devised a second set of weights ("consistent") so that it was possible to reweight the 1995 survey in a consistent manner. All the estimates in the report use the "consistent" set of weights. If a future family leave survey is conducted, then we recommend that "preferred" weights be calculated and used to produce revised estimates for 2000, which can then be compared to future survey estimates.

This section discusses the following issues:

- Method used to create the 2000 consistent weights;
- Modifications made to the 1995 weights to make them consistent with the 2000 consistent weights;
- Suggested method to create preferred weights for future comparisons; and
- Suggested weighting adjustment to reduce bias caused by changing respondents among the leave-takers, leave-needers, and employed-only categories.

2.6.1 Consistent Weighting Method

The weighting process entailed two major calculations. First, weights were calculated at the household level for the screening interview. Weights were then calculated at the person level, for the extended interview of persons who were identified as leave-takers, leave-needers, or employed-only. For both calculations, separate weighting steps were taken to calculate non-response adjustments and base sampling weights.

Household Weights. The initial weight was the sample selection or base weight. This weight was the inverse of the probability of selection of a telephone number. Since all telephone numbers in eligible banks of telephone numbers had the same probability of selection, base weights were all equal. The weight was equal to the total number of telephone numbers (including nonworking and business) in all eligible 100-banks, divided by the number of sample telephone numbers. This weight was 9729.47755.

The second step in producing a household weight was to make an adjustment for non-response at the screener level. These adjustments were done separately for four categories of telephone numbers, defined by whether the average rent for the telephone exchange containing a telephone number was above or below average, and whether the median income for the telephone exchange was above or below average. The non-response adjustment factor consisted of the ratio of all residential telephone numbers (all numbers except for non-working and business) divided by telephone numbers for which a screening interview was completed. Telephone numbers for which we were able to get only an answering machine or for which there was only a ring no answer on repeated calls[4] could not be definitely classified as residential. Based on past studies, we estimated the proportions of such telephone numbers that are residential. We then multiplied these proportions by the quantity of answering machines and ring no answers before adding them into the numerator of the non-response adjustment factor.

The final weighting step was to make an adjustment for households that contained more than one residential telephone number, and thus had more than one chance of selection for the sample. Ordinarily, a weighting adjustment of ½ would be made for households with more than one telephone number and no adjustment would be made for other households. However, this type of adjustment was not made for the 1995 survey. Consequently, to make

[4] Some ring no answer numbers can be determined to be definitely non-working because of a distinctive tri-tone ring, but some non-working numbers do not have this ring.

weighting as consistent as possible for the two surveys, we used a simplified adjustment in which weights for all households were decreased based on the percentage of households that had more than one telephone number.

The product of the base weight, the non-response adjustment factor, and the global multiple phone adjustment factor constituted the final household screener weight.

Person Weights. Interviews were attempted for all persons identified as leave takers or needers, and for a subsample of persons identified as employed-only (i.e., employed but not having taken or needed a leave). As with the household weights, three steps were taken to construct the person weights. The first step was to determine the inverse of this subsampling rate for other employed persons. Note that this factor was applied for all persons identified in the screening interview as other employed, even if they were later determined to be takers or needers. The factor was not applied to persons identified as takers or needers in the screening interview even if they were later determined to be other employed. The subsampling rate for other employed varied for different release groups of telephone numbers, but a single adjustment factor, based on the average subsampling rate, was used for all persons.

The second step in the person weighting was adjusting for non-respondents from households that completed a screening interview. The non-response adjustment factors were applied separately for takers, needers, and other employed, and by gender and age groups. An adjustment factor was calculated as the ratio of the weighted estimate for all persons for whom an interview was attempted, divided by the weighted estimate of persons for whom an interview was completed.

The final step in person weighting was to do a poststratification adjustment with population figures obtained from the Current Population Survey (CPS) conducted by the Census Bureau. Controlling survey estimates to the CPS is commonly done to reduce both sampling error and bias. In many surveys, males, particularly young adults, tend to be missed at higher rates than females, so poststratification factors for young adult males tend to be greater than for other groups. Thus, the application of poststratification factors tends to reduce biases in survey estimates resulting from the higher miss rates.

Control totals from August 2000 CPS for gender crossed by age were examined. We applied the poststratification factors determined for the 2000 survey to the 1995 survey as well. Thus, it was important to define categories such that it would be reasonable to expect that

about the same factors would be applicable for 1995. We found that there were systematic differences between males and females, but only apparently random variation in factors among age groups. We therefore decided to poststratify by gender only. Note that to determine the poststratification factors, it was necessary to include in the denominator of the adjustment factor persons who were not employed (and thus ineligible for the survey) as well as employed persons.

The final person weight was then determined as the product of the final household weight, the employed-person subsampling factor (when applicable), the person non-response adjustment factor, and the poststratification adjustment factor.

2.6.2 Modification of 1995 Weights

The weights used for the 1995 survey included neither a multiple phone adjustment nor poststratification. Since consistent weighting procedures are critical for comparing the 1995 and 2000 surveys, we re-weighted 1995 data using the consistent weighting approach and produced revised estimates using the new weights. It was not possible to do a multiple phone adjustment for only those households with more than one residential telephone, since that information was not available. Instead, we obtained the proportion of households found to have multiple residential telephones in another survey that Westat conducted in 1995, the National Household Education Survey (NHES). This proportion was used to make a global adjustment to all final weights from the 1995 survey.

Next, we assumed that the level of poststratification adjustment factors by gender would be about the same for the 1995 survey as for the 2000 survey. It was not possible to directly determine poststratification adjustment factors for 1995, because no data records were available for unemployed persons. The multiple phone and poststratification factors were applied to the original 1995 survey weights to obtain the modified 1995 weights. These modified weights were used for all comparisons of 1995 and 2000 data.

2.6.3 Preferred Weighting Method

The preferred weighting method differs in only one respect from the consistent weighting method described above. We have not produced weights using this method. However, if future surveys of leave takers are conducted, we strongly recommend that 2000

estimates be made based on weights from the preferred weighting method for comparison to the future survey.

The preferred weighting method uses a different multiple phone adjustment factor. A multiple phone adjustment of ½ is applied to each telephone number for which the household has more than residential telephone number. This multiple phone adjustment is applied instead of the global adjustment described above as the last step in the household weights.

2.6.4 Switching among Takers, Needers, Employed-only, and Not Employed

Persons who were reported in the household interview as leave-takers, needers, or employed-only were eligible for extended interviews. The extended interviews began by confirming that the person really was a taker (or a needer or employed-only, if so designated). Some persons said they were in a different category. Interviews were conducted according to the category determined in the extended interview. It should be noted that some percentage of the persons who were reported as not employed in the screening interview would have said they were takers, needers, or other employed had we confirmed their status in an extended interview. Consequently, a small degree of undercount occurred for each of the groups of interest. If we assumed that the counts from the screening interview were unbiased, or at least closer to the truth than the counts resulting from the extended interview, then a weighting adjustment could be made that would reduce bias. This weighting adjustment would control estimates back to the levels obtained in the screening interview. Such a weighting adjustment might be considered for future surveys.

2.7 Variance Estimation

Appendix B of *Balancing the Needs* provides the standard errors and unweighted sample sizes for each of the estimates published in the report. The standard errors were computed using replicate variance estimation methods. The program used to estimate the standard errors was a Westat-authored program, WESVAR 4.0. The variance estimation procedures accounted for both the complex sample design and the use of weights in the estimation process.

2.8 Definitions of Key Study Measures

The employee survey identified persons who were employed between January 1, 1999 and the time the interview was completed (between 18 and 20 months). This subsection describes how FMLA-related estimates were defined for survey analysis, including the following:

- Coverage and eligibility under FMLA;
- Employee leave status; and
- Leaves taken under FMLA.

2.8.1 Coverage and Eligibility

To determine coverage under the FMLA, the respondent was asked whether his/her employer had at least 50 employees within 75 miles of the worksite (Question C15). If the answer to this question was "yes," the respondent was defined as covered under the Act. This operational definition is not in precise conformance with the requirements of the FMLA. The Act defines an employee as working for a covered employer when the business has at least 50 employees. The definition used on the survey, however, did not include those situations when a worksite did not have 50 employees within a 75-mile radius, but was part of a business that did have employees at other worksites and thus would have met this criteria. This may have pushed the estimate of the number of covered employees downward.

Employees were classified as being eligible under the Act if the respondent reported working at least 1,250 hours in the previous 12 months (Question C17, C18, or C19a) and had worked for the same employer for at least 12 months (Question C16 or C19).

2.8.2 Employee Leave Status: Leave-Takers, Leave-Needers, and Employed Only

To determine whether the leave-taker was covered under the FMLA, the respondent was asked about the size of their employer at the time of the longest leave (Question C15). To determine eligibility, the respondent was asked about the number of hours worked at the time of the longest leave (Question C17, C18, or C19a) and whether they had worked for the same employer for at least 12 months (Question C16 or C19).

A similar definition was followed for leave-needers. These respondents were asked about all of the times they needed but did not take leave during the survey reference period (Question B1). If more than one leave was needed, the respondent was asked detailed questions about their most recent need for leave. To determine whether the person was covered under the FMLA, the respondent was asked about the size of the employer he/she was

working for at the time the most recent leave was needed (Question C15). To determine eligibility under FMLA, the respondent was asked about the number of hours worked for the employer at the time the most recent leave was needed (Question C17, C18, or C19a) and whether they had worked for the same employer for at least 12 months (Question C16 or C19).

The employed-only population was defined as all who were employed but did not take or need leave between January 1, 1999 and the survey. To determine the coverage and eligibility of the "employed-only" group, the respondent was asked about his/her current employer (Question C15). If the respondent was not currently employed, he/she was asked about the employer he/she had worked for the longest period of time since January 1, 1999 (Question C16-C19a).

2.8.3 Leaves Taken Under the FMLA

Leaves were defined as taken under the FMLA if the respondent answered "yes" to whether he/she had heard of the FMLA (Question C3), and reported that the longest leave was taken under the FMLA (Question C6). The number of persons that took leave under the FMLA was estimated by counting those persons who:

- Reported taking leave under the FMLA;
- Were defined as working in a covered worksite; and
- Were classified as being eligible under the law.

2.9 Comparing Data from the 1995 and 2000 Surveys of Employees

As noted above, efforts were made to keep the 1995 and 2000 surveys as comparable as possible. Nevertheless, the comparisons between the surveys may still have been affected by several differences between the two surveys. This section discusses the following issues:

- Differences in the response rates;
- Differences in question wording for key items; and
- Implications of these differences for comparing the 1995 and 2000 surveys.

2.9.1 Differences in Response Rates

The 1995 survey had a combined response rates of 73.1 percent for leave-takers, 75.9 percent for leave-needers, and 70.6 percent for those who were employed but did not take any leave. This is between 11 and 16 percentage points higher than the 2000 survey, depending on the group.[5] The differences in response rates may have led to different patterns of non-response error across the two surveys.[6]

In order to investigate the extent that comparisons were affected by the response rate differentials, several analyses related to the non-response were conducted. Each provides a different perspective on the potential non-response problems in the 2000 survey. The analyses, and the non-response problem they are designed to address, include:

- **Demographic distribution comparison.** This should reveal any differences in the types of respondents captured in each survey.
- **Non-response follow-up survey.** A survey of a sample of non-respondents to the 2000 survey was completed. This provides a direct measure of a sample of persons that the main survey missed.
- **Level-of-effort comparison.** A comparison of differences between respondents to the 2000 survey by the level of effort it took to complete the interview. The assumption in this analysis is that those that required the most effort to interview resemble those that the survey was unable to interview.

In the section that follows, the comparison of the 1995 and 2000 demographic distributions is discussed. In the last part of the section, the results from the other two analyses are described, along with conclusions related to the potential non-response error.

---

[5] It is important to note the difference between non-response error and a low response rate. A low response rate increases the chances that significant non-response error exists. However, non-response error occurs only when the non-respondents actually differ from the respondents along the characteristics that are important to the survey. For example, a recent analysis comparing two surveys which differed by approximately 20 percentage points did not find significant evidence of more non-response error for the survey with the lower response rate (Keeter, et al., 2000). Similarly, analysis of a large survey on welfare reform came to similar conclusions when comparing results among estimates based on response rates that differed by as much as 20 percentage points (Groves, et al., 1997).

[6] For example, it is possible that the 2000 survey missed more employed males in the populations that took leave for family and medical reasons. This would affect the comparison of leave-takers by gender between the 1995 and 2000 surveys. The extent that this may be the case depends on the differentials in response rates by gender between the two surveys.

Comparing Demographic Distributions

Table 2-1 displays the unweighted demographic distributions of the three groups. The unweighted responses are used so that none of the non-response or post-stratification adjustments embedded in the weighting influence the observed distributions. The "all employees" column was standardized to the distribution observed in the 2000 survey.[7]

Table 2-1. Demographic Distributions by Study Groups (Unweighted Sample): 1995 and 2000 Surveys

	Percent of Leave-Takers		Percent of Leave-Needers		Percent of All Employees	
	1995 Survey	2000 Survey	1995 Survey	2000 Survey	1995 Survey	2000 Survey
Gender * xx						
Female	58.7%	62.5%	50.3%	57.6%	48.0%	52.5%
Male	41.3%	37.5%	49.7%	42.4%	52.0%	47.5%
Age * +						
18-24	12.2%	7.1%	9.6%	8.0%	12.9%	13.0%
25-34	31.5%	26.8%	29.4%	25.9%	23.2%	21.2%
35-49	38.5%	41.1%	39.6%	42.8%	42.6%	42.0%
50-64	14.5%	22.6%	20.3%	21.4%	18.2%	21.1%
65+	3.3%	2.4%	1.1%	2.0%	3.2%	2.7%
Race * ++						
White, Non-Hispanic	79.4%	77.2%	71.0%	74.7%	81.1%	78.5%
Black, Non-Hispanic	10.6%	10.2%	17.5%	12.9%	9.5%	9.5%
Hispanic	8.1%	7.0%	8.7%	9.4%	7.5%	6.9%
Asian	(1)	2.5%	(1)	2.0%	(1)	2.7%
Other	1.9%	3.1%	2.7%	1.0%	1.9%	2.4%
Married xx						
Married	72.7%	75.2%	63.1%	68.5%	70.9%	69.8%
Separated/Divorced	15.2%	14.0%	23.5%	20.2%	13.1%	10.2%
Never married	12.0%	10.8%	13.4%	11.3%	16.0%	20.0%

[7] For this purpose, the leave-takers were counted as 16.5 percent of the total employed population, leave-needers as 2.4 percent and employed-only as 81.1 percent. These percentages approximate what was observed for the 2000 distribution across these groups.

Table 2.1 Demographic Distributions by Study Groups (Unweighted Sample): 1995 and 2000 Surveys (continued)

	Percent of Leave-Takers		Percent of Leave-Needers		Percent of All Employees	
	1995 Survey	2000 Survey	1995 Survey	2000 Survey	1995 Survey	2000 Survey
Family with Children						
No	43.3%	42.2%	47.1%	45.8%	55.0%	57.1%
Yes	56.7%	57.8%	52.9%	54.2%	45.0%	42.9%
Education ** + xx						
Less than high school	8.7%	5.0%	9.6%	6.9%	8.4%	5.4%
High school graduate	26.9%	28.0%	27.3%	28.1%	28.8%	29.4%
Some college	31.0%	31.3%	34.2%	26.6%	29.8%	29.2%
College graduate	20.8%	24.0%	16.6%	28.1%	20.9%	25.4%
Graduate school	12.6%	11.6%	12.3%	10.3%	12.1%	10.7%
Income ** ++ xx						
Less than $20,000	19.6%	13.5%	24.3%	15.5%	18.8%	14.8%
$20,000 to less than $30,000	17.3%	12.3%	19.2%	16.6%	16.1%	12.3%
$30,000 to less than $50,000	29.9%	24.7%	28.8%	24.1%	32.2%	25.7%
$50,000 to less than $75,000	19.6%	26.7%	15.2%	27.8%	20.2%	24.6%
$75,000 to less than $100,000	9.0%	12.4%	9.0%	9.1%	8.2%	11.9%
$100,000 or more	4.6%	10.3%	3.4%	6.9%	4.6%	10.6%
Compensation						
Salaried	36.5%	36.4%	29.4%	25.2%	37.0%	36.5%
Hourly	55.0%	53.6%	61.0%	59.9%	51.6%	52.4%
Other	8.5%	10.0%	9.6%	14.8%	11.5%	11.1%

(1) Asian was not a race category in the 1995 survey. Asians are included in "All Others."
* Difference between years for leave-takers is significant at p<.10; ** is significant at p<.05.
\+ Difference between years for leave-needers is significant at p<.10; ++ is significant at p<.05.
x Difference between years for all employees is significant at p<.10; xx is significant at p<.05.
Note: "All employees" was calculated by weighting leave-takers by 16.5%, leave-needers by 2.4% and employed-only by 81.1%. Column percents may not total to 100% due to rounding.

These distributions show relatively small differences across the demographic groups. The largest differences are for gender and income. The 2000 survey found more females and higher income groups. Both of these differences can be partially explained by trends over the five year period between the two surveys. As noted in the introduction of the full study report, women constitute a greater proportion of the workforce in 2000, relative to 1995. Similarly, the higher income of those in 2000 reflects inflation and actual growth in income over this time period.

There are several differences that are not as easily explained by trends in employment. Leave-takers for the 2000 Survey were more likely to be in the older age groups. In addition, the 2000 survey has a lower proportion of persons that are white (for employees 78.5% vs. 81.1%), a lower proportion with less than high school education (5.4% vs. 8.4%), and a higher proportion that are never-married (20% vs. 16%). With the exception of education, none of these differences are particularly large.

Overall, therefore, the 1995 and 2000 survey samples differ in terms of demographics. However, considering shifts in the economy over the time period, these differences are relatively small and do not indicate large differences in non-response error between the 1995 and 2000 employee surveys.

Non-Response Survey

A survey of non-respondents was carried out after the main survey was completed. The survey attempted to interview a subset of non-respondents to the screener, as well as a subset of non-respondents who had been identified as a leave-taker from a completed screener.

For the survey of non-respondents, all households and persons were sent a letter with a $5 incentive using express mail prior to calling to request an interview. In order to conduct this mailing, the sample was restricted to those households for which an address was available.[8] Approximately half of the non-respondents to the main survey had an address and were included in the non-response study.

Screener interviews were attempted with 1738 households who had not completed a screener and for whom the survey team had an address. If a screener was completed with one of these households and a leave-taker was identified, the interviewer proceeded to attempt to complete an interview with that person. A total of 430 screening interviews were completed during the study for a response rate of 24%. From these screeners, an additional 38 interviews were completed with leave-takers.

[8] At the beginning of the survey all of the sampled telephone numbers were sent to a service that attempted to match the number to an address. When a match occurred, the address was used to send different mailings to respondents during the survey period (e.g., pre-notification letter; refusal conversion letters). This same address was used to send the $5 for the survey of nonrespondents.

Interviews were attempted with 116 persons who had been identified as a leave-taker in the main study and for whom the survey team had an address. Of these 116, a total of 28 leave-taker interviews were completed during the non-response follow-up survey for a response rate of 24%.

The response rates imply that the non-response follow-up accounts for slightly over half of the difference between the rates of the 1995 and 2000 Surveys. The screener response rate for the 2000 Survey was 69.3 percent. The 24 percent rate achieved on the non-response follow-up, therefore, increases this by approximately 7.5 percent to 76.8 percent.[9] The 1995 screener response rate was 80.3 percent. A similar calculation shows that the non-response survey for the extended interview increased the 83.9 percent achieved on the main survey approximately 3.9 percent.[10] This raises the response rate after the non-response survey to approximately 87.8 percent. This is about half the difference between the 1995 rate of 91.0 and the 83.9 rate for the 2000 Survey. This is interesting to note, since a primary concern with the 2000 Survey is whether the lower response rate affected comparability with the 1995 Survey. Even though the response rates for the non-response follow-up are low, it is still possible to use these data to assess comparability between 1995 and 2000.

Analysis of the data from the non-response survey concentrated primarily on the results at the screener level. The small number of leave-taker interviews completed did not allow for detailed analysis at that level. The results at the screener level for respondent status, age and gender are shown in Table 2-2. The first column of this table provides the results from the survey of non-respondents. The second column provides the data from the main survey. As can be seen, there is a slight tendency for the non-response survey to have more adults who were not employed during the survey period and were thus ineligible for the survey (24.8% vs. 21.8%). This difference, however, is not highly significant (significant at .15 using a two-tailed test).

[9] The screener non-response rate for 2000 was approximately 31 percent (1 – 69% = 31%). The non-response survey interviewed 24% of this group, resulting in an increase of .31 x .24 = 7.5 percent.

[10] Following the logic for the screener, the non-response rate for the leave-taker interview was 16.1 percent (1-.839 = .161). The non-response survey interviewed 24% of the non-respondents, resulting in an increase of .161 x .24 = 3.9 percent.

Table 2-2. Leave-Taker Status, Age and Gender of Household Members Identified at the Screener by Type of Survey: 2000 Survey

	Non-respondent Survey	Main Survey
Screening Status		
Under 18	22.0%	23.7%
Leave-needer	1.3%	1.6%
Leave-taker	6.7%	9.3%
Employed only	45.2%	43.6%
Adult, not employed	24.7%	21.8%
Age of Household Members		
Less than 18	20.6%	23.2%
18-24	8.7%	9.0%
25-64	54.9%	54.1%
65+	15.8%	13.7%
Gender of Household Members		
Female	53.5%	51.7%
Male	46.5%	48.3%

Note: Based on unweighted data. Column percents may not add to 100% due to rounding.

The number of completed extended interviews from the non-response survey are too few to draw definitive conclusions. For analysis purposes, it is the 28 interviews completed with those that did not respond to the leave-taker interviews that are directly comparable to the main survey interviews. Comparing results of these with those completed during the main survey indicated that the non-respondents were more likely to be male (60.7 percent male vs. 37.2) and have children (78.6% vs. 57.8%). No differences were found with respect to other demographics or several outcome variables (coverage, covered and eligible, heard of FMLA, took leave for own health reason).

Comparisons by Level of Effort

Analyses across different measures of the level of effort to complete interviews were conducted. This type of analysis provides one way of assessing potential non-response error. If those that are harder to reach differ substantially from those that are relatively easy to reach, then one might conclude that there is potential non-response error. This assumes that the persons that are hardest to interview are similar to those that were never interviewed at all.

Level of effort was measured in two ways, including:

- Whether or not the respondent initially refused to do the interview; and
- Number of times the respondent was contacted to complete the interview.

To provide some perspective, approximately two-thirds of the non-response at the screener level was due to refusals and one-third was due to non-contacts. Refusers, therefore, carry greater implications with respect to non-response error from the screener. The opposite is the case at the extended level, where hard to contact respondents constituted two-thirds of the non-response and refusers constituted only one-third.

Screener Responses by Measures of Level of Effort. Table 2-3 provides level-of-effort measures for the screener, including number of contacts[11] and refusal conversions.[12] Screener measures available for analysis consist of the roster information collected during this portion of the interview. This includes classifying each household member into one of five screening status groups: (1) under 18 and not eligible, (2) leave-taker, (3) leave-needer, (4) employed only and (5) adult but not employed. It also includes the age and gender of each individual. As can be seen, significant effects were found for the screening status variable and age of household member for both contacts and refusals.

[11] Interviews requiring 1-3 contacts at the screener level represent approximately 84% of the extended interviews completed. Those requiring 4 or more contacts represent approximately 16% of the extended interviews completed.

[12] Interviews with households that initially cooperated on the screener represent approximately 77% of the extended interviews completed. Those that refused the screener at least once represent 23% of the extended interviews completed.

Table 2-3. Leave-Taker Status, Age and Gender of Household Members Identified at the Screener by Level of Effort to Complete the Screener: 2000 Survey

	Extent of Contact		Refusals	
	Percent with 1-3 Contacts	Percent with 4 or More Contacts	Percent of Initial Cooperators	Percent of Converted Refusers
Screening Status ** ++				
Under 18	23.2%	28.8%	24.7%	22.7%
Leave-needer	9.4%	10.0%	9.6%	9.1%
Leave-taker	1.8%	1.8%	1.8%	1.7%
Employed only	43.8%	41.9%	43.9%	42.3%
Adult, not employed	21.8%	17.5%	20.0%	24.1%
Age of Household Members ** ++				
Less than 18	22.7%	28.0%	24.4%	21.6%
18-24	9.2%	10.9%	9.6%	9.4%
25-64	54.6%	53.8%	54.9%	52.9%
65+	13.5%	7.3%	11.1%	16.1%
Gender of Household Members				
Female	51.9%	52.3%	51.7%	52.4%
Male	48.1%	47.7%	48.2%	47.6%

** Difference between 1-3 and 4 or more contact groups is significant at p<.05.
++ Difference between initial cooperators and refusers is significant at p<.05.
Note: Based on unweighted data.

The significant effects for the contact variable indicates that households with younger persons required more effort in order to complete the screener. This is indicated by the larger percentage of persons enumerated that were under 18 years old (23.3% for easy to contact vs. 28.8% for hard to contact). Also, there were fewer adults who were not employed in the hard to contact group, and about half as many persons enumerated that were 65 or older (13.5% vs. 7.3%). Overall, this would indicate that the screener may have under-estimated the number of people that are employed (and perhaps took leave) over the 18 month period.

The opposite is suggested when looking at the refusers. In this case, those who refused at least once were more likely to not be employed and in the older age groups. This last pattern is consistent with observations from the interviewers who reported that those who were retired tended to refuse to do the interview because they did not believe it was relevant to them. This general pattern indicates that the screener was more likely to over-estimate the number of employed persons.

The differences between the contact and refusal measures are consistent with other research on survey non-response. The reasons an individual is difficult to contact may be different than those for refusing to do an interview. The former is likely to be driven by the person's schedule and other demands placed upon his/her time. Refusing may be a function of this, but it may also be related to the particular topic of the survey. In this particular instance, individuals may have had a tendency to refuse because they perceived the topic as not relevant to them

The implication of these two opposing patterns is not entirely clear. To some degree, the patterns cancel each other out and result in less overall bias. However, since the refusals account for about two-thirds of the screener non-response, one might speculate that the overall tendency is for the screener non-response to dominate a bit more and result in a bias in the direction of missing persons that are not employed and over-estimating the number of leave-takers. However, the size of these differences is quite small. Even though they are statistically significant, the effects are not large and do not imply a large bias in this direction.

Extended Responses by Measures of Level of Effort. The level-of-effort measures were also examined using data from the extended interview. This includes both descriptive (e.g., demographic) and analytic variables. When examining the demographic data for all employees, the following patterns emerge (data not shown):

- **Gender**. Males were more likely to refuse and the hardest to contact.
- **Age**. Young persons (age 18-24) were the hardest to contact.
- **Race/ethnicity**. Whites were more likely to refuse than non-whites. Non-whites were more difficult to contact.
- **Marital Status**. Married persons were more likely to refuse to do the interview and never married persons were more difficult to contact.
- **Children**. Employees with children were harder to contact.
- **Education**. Employees that had less than a college degree were harder to contact.
- **Income**. Employees that refused were more likely to have incomes over $30,000. Employees most difficult to contact were more likely to have incomes less than $20,000.

Similar patterns were evident when examining demographic characteristics for leave-takers. That is, the profile of persons that required the most effort to interview among leave-takers were males, young persons, whites (to refuse), non-whites (to contact), the never

married, those with children. and those with less than a college degree. No differences were found with respect to income.

These patterns can be interpreted in light of those found when comparing the 1995 and 2000 demographic distributions. With respect to gender, these results imply that while some of the changes observed between the 1995 and 2000 are due to trends in employment (i.e., more women being employed in 2000), there also seemed to be higher non-response for males in the 2000 survey. The patterns by age would imply that there may be some under-enumeration of younger workers in the 2000 survey. This is consistent with the trend discussed in Chapter 1 of *Balancing the Needs*, which pointed out that the proportion of younger workers went up between 1995 and 2000, while this is not reflected when comparing the 1995 and 2000 surveys. The results for race/ethnicity imply a slight under-estimate of white employees, as also seemed to be indicated when comparing 1995 and 2000 demographics.

Similar patterns can be observed for the other relationships described above. For example, the tendency for married persons to refuse to do the interview may explain the larger proportion of "never married" respondents in 2000 relative to 1995. None of the differences observed above are particularly large. This is also consistent with the 1995 and 2000 demographic comparisons, which found differences, but very little that would indicate a large relative bias when comparing the 1995 and 2000 surveys.

How does this translate to variables that are of interest to the survey? Tables 2-4 and 2-5 provide data for four different measures relevant to the analyses discussed in the report for all employees and for leave-takers: (1) working in a covered worksite, (2) covered and eligible employee, (3) ever heard of FMLA and (4) whether or not the leave was taken for own serious health reason (leave-takers only). Table C-5 provides these data for level-of-effort measures at the screener level for all employees and for leave-takers. The first two columns of the table show estimates of the analytic variables by the number of contacts (1-3 vs. 4 or more). The third and fourth columns break out these variables by whether the household ever refused to do the screener (initial cooperator vs. refused at least once).

Table 2-4. Analytic Outcomes for Employees and Leave-Takers by Level of Effort to Complete the Screening Interview: 2000 Employee Survey

	Extent of Contact		Refusals	
	Percent with 1-3 Contacts	Percent with 4 or More Contacts	Percent of Initial Cooperators	Percent of Converted Refusers
All Employee				
Covered Worksite **				
Yes	75.6%	82.5%	77.1%	75.1%
No	24.4%	17.5%	22.9%	24.9%
Covered and Eligible **				
Yes	60.9%	66.6%	61.9%	61.2%
No	39.1%	33.4%	38.1%	38.8%
Heard of FMLA?				
Yes	59.1%	58.6%	58.2%	61.9%
No	40.9%	41.4%	41.8%	38.1%
Leave-Takers				
Covered Worksite **				
Yes	77.9%	69.2%	75.8%	77.2%
No	22.1%	30.8%	24.2%	22.8%
Covered and Eligible				
Yes	66.3%	59.5%	63.7%	68.8%
No	33.7%	40.1%	36.3%	31.2%
Heard of FMLA?				
Yes	64.8%	56.6%	62.6%	64.6%
No	35.2%	43.4%	37.4%	35.4%
Leave for Own Health *				
Yes	49.0%	40.4%	48.3%	44.2%
No	51.0%	59.6%	51.7%	55.8%

* Difference between 1-3 and 4 or more contact groups is significant at p<.10.
** Difference between 1-3 and 4 or more contact groups is significant at p<.05.
Note: Based on weighted data.

The data in Table 2-4 do not indicate significant differentials across these four different outcome measures. For employees, two of the six level-of-effort measures are statistically significant. None of the measures related to refusing the screener are significant, while two of the three measures related to contacting the respondent are significant. With respect to the significant effects, those in covered worksites and those that are covered and eligible required more contacts to complete the interview than those that are not covered or covered and eligible. Generally, a similar pattern emerges for the leave-takers. None of the differences for the refusers is statistically significant, while two of the three measures for number

of contacts are significant. In this case, those that are not in covered worksites and those that took leave for a reason other than their own health required more effort to interview.

When examining these level-of-effort measures at the extended level (Table 2-5), a similar pattern emerges. There are no significant differences for the refusers. There are two significant differences for the "contact" variables. Employees that are difficult to contact are more likely not to have heard of the FMLA (60.3% vs. 51.8%). Among leave-takers, those persons requiring the most contacts to complete the interview are less likely to have taken a leave for reasons of their own health (50.3% vs. 30.7%). This is consistent with the pattern observed for the screener non-response as well.

Overall, the level-of-effort measures reflect many of the patterns observed in the demographic and non-response analysis, especially when concentrating on the effects on refusers. The most consistent findings across these analyses is that those most likely to be difficult to interview were males, whites and young persons. Other demographic groups were also suggested as being more likely to be non-respondents, including those married, those with children and those with less than a college degree. When examining several of the analytic variables, very few large differences were observed. There was some indication that individuals taking leave for a non-health related reason were more difficult to contact. This, however, did not carry over to refusers.

Table 2-5. Analytic Outcomes for Employees and Leave-Takers by Level of Effort to Complete the Extended Interview: 2000 Employee Survey

	Extent of Contact		Refusals	
	Percent with 1-4 Contacts	Percent with 5 or More Contacts	Percent of Initial Cooperators	Percent of Converted Refusers
All Employee				
Covered Worksite				
Yes	76.6%	76.8%	76.7%	75.6%
No	23.4%	23.2%	23.3%	24.4%
Covered and Eligible				
Yes	61.3%	64.5%	61.7%	61.9%
No	38.7%	35.5%	38.3%	38.1%
Heard of FMLA? **				
Yes	60.3%	51.8%	59.2%	56.9%
No	39.7%	48.2%	40.8%	43.1%
Leave-Takers				
Covered Worksite				
Yes	77.7%	67.5%	76.1%	76.5%
No	22.3%	32.5%	23.9%	23.5%
Covered and Eligible				
Yes	66.3%	57.8%	64.9%	67.6%
No	33.7%	42.2%	35.1%	32.4%
Heard of FMLA?				
Yes	65.5%	50.2%	63.3%	58.5%
No	34.5%	49.8%	36.7%	41.5%
Leave for Own Health **				
Yes	50.3%	30.7%	47.0%	53.8%
No	49.7%	69.3%	53.0%	46.2%

** Difference between 1-4 and 5 or more contact groups is significant at p<.05.
Note: Based on weighted data.

Summary of Non-response and Comparisons to the 1995 Survey

Comparing the 1995 and 2000 surveys should be done with caution. Response rates differed from 11 to 16 percent between the two surveys. If this response rate indicates that the 2000 survey missed a significant number of persons in important analytic groups who had been included by the 1995 survey, then bias may be introduced when comparing the two surveys. While the analyses summarized above are standard ways to assess potential non-response bias, these methods all rely on untested assumptions about those persons that are never interviewed.

With this caveat in mind, the analyses discussed above provide very little indication that significant non-response bias exists when comparing the 1995 and 2000 employee surveys. The comparison of the demographics between the two surveys does not indicate large differences across groups for either leave-takers or employees. The largest differences were by gender and income. The 2000 survey had a higher proportion of women and a higher proportion of persons in the high income groups. Both of these, however, can be partly explained by trends in the economy. More women have entered the labor force over the last five years and incomes have risen both due to inflation, as well as real growth. There were other differences observed in the demographic distributions, although very few of these were extremely large. These differences found that the 2000 survey had a higher proportion of leave-takers in the older age groups and a higher proportion of employees were non-white, higher educated and never married.

The survey of non-respondents also provided little evidence that there were large differences in the two surveys due to differential non-response. The results from the screener indicated that there may be a small bias in the 2000 survey (relative to the 1995 survey) towards overestimating the number of employed individuals.

Finally, the comparisons by the level of effort were consistent with the demographic and non-response survey. At the screener, there was a tendency for persons that refused to complete the screener to be elderly and not employed. At the extended level, no significant effects were found when comparing the initial cooperators and the refusers. Several significant effects were found by the number of contacts required to complete the interview. One of these effects was consistent across both the screener and extended levels --- the reason for leave reported by leave-takers. This difference indicated that those most difficult to contact were less likely to report taking a leave for their own serious health reason.

2.9.2 Differences in Survey Instruments

The 1995 and 2000 survey instruments are very similar. In most cases, comparable items are worded identically in both surveys. The primary differences between the two surveys are additional questions included in the 2000 survey. For example, the 2000 survey included items about the details for the longest and second longest leaves reported by the respondent. The 1995 survey only asked detailed questions about the longest leave.

There are two exceptions worth noting. First, the screener items used to classify respondents as a leave-taker or leave-needer were modified in 2000.[13] In 1995, the item read:

> *Since January 1, 1994, have you taken leave from work to care for a newborn, newly adopted, or new foster child; for your own serious health condition or the serious health condition of your child, spouse, or parent that lasted more than three days or required an overnight hospital stay?*

In 2000, the item was changed to read:

Since January 1, 1999, have you taken leave from work:

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

One difference between the two surveys is that the 2000 item included the extra condition "for reasons related to your or a family member's pregnancy" (see second bullet above). This change was made to clarify that pregnancy disability leave is covered by the FMLA.

A second difference is the modification of the sentences defining a serious health condition. In 2000, this definition was split out as a separate sentence (see second sentence in the last bullet above). In 1995, this was stated as part of the last phrase of the item. This change was made to clarify that such leave is conditioned upon taking time off for more than three days or for an overnight hospital stay.

The second questionnaire item that was changed asked about the size of the establishment for which the respondent worked. These items were used to classify respondents into a covered and non-covered status with respect to the FMLA. In 1995, this item read:

At the place where you work(ed) (i.e., the site—store, building) would you say there were fewer than 50 permanent employees or 50 or more permanent employees?

---

[13] These items were asked during the screener (referring to each person living in the household) and the extended interview (referring to the respondent).

In 2000, the word "permanent" was dropped from the item. This change was made because the Act does not require that employees be permanent for the establishment to be covered by the Act.

2.9.3 Caveats for Comparing the 1995 and 2000 Survey of Employees

Given the differences in response rates between 1995 and 2000, as well as the above questionnaire changes, some caution should be exercised when estimating change between the two surveys. The areas of concern related to the analyses discussed in this report are described below.

Changes in the Number of Employees

Households that refused to complete the 2000 screener tended to consist of persons that were not employed during the reference period. All other things being equal, this would lead to a higher estimate of the total number of employed persons in the 2000 survey. As a result, comparing the 1995 and 2000 surveys may overestimate the amount of growth in employment over this time period. This is reflected in changes observed in the Current Population Survey (CPS). The CPS, which measures employment over a one week period in each month, estimated growth in employment from 125 million in 1995 to 133 million in 2000 (difference of 8 million). Comparing the two FMLA surveys, which measure employment over an 18-20 month period, the increase is from 127 million to 144 million, a growth of approximately 17 million.

A larger estimate of the growth in the number of employees may lead to overestimates of growth in important subgroups, such as covered employees and covered and eligible employees.

Estimates of Covered Employees

The proportion of persons who were covered by the FMLA increased from 66 percent to 77 percent, based on estimates from the 1995 and 2000 Surveys of Employees. This increase in employee coverage contrasts with estimates from the 1995/2000 Survey of Establishments and data published by the Bureau of Labor Statistics, neither of which found a change in the proportion covered by FMLA.

The change observed between 1995 and 2000 may reflect the more inclusive wording of the item on establishment size in the 2000 Survey of Employees,[14] rather than a true increase in covered employees. A comparison of this estimate to that from the 2000 Survey of Establishments suggests that the estimate from the Employee survey is too high, although it is difficult to estimate the magnitude of any over-estimate because the two sources of information are not directly comparable.

Regardless of the precise accuracy of the number of covered employees estimated from the 2000 Survey of Employees, it is likely that the estimate of change between 1995 and 2000 using these data is an over-estimate.

Covered and Eligible Employees

The estimate of covered and eligible employees are computed by restricting the covered population to those that worked at least 1,250 hours over the previous 12 month period and had worked for the same employer for 12 months or longer.[15] As noted above, the estimate of change for the proportion and number of covered employees may be an over-estimate. Consequently, the estimate of change for the number of covered and eligible employees may also be an overestimate.

Covered and Eligible Leave-Takers

The estimates of covered and eligible leave-takers is computed by restricting the covered leave-takers to those that worked at least 1,250 hours over the previous 12 month period and had worked for the same employer for 12 months or longer. As noted above, the estimate derived from the 2000 Survey of Employees for the proportion and number of covered employees may be an over-estimate. Consequently, the estimate for the number of covered and eligible leave-takers may also be an over-estimate. Comparisons between 1995 and 2000 of these also over-estimate the change.

---

[14] The definition used in the survey is not in precise conformance with the requirements of the FMLA. The Act defines an employer as covered only when the employer has 50 or more employees *for at least 20 workweeks* in the current or preceding calendar year. The survey, however, counted as covered *all* establishments with at least 50 employees within 75 miles of the sampled location. Thus, the survey did *not* count as covered those employers with at least 50 employees beyond 75 miles of the sampled location.

[15] Employee eligibility also requires employees to work at a location where at least 50 employees are employed within 75 miles. This part of the eligibility requirement was actually applied when classifying *establishments* as covered or not covered.

Characteristics of Leave-Takers

The non-response analysis suggested that some of the changes in the characteristics of leave-takers were due to differential non-response patterns across demographic groups. While the analysis did not find a great deal of evidence of a large non-response bias in this regard, it does not totally rule it out either. For example, some of the observed increase in the proportion of female leave-takers may be due to the lower response rate for males in the 2000 survey.

The change in the screening instrument (described above) may have also affected the reasons respondents reported they took leave. As noted in the full study report, the proportion of persons that reported taking leave for reasons of their own health decreased between 1995 and 2000 (61.4% in 1995 vs. 47.2% in 2000; see Table 2.7). One possible explanation for this pattern is the addition of the extra phrase in the screener referring to "...reasons related to your or a family member's pregnancy" (see above discussion). Perhaps some respondents in the 1995 survey reported pregnancy-related leave as being for their "own health."

Analysis of the data from the survey does not seem to indicate that this change was major reason for the decrease in the proportion reporting a leave for a serious health condition. First, while the reasons for leave shifted, the demographic distributions of the samples did not shift significantly (as discussed above). If the screener was the cause of the shift, one would have expected those individuals who tend to take this type of leave (e.g., women, married persons) would constitute a larger portion of the sample than expected. More women were interviewed in 2000. However, most of this increase can be explained by changes in the labor force. Furthermore, the decrease in the number of leaves taken for personal health reasons occurred across almost all demographic groups (see Chapter 2 of *Balancing the Needs*). For example, while females are less likely to take leave for their own health condition, both females and males showed a significant decrease in taking this type of leave between 1995 and 2000. If the screener affected how leave-takers were initially identified, this effect seems to have occurred across all demographic groups.

Second, if this change were responsible for the decrease in leaves taken for the employee's own health, then one would expect the shift to occur primarily in the categories related to pregnancy and maternity. However, as described in Chapter 2 of *Balancing the Needs*, this shift affected the proportion of persons taking leave for other reasons, including the care of ill parents and spouses (e.g., Table 2.7 displayed on page 2-7).

It is also possible that the shift in the distribution of the reasons for leave is at least partly due to re-structuring the screening item that defined a "serious health condition." As noted above (section 1.6.2), the statement defining a serious health condition was changed by separating the definition into a separate sentence to clarify that this type of leave is dependent on specific conditions. These conditions were in the 1995 screener question, but were not separated into a separate sentence as in 2000. One possible scenario is that this change resulted in respondents in 2000 using a more restrictive definition for "serious health condition" (e.g., reporting leave as a result of more severe conditions). If true, then respondents may have reported relatively fewer leaves for their own illnesses in 2000 relative to 1995.

If the restructuring of the item in 2000 had this type of effect, one would have expected some change in the demographic distribution of leave-takers between the two surveys, since certain groups are more likely to take leave for this type of reason (e.g., younger, males, non-married employees). As noted above, the changes observed in the demographics were relatively small. This suggests that the change in the screener question is not clearly related to the changes observed in the data from the two surveys. However, if respondents in 2000 used a more restrictive definition of "serious health condition" when recalling whether or not they had taken leave, one would also expect leave to have been for a longer duration in 2000, relative to the 1995 survey period. As reported in Chapter 2 of the final report, leaves did not appear to have been taken for a longer period in 2000, as compared to 1995. But we recently made this comparison for those reporting their leaves had been for their own health, and found that leaves for this reason reported in 2000 were in fact significantly longer than leaves reported in 1995. Specifically, the median length of leaves for "own health" in 2000 was 14.6 days, whereas in 1995 it was 10 days ($p<.10$). Furthermore, 30.9 percent of leaves for one's own health reported in 2000 were for more than 30 days. In 1995, 24.2 percent of these leaves lasted more than 30 days ($p<.05$). These findings suggest that persons reporting on leaves for their own health used a more restrictive definition of "serious health condition" in 2000, as compared to respondents in 1995.

In summary, it is possible that changing the screening question, and the associated questions on the extended interview, did affect the reasons employees reported for taking leave. Findings from the analyses reported above were mixed, but indicate some evidence for this possibility. As noted in chapter 2 of the final report, there is no clear substantive explanation for why there was a decrease in reports of leave for one's own health between the 1995 and 2000 surveys. Findings reported above with respect to length of leave suggest that the cause of the decrease is, at least in part, a methodological one.

Chapter 3

2000 Survey of Establishments

The 2000 Survey of Establishments, like the 2000 Survey of Employees, was conducted to obtain estimates of the use of leave under the Family and Medical Leave Act since January 1, 1999, and to examine the Act's impact on U.S. businesses. The 2000 Survey of Establishments was designed to represent U.S. private business establishments. It excluded government and quasi-government organizations (e.g., schools, post offices). For purposes of the sample, an establishment was defined as the business located at a particular address or location. Data were collected with respect to this location, even if the employer had other locations. The content of the establishment interview was based largely on the 1995 Survey of Establishments, with the addition of items to explore emergent issues in family and medical leave. Appendix E in *Balancing the Needs* provides the 2000 establishment survey instrument. The human resources director or the person responsible for the company's benefits plan was selected to be the respondent for each establishment. A total of 1,839 interviews were completed. The final weighted response rate for the survey was 65.0%.

This chapter describes the activities Westat undertook to conduct the 2000 Survey of Establishments. These activities included:

- Procedures for data collection;
- Sample design;
- Survey response rate calculations;
- Weighting;
- Variance estimation; and
- Level of effort analyses.

3.1 Data Collection Procedures

In order to collect the data for the 2000 Survey of Establishments, Westat completed the following activities:

- Revised the 1995 questionnaire;
- Developed the computer assisted telephone interviewing (CATI) system;
- Developed a screening instrument;
- Trained interviewers to use the paper and pencil screening questionnaire and the CATI questionnaire;
- Developed informational materials to be mailed out to respondents; and
- Conducted screening interviews and extended interviews with establishment personnel.

These activities are described in greater detail in the sections that follow.

3.1.1 Instrument Revision

The paramount standard used in making instrument revisions was to preserve, as much as possible, the comparability of the 2000 data to that collected in 1995. Therefore, survey design decisions were made with the highest priority placed on retaining the original order and wording of items. Using the comparability standard, instrument revisions took into account four sources of input. First, problems with particular items that had been identified during the 1995 data collection were considered. Second, items that were no longer relevant or appropriate were modified or deleted, while items to address emerging leave issues (e.g., intermittent leave) were added. Third, public comment on the instrument during the OMB clearance process was considered. Finally, the revised instrument was subjected to cognitive testing, in order to assess how well respondents could comprehend the questions and generate appropriate responses. These cognitive interviews were conducted with employers from 6 FMLA-covered establishments and 3 non-covered establishments in May 2000.

Based on all these inputs, four types of changes were made to the 1995 instrument. First, responses to a series of items about leave benefits were reordered (Q6 series; see Appendix E in *Balancing the Needs*). This change was made to help respondents understand the question better by grouping together the health related conditions. Second, the wording was changed on several items (Q19f; Q23aa; Q27; Q28a; Q28f; Q31; Q33a; see Appendix E in *Balancing the Needs*) to clarify the item intent, provide more appropriate examples, or to

eliminate words that were no longer relevant in 2000. For example, an item in the 1995 instrument asked about leave policies that were "pre-existing" to the enactment of the FMLA. The item was changed to read "other leave policies" rather than "pre-existing leave policies." Third, some items were separated (Q9a-c) or combined (Q11b) to simplify the response task. Finally, items that referred to specific dates were updated to current reference periods (e.g., Q17, Q35).

3.1.2 CATI Programming

Programming the instrument began in early May 2000. The CATI system was tested at multiple levels to ensure quality control. These tests included determining whether appropriate skips were followed, measuring the time it took to administer the questionnaire, and checking the management system and the relationships between systems to insure that the whole CATI system worked.

3.1.3 Screening Instrument

Screening activities were conducted to confirm that a sampled establishment still existed and obtained the correct name, address, and phone number for the business. The screening process also identified the person most knowledgeable about employee benefits for that establishment. This person was then recruited for the extended interview. Screening also confirmed that the establishment did in fact have employees and was neither a government nor a quasi-governmental organization.

3.1.4 Interviewer Training

Interviewers were trained to conduct both the paper-and-pencil screening interview and the CATI extended interview. Training included sessions on contacting business establishments, refusal avoidance measures and extensive practice with administering the interview. Westat staff prepared training materials to familiarize interviewers with all aspects of the task. These materials included an interviewer's manual that described the background and purpose of the project, provided answers to commonly asked questions, and presented question by question specifications along with the actual screening instrument.

A total of 17 interviewers were trained for the screening portion of the study, while 13 were trained to conduct the extended interview. Training sessions, which lasted 8 hours for the screening interview and 16 hours for extended interview, included:

- A lengthy introduction to the project, including background on the Family and Medical Leave Act and the purpose and goals of the 2000 Survey of Establishments.
- Interactive lectures, presenting the basic objectives of the questionnaire and discussing possible contact scenarios.
- Extensive discussions of procedures for contacting business establishments, including best times to call, dealing with disconnected telephone numbers and language problems, responding to reluctant respondents, and processing establishments that have gone out of business. These included discussions of interviewers' previous successes and failures when making such calls.
- Dyad role-playing, where one interviewer played the role of the interviewer while another acted as the respondent. The trainers and other supervisory personnel monitored interviewers and problems were discussed at the conclusion of each role-playing session. All interviewers successfully completed role-playing before they were allowed to begin interviewing.
- Refusal avoidance, which discussed how to handle respondents who were not inclined to participate.

Training for the screening interview took place in late May 2000, while the extended interview training occurred on July 12-13, 2000.

3.1.5 Pre-Notification Materials

Following the screening interview and prior to the commencement of the main interview, the person identified in the screening interview as the most knowledgeable person about employee benefits at the establishment was sent a package with information about the study. This package of materials included a letter from U.S. Secretary of Labor Alexis M. Herman, background material on the study, and an explanation of data items that would be requested (e.g., number of leavetakers). This explanation encouraged respondents to review their leave records before the interview, in order to obtain more precise numbers. Pre-notification materials are displayed in Appendix x.

3.1.6 Interviewing

The screening portion of the study began on June 6, 2000, and continued for approximately 8 weeks. If an establishment was eligible, pre-notification materials were sent to

the identified most-knowledgeable person prior to calling that person to conduct the extended interview.

Project staff reviewed all problem cases that were identified during screening. Cases with discrepancies were returned to the telephone center for a clarification call. In order to maximize the screener response rate, screening interviewers attempted to convert all cases that originally refused.

The interviewing for the data collection portion of the FMLA Employer study began on July 13, 2000, and lasted for approximately 10 weeks. A total of 1,839 interviews were completed.

3.2 Sample Design

The sample design for the 2000 Survey of Establishments mirrored that used for the 1995 survey, in order to keep the two survey results as comparable as possible. This section discusses the following sample design issues:

- Sample universe;
- Sampling frame;
- Sampling strata; and
- Sample size.

3.2.1 Sample Universe

As in 1995, the 2000 sample was planned to represent all private business establishments except those of self-employed individuals without other employees. This excluded government and quasi-government units (e.g., government offices, public educational institutions, post offices). This sample universe differed from that used for the 2000 Survey of Employees, which included both private and public employees.

3.2.2 Sampling Frame

Westat obtained the Dun and Bradstreet's Dun's Market Identifiers (DMI) to use as the sampling frame. The DMI is considered to be the most comprehensive commercially available list of U.S. businesses, and had been used for the 1995 Survey of Establishments.

Most out-of-scope establishments could be identified using the information contained within the DMI. These out-of-scope establishments were deleted from the frame prior to drawing the final sample.

3.2.3 Sampling Strata

Once out-of-scope establishments were deleted from the frame, it was stratified by size and industry. Stratification by establishment size was done for two reasons. First, stratification allowed us to obtain a more efficient sample. The distribution of establishments by size is highly skewed, with many more small establishments than large establishments. If a simple random sample had been drawn, the number of large establishments selected may have been too small to produce stable employee-based estimates (e.g., number of leave-takers). Stratifying the frame by size and sampling within each size class thus produced a more efficient sample. Second, separate analyses were planned for establishments with 25 to 49 employees and 50 to 99 employees, so the size classes which included these establishments were also over-sampled. All estimates in the final report were adjusted for the over-sampling by weighting establishments by their probability of selection. Weighting is described in greater detail in Section 3.3.

Stratification of the sample frame by industry group was done in order to insure an equal distribution of the sample across these groups. Five industry groups were formed by grouping establishments using their Standard Industrial Classification (SIC) codes as given below:

(1) SIC group I: Agriculture, forestry, and fishery (SICs 01-09); Mining (SICs 10-14; Construction (SICs 15-17);

(2) SIC group II: Manufacturing (SICs 20-39);

(3) SIC group III: Transportation, communication, and utilities (SICs 40-49 except SIC 43, US postal service); Wholesale (SICs 50-51); Finance, insurance, and real estate (SICs 60-67);

(4) SIC group IV: Retail (SICs 52-59);

(5) SIC group V: Services (SICs 70-89 except public units from SIC 82).

In the 1995 survey, the industry stratification was implicit in the sense that establishments were selected from the list ordered by SIC within size classes. It was made explicit for the 2000 survey to follow more closely the stratified simple random sampling design.

The sampling strata thus were defined by cross-classification of the size classes and SIC groups.

3.2.4 Sample Size

The total target sample size was set at 1,800. The target sample size was determined based on the eligibility and response rates from the 1995 survey and calculations. The 1995 Survey obtained an eligibility proportion of 90 percent and a response rate of 70 percent. This resulted in the sample size of 3,000.

The overall sample size was allocated to the employment size classes using the $\sqrt{X}$ -proportional allocation method, where X is the aggregate number of employees of a given size class. This allocation allows sampling large size establishments more heavily (in terms of the sampling fraction) but not as extreme as with the X-proportional allocation.

The sample sizes for the 20-49 and 50-99 size classes were further increased over the normally allocated sample size to meet the special need for the size classes. The sample sizes of the SIC groups within each employment size class were then determined using the simple proportional allocation method, which allocates proportionally to the numbers of establishments in the SIC groups within the size class.

3.3 Response Rates

The weighted response rate for the 2000 Survey of Establishments was 65.2 percent. This is 8.2 percent lower than the 73.2 percent achieved in 1995. The paragraphs that follow describe how the response rates were calculated for both the screening interview and the extended interview.

3.3.1 Screening Response Rates

In order to determine the screener response rate, several decisions and computations were made, including the following:

- Classify result codes into categories;
- Determine unit eligibility;
- Calculate unweighted and weighted screener response rates; and

- Make non-response adjustments to each size-industry stratum cell.

Result Code Classification

Screening produced a variety of results, such as complete, ring no-answer, ineligible for various reasons, or non-working number. The screener result codes were classified into four categories: respondents; non-respondents with known eligibility; out of scope units; and units with unknown eligibility. These result codes, and the number of units so screened, are displayed in Table 3-1.

Table 3-1. Definition of Screener Response Status Categories

Response Status Category	Screener Result Code	Number of Units
1. Respondent	C: Complete	2,660
2. Nonrespondent with known eligibility	RB: Final refusal	33
	LP: Language problem	1
	NA: Ring-no-answer	12
	NM: Answering machine	17
	MC: Maximum calls	0
3. Out of scope	I: Ineligible	11
	S1: Out of business	49
	S2: Government	157
	OD: Duplicate	2
	OO: Other	2
4. Unit with unknown eligibility	NL: Non-locatable	18
	NW: Non-working	38
Total		**3,000**

Determining Unit Eligibility

Unit eligibility is determined based on the results of screening efforts. Treatment of units with unknown eligibility required an assumption about their actual eligibility. They could have been considered as all eligible, all ineligible, or some combination of eligible and ineligible. Based on Westat's experience with other telephone surveys of businesses, 10 percent of units with unknown eligibility were assumed to be eligible. This assumption was used in making the screener non-response adjustments, described below.

It is arguable whether the eligibility of the ring no-answer (NA) cases is known or not. Strictly speaking, their eligibility is unknown. However, following Westat's usual practice in conducting telephone surveys, these cases were assumed to be eligible for the purpose of screener non-response rate calculation (and weight adjustment). The effect of this assumption

on the response rate was small, as will be described later, but produced a slightly conservative response rate.

Calculation of Screener Response Rates

Based on the assumptions discussed above, two response rates were calculated using the formulae given below:

1. Unweighted response rate: $\mathrm{UR} = \dfrac{R_1}{R_1 + R_2 + 0.1R_4} = \dfrac{2660}{2660 + 63 + 0.1 \times 56} = .975$;

2. Weighted response rate: $\mathrm{WR} = \dfrac{R_1 \sum_{i=1}^{R_1} w_i}{R_1 \sum_{i=1}^{R_1} w_i + R_2 \sum_{i=1}^{R_2} w_i + 0.1R_4 \sum_{i=1}^{R_4} w_i} = 0.937$,

where R_k, $k = 1, 2, 3, \text{or } 4$, is the number of screener sample units in response status category k and w_i is the screener base weight, which is defined as the inverse of selection probability of unit i.

The weighted rate is lower than the unweighted, which indicates that small sized units (with large screener base weight) did not respond more often than the large sized units. To see more clearly differential response rates at different levels of aggregation, unweighted and weighted response rates, calculated by industry group, are displayed in Tables 3-2 and 3-3. As these tables illustrate, Industry Group I had substantially lower response rates. To handle the differential response rates, Westat calculated a screener non-response weight adjustment separately for each stratum cell.

Table 3-2. Unweighted Screener Response Rate (%) by Industry Group

Industry Group[1]					All
I	II	III	VI	V	
90.6	99.3	98.2	98.0	97.5	97.5

Note 1: Definition of Industry Groups

(1) SIC group I: Agriculture, forestry, and fishery (SICs 01-09); Mining (SICs 10-14; Construction (SICs 15-17);

(2) SIC group II: Manufacturing (SICs 20-39);

(3) SIC group III: Transportation, communication, and utilities (SICs 40-49 except SIC 43); Wholesale (SICs 50-51); Finance, insurance, and real estate (SICs 60-67);

(4) SIC group IV: Retail (SICs 52-59);

(5) SIC group V: Services (SICs 70-89 except public units from SIC 82).

Table 3-3. Weighted Screener Response Rate (%) by Industry Group

Industry Group					All Establishments
I	II	III	VI	V	
77.9	99.7	96.9	93.7	95.7	93.7

Both weighted and unweighted response rates provide important information for assessing survey results. The unweighted response rate shows the effectiveness of the screener interviewing process if all the units are treated in the same manner regardless of the size. The weighted rate gives an estimate of what percentage of the survey universe is represented by the respondents. Nonresponse weight adjustment is a process to force the respondents to represent the whole universe. In the process, differential nonresponse rates should be appropriately taken care of to avoid bias in survey estimates that could result otherwise. The stratum cell level adjustment is believed to be such a procedure.

Screener Non-Response Adjustment

As described above, the weighted screener response rate for this study was 93.7%. To perform screener nonresponse weight adjustment, the nonresponse cells were first created. The cells were defined by five industry codes crossed by employment size categories used to define the sample design strata. The screener result codes were then classified into the four categories listed above. For purposes of this study, as was done with the 1995 FMLA study, the no answer and answering machine cases were deemed eligible for the purpose of screener non-response adjustment. The cases with unknown eligibility (non-locatable and non-working) again as in 1995 were treated the same, where 10% were assumed to be eligible.

3.3.2 Extended Interview Response Rate

In order to determine the extended interview response rate, we:

- Classified result codes; and
- Calculated unweighted and weighted extended interview response rates.

Table 3-4 shows the frequency distribution of combined dispositions, grouped into the four categories discussed above. Even the most careful screening cannot eliminate all ineligible units, and a considerable number of units were identified as ineligible at the extended interview. In addition to units identified as ineligible during screening, 38 units were identified as

non-locatable during the extended interview. These units were classified as eligible non-respondents, since they had been eligible at the screener. Consistent with the way in which units of unknown eligibility were handled in the screener response rate computation, 10 percent of these units were assumed to be eligible.

Table 3-4. Frequency Distribution by Disposition Codes of the Screener and the Main Interview

Response Category	Screener Disposition	Main Interview Disposition	Frequency	Category Total
1 – Respondent	C	CD	24	1,839
	C	CO	1,789	
	C	CR	26	
2 – Non-respondent with known eligibility	C	LP	1	780
	C	MC	188	
	C	MR	16	
	C	NL	38	
	C	R3	1	
	C	RB	414	
	C	RM	59	
	LP	-	1	
	NA	-	12	
	NM	-	17	
	RB	-	33	
3 – Out of scope non-respondent	C	I1	18	326
	C	IE	82	
	C	OD	2	
	C	OO	2	
	I	-	11	
	NL	OO	1	
	OD	OD	2	
	OO	-	2	
	S1	-	49	
	S2	OG	1	
	S2	-	156	
4 – Non-respondent with unknown eligibility	NL	-	17	55
	NW	-	38	
Total			**3,000**	**3,000**

The unweighted overall response rate was computed as:

$$\text{UR} = \frac{R_1}{R_1 + R_2 + 0.1R_4} = \frac{1839}{1839 + 780 + 0.1 \times 55} = 0.701;$$

the weighted response rate (using the screener base weight) was calculated as:

$$\text{WR} = \frac{R_1 \sum_{i=1}^{R_1} w_i}{R_1 \sum_{i=1}^{R_1} w_i + R_2 \sum_{i=1}^{R_2} w_i + 0.1R_4 \sum_{i=1}^{R_4} w_i} = 0.650,$$

where R_k, $k = 1, 2, 3, \text{or } 4$, is the number of sample units in response Category *k* and w_i is the screener base weight, which is defined as the inverse of selection probability of unit *i*.

Based on the set of screener respondents, the resulting main interview response rate was 71.9 percent.

3.4 Weighting and Estimation Procedures

Weighting the 2000 Survey of Establishments data entailed three steps:

- Assigning sampling base weights equal to the reciprocal of the probabilities of selection;
- Adjusting the base weights to compensate for nonresponse in the screener and extended interviews; and
- Post-stratifying the weights so that weighted counts from the survey matched Bureau of Labor Statistics (BLS) establishment counts within broad size and industry groups.

These steps are described in the paragraphs that follow.

3.4.1 Assigning Base Weights

Since the sample design was a stratified simple random sampling design, each design stratum had the same sampling base weight (for all the establishments in the stratum) that was defined by the stratum population size divided by the stratum sample size.

3.4.2 Adjusting Base Weights

The base weights of the screener respondents were inflated by the inverse of the response rate of the eligible establishments within each design stratum. These screener nonresponse adjusted weights were further adjusted for main interview non-response, again within each design stratum. All non-respondents were assumed eligible, since all establishments in the main interview had been screened as eligible.

As commonly seen in an establishment survey, some outliers were very influential in estimation of totals by having a large employment size and a large sampling weight due to out-dated or erroneous size measure. The procedure used to handle such outliers treated 31 outliers by modifying their weights as if they had been selected from the correct size classes. This kind of outlier treatment causes some bias in an estimate but it also reduces a large amount of the variance of the estimate so that the mean square error of the estimate is much smaller than that which would be obtained without outlier treatment.

3.4.3 Post-Stratifying Weights

Outlier-treated weights were further adjusted by poststratifcation ratio adjustment using BLS establishment counts for the first quarter of 1999. The BLS counts were first modified to correct coverage differences between the BLS program from which the BLS counts were obtained and the 2000 Survey of Establishments. For farm establishments and railroad transportation establishments, which the BLS program did not cover but the FMLA did, the DMI counts were used. The number of public educational establishments (a type of quasi-government establishments), which the FMLA did not cover but the BLS counts included, was estimated using DMI counts and the estimates based on the number of such establishments identified from the sample. For U.S. postal establishments, which are another type of quasi-government establishment not covered by the FMLA, a separate BLS count was available so that the BLS count, which included the sector, was modified.

The poststrata were defined as broad cross-classifications of 4 aggregated size classes (1-24, 25-49, 50-249, and 250+) and 3 combined industry groups. The poststratification ratio adjustment forced the sum of weights agree to the BLS total for each of the 12 poststrata. This final weighting process produced the final weights, which were used in estimation.

3.5 Variance Estimation

Appendix B of *Balancing the Needs* displays sandard errors and unweighted sample sizes for all estimates published in the report. Three types of estimators were used for three types of population values of interest, total, ratio, and proportion. An estimate of a population total (e.g., the total number of leave takers) was obtained as the weighted sum of the value of the variable of interest (e.g., establishment's reported number of leave takers). An estimate of a population ratio was defined as the ratio of two estimated totals. A population proportion of establishments with a certain attribute was estimated by the ratio of an estimated total number of establishments with the attribute to the sum of the final weights of all the sample establishments.

Using WESVAR 4.0, the sampling error of an estimate was estimated by the jackknife replication method, for which 40 replicates were created in such a way that stable sampling error estimates can be obtained for subpopulation estimates used in the analysis. The variance of an estimate was estimated by the sum of squared deviations of replicate estimates from the full sample estimate. The sampling error was then estimated by taking the square root of the variance estimate.

3.6 Level of Effort Analyses

This section examines the quality of data obtained in the 2000 Survey of Establishments. During data collection, interviewers encountered respondents who repeatedly put off being interviewed by avoiding phone calls or breaking appointments. These respondents required the project to make many callbacks before they completed the interview or gave a final refusal, and some were never interviewed by the end of the field period.

3.6.1 Analytic Approach

The analysis used the number of calls to complete as a measure of level of effort. We examined how level of effort was associated with establishment characteristics, and how level of effort affected several measures of data quality. We also explored how level of effort affected key study measures, including FMLA leave use, administrative burden, and impact of FMLA. Table 3-5 displays the measures used in this analysis.

Table 3-5. Measures Used in Level of Effort Analysis

Level of Effort	**Calls to Complete:** Number of calls to complete extended interview (weighted mean=2.81).
Establishment Characteristics	**SIC:** Standard Industrial Classification. **Size:** Number of employees at establishment. **FMLA Coverage Status:** Whether the establishment was covered by the Family and Medical Leave Act. **Workforce Unionization:** Proportion of employees that belong to a union. **Workforce Gender:** Proportion of employees that are female.
Respondent Characteristics	**Tenure:** Number of years respondent has worked for the establishment.
Data Quality	**DK Response Count:** Number of items that the respondent could not answer (weighted mean=3.13 items). **DK Response on FMLA Leave Use Item:** The proportion of respondents who gave a DK response on item asking the number of employees who had used FMLA leave. **Record Check:** Whether respondent had checked business records for information needed in the survey.
Establishment Outcomes	**Use of FMLA Leave:** The number of employees per 100 that used FMLA leave, collapsed into three categories: no use; low use, defined as more than 0% and less than 5%; and high use, defined as more than 5%. This measure only includes respondents who gave a valid response on this item. **Administrative Burden:** An index constructed from 8 items assessing the level of burden for administering the provisions of the Family and Medical Leave Act. Scores range from 1 to 5; lower scores indicate lower burden (mean score=2.3). **Business Impact:** An index constructed from 8 items assessing the impact of the Family and Medical Leave Act on establishment outcomes (e.g., business productivity; employee morale). Index scores range from –1, indicating more negative impact, to +1, indicating more positive impact (mean score=-0.01).

To explore the relationship between establishment characteristics and level of effort, we compared the mean number of extended calls for establishments of different size, SIC, workforce gender, and workforce unionization. We used the same approach to look at respondent characteristics.

Next, to explore the effects of level of effort on data quality, we combined respondents into level of effort groups based on the number of calls required to complete their interview and compared groups on three measures of data quality (overall number of DK responses; DK response on FMLA leave use question; and records consultation). Level of effort groups were defined as follows:

- **Low effort group**—Respondents who required only one call to complete the interview;
- **Moderate effort group**—Respondents requiring two to four calls to complete; and
- **High effort group**—Respondents that required five or more calls to complete.

Finally, to assess whether level of effort is related to some of the important study measures, we compared effort groups on reported level of FMLA leave use, mean administrative burden and business impact. We also calculated the estimates we would have obtained for these measures depending on how many callbacks were made to complete the interview.

3.6.2 Establishment Characteristics Associated with Level of Effort

Establishment Size. Generally speaking, larger establishments required more calls to complete, as Table 3-6 illustrates. Establishments with ten or fewer employees and those with 11 to 25 employees required a little less than four calls on average to complete the extended interview. This was significantly fewer calls compared to establishments with 50 or more employees, which required between 4.9 and 8.1 calls, on average. Establishments in the 50 to 99 size group required significantly fewer calls than did establishments in the over-1000 group.

Table 3-6. Establishment Size and Level of Effort
(weighted data)

Number of Employees	Mean Calls to Complete	SE	Significant Differences[1]
1-10 (1)	3.66	0.386	4,5,6,7,8
11-25 (2)	3.67	0.399	4,5,6,7,8
26-49 (3)	5.48	2.068	
50-99 (4)	4.88	0.456	8
100-249 (5)	5.48	0.655	8
250-499 (6)	5.82	0.802	
500-999 (7)	8.12	2.242	
1000+ (8)	7.37	0.775	

[1] Numbers indicate significant differences between group pairs.

Establishment SIC. Establishment SIC did not have an effect on the number of calls required to complete the extended interview. These results are displayed in Table 3-7.

Table 3-7. Establishment Standard Industrial Classification and Level of Effort
(weighted data)

SIC[1]	Mean Calls to Complete	SE	Significant Differences[1]
Manufacturing	3.04	0.662	
Retail	3.91	0.662	
Service	3.22	0.365	4
All others	4.37	0.576	

[1] Numbers indicate significant differences between group pairs.

FMLA Coverage Status. Non-covered establishments required significantly fewer calls to complete. Mean number of calls for covered and non-covered establishments are displayed in Table 3-8.

Table 3-8. FMLA Coverage Status and Level of Effort
(weighted data)

FMLA Coverage Status[1]	Mean Calls to Complete	SE
Covered	4.85	.374
Not covered	3.60	.344

[1] Groups are significantly different ($p<.05$).

Since being covered by the FMLA is partly determined by size, we controlled for establishment size and SIC using multiple regression. Results show that, on average and holding constant size and SIC, covered establishments required about 1.2 more calls to complete than did non-covered establishments. These results also indicate that respondents from manufacturing establishments required significantly fewer calls, compared to respondents working for Other-classified establishments. This is interesting considering that size is positively associated with level of effort, and manufacturing establishments tend to be larger. These results are displayed in Table 3-9.

Table 3-9. Establishment Characteristics and Level of Effort

Dependent: Calls to Complete				
Parameter	**Parameter Estimate**	**Standard Error of Estimate**	**Test for H0: Parameter=0**	**Prob>|T|**
INTERCEPT	4.11	0.597	6.89	0
Number of employees	0	0.001	2.97	0.0050
Covered by FMLA	1.24	0.480	2.58	0.0135
SIC=Manufacturing	-2.04	0.625	-3.27	0.0022
SIC=Retail	-0.38	0.889	-0.43	0.6672
SIC=Service	-1.04	0.601	-1.72	0.0924
R_SQUARE	0.02			

Workforce Unionization. Establishments with either a high or low proportion of unionized workers required more calls to complete the interview, on average, compared to establishments with a medium proportion of unionized workers. However, the only significant difference was between establishments with less than 25 percent unionized workforce and establishments with 25 to 49 percent unionized workforce. These results are displayed in Table 3-10.

Table 3-10. Workforce Unionization and Level of Effort (weighted data)

Proportion of Workforce Unionized	Mean Calls to Complete	SE	Significant Differences[1]
Less than 25% unionized	3.82	0.331	2
25-49% unionized	1.77	0.979	
50-74% unionized	2.37	0.925	
75-100% unionized	3.91	0.724	

[1] Numbers indicate significant differences between group pairs.

Workforce Gender. The proportion of an establishment's workforce that is female was not significantly related to the number of calls needed to complete an interview. Table 3-11 displays these results.

Table 3-11. Workforce Gender and Level of Effort (weighted data)

Proportion of Workforce Female[1]	Mean Calls to Complete	SE
Less than 25% female	3.83	0.542
25-49% female	3.73	0.991
50-74% female	4.00	0.670
75-100% female	3.29	0.519

[1] No significant differences between groups.

Respondent Job Tenure. As Table 3-12 illustrates, the length of time that a respondent had held her current position was not related to the number of calls needed to complete the extended interview.

Table 3-12. Respondent Job Tenure and Level of Effort (weighted data)

Years respondent has held current position[1]	Mean Calls to Complete	SE	Significant Differences[1]
1 year	4.58	0.693	3
2-4 years	4.54	0.735	
5 or more years	3.22	0.399	

[1] Numbers indicate significant differences between group pairs.

Level of Effort Effects on Data Quality

In this section, we assess the effects of level of effort on several indicators of data quality, which include:

- Number of items for which the respondent could not provide an answer;
- Whether the respondent could provide the number of employees who had used FMLA leave;
- Whether the respondent consulted records prior to the interview; and
- Estimates of critical study outcomes such as administrative burden and business impact.

Unanswered Items. One measure of data quality is the number of items for which the respondent could not provide an answer (DK responses). We combined respondents into level of effort groups based on the number of calls required to complete their interview, as described above. We then computed the mean number of DK responses for each reluctance group. As Table 3-13 illustrates, respondents in the no reluctance group had on average about 2 fewer DK responses, compared to both of the other two groups.

Table 3-13. DK Responses Among Level of Effort Groups
(weighted data)

Level of Effort Group[1]	Mean Number of DK Responses	SE	Significant Differences[1]
1 - Low effort	2.04	0.322	2,3
2 - Moderate effort	3.88	0.523	
3 - High effort	3.97	0.704	

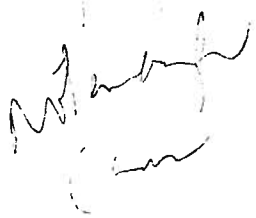
[1] Numbers indicate significant differences between group pairs.

Reporting Leave Use. While the index of DK responses gives an idea of the overall trouble respondents had with answering questions, one item was particularly important: number of employees who had used FMLA leave had an important influence, and almost 50 percent of all respondents could not provide an answer to this item. We compared the proportion of respondents in each level of effort group that could not answer the leave use item. As Table 3-14 illustrates, a significantly higher proportion of the high effort group could not give a response on the leave use item, compared to the proportion in the low effort group.

Table 3-14. Reporting FMLA Leave Use and Level of Effort
(weighted data)

	Low Effort Group		Moderate Effort Group		High Effort Group	
	%	SE	%	SE	%	SE
Percent of respondents that did not know FMLA leave use	2.97	1.225	21.76	16.491	19.1[1]	3.765

[1] Significantly different from the no reluctance group.



Consulting Records. Another measure of data quality is whether the respondent consulted business records prior to being interviewed. For this survey, every respondent was sent a pre-notification letter that explained the purpose of the study and requested that the respondent consult business records to obtain the number of employees who had used unpaid family or medical leave between 1999 and the interview date. Theoretically, respondents who consulted records could give more accurate counts, thus producing better estimates.

Level of effort groups did not differ in their likelihood of consulting records to get information for the survey. In fact, regardless of the number of calls required to complete the interview, the majority of respondents did not consult records. These results are displayed in Table 3-15.

Table 3-15. Consulting Records and Level of Effort (weighted data)

	Low Effort Group[1]		Moderate Effort Group[1]		High Effort Group[1]	
	%	SE	%	SE	%	SE
Respondent consulted records	33.4	4.319	31.36	5.024	38.7	6.765
Respondent did not consult records	66.6	4.319	68.64	5.024	61.3	6.765

[1] No significant differences between groups.

Key Measures of Study Interest: Leave Use, Administrative Burden, and Business Impact. The analyses reported in Section 3.7.1 suggest that a high level of effort is required to complete interviews with large, non-manufacturing establishments. Unfortunately for this particular survey, respondents at establishments covered by the FMLA also required a higher level of effort. If level of effort affected data quality, then the survey may have produced biased estimates on the measures of interest, which included the amount of leave taken under the FMLA, the burden on businesses of administering the FMLA, and the Act's impact on business outcomes such as productivity and profitability.

To assess the impact of level of effort on these key measures, we compared the three level of effort groups on each of these measures. For leave use, we compared effort groups in terms of the proportion in each of four leave use categories: no use, low use, and high

use. The proportion of respondents with no, low, and high use of FMLA leave by employees did not differ significantly among the level of effort groups (Table 3-16).

Table 3-16. Reported Leave Use and Level of Effort
(weighted data)

Level of FMLA Leave Used by Employees[1]	Low Effort Group		Moderate Effort Group		High Effort Group	
	%	SE	%	SE	%	SE
No leave use	65.26	8.987	52.73	11.028	63.56	7.950
Low leave use	7.07	1.518	12.54	2.959	10.38	2.773
High leave use	27.67	8.533	34.73	11.410	26.07	7.459

[1] No significant differences between groups.

For the outcomes of administrative burden and business impact of FMLA, we compared effort group mean scores on the two indexes. In terms of mean reported administrative burden and business impact, displayed in Table 3-17, we did not find any significant level of effort group differences.

Table 3-17. The Impact Of Level of Effort On Administrative Burden and Business Impact
(weighted data)

Level of Effort Group[1]	Mean Administrative Burden	SE	Mean Business Impact	SE
Low Effort	2.52	0.187	-0.01	0.021
Moderate Effort	2.19	0.129	-0.06	0.072
High Effort	2.27	0.147	0.02	0.011

[1] No significant differences between groups.

Discussion

These analyses show that, in the 2000 Survey of Establishments, level of effort was associated with some establishment characteristics (size, industry, and unionization of workforce) and respondent job tenure. Compared to their easier-to-interview counterparts, respondents requiring more calls to complete gave more DK responses and were less likely to know how much FMLA leave had been used. However, they did not give more negative assessments of FMLA's impact on their businesses.

The results of these analyses have several implications for establishment estimates reported in *Balancing the Needs*. First, level of effort was related to some key establishment characteristics, particularly establishment size, but also SIC. The relationship with size may be partly a result of needing to find the appropriate respondent in larger establishments. This might occur, for example, when the first respondent in the establishment refers the interviewer to someone else in the central office. Although the number of calls to complete used in this analysis did not include initial screening calls made to get a respondent's name, some of the calls included in our measure likely were made in order to find the appropriate respondent. The relationship between size and SIC consequently had a potential for biasing estimates.

Second, level of effort had a negative effect on two related measures of data quality. These included the number of all items for which the respondent could not provide an answer, and the proportion of respondents who did not know how much family and medical leave had been taken by their employees. However, DK responses on individual items will have produced biased estimates only if those who responded to an item were systematically different in some way from those who did not respond to it.

Third, level of effort did not appear to be related to whether respondents checked establishment records for information needed in the survey. In fact, a substantial majority of all respondents said they had not checked records, even though the advance letter had requested checking establishment records for family and medical leave use. So while the results suggest that estimates of leave use may be not be as precise as desired for a study such as this, at least the imprecision does not seem to be systematically influenced by level of effort.

Finally, level of effort did not appear to be directly related to other estimates of key outcome measures of study interest, including perceptions of administrative burden and business impact of complying with the Family and Medical Leave Act. It could be that level of effort could be indirectly related to these outcome measures. Alternatively, level of effort may be only related to specific items within the burden or impact indexes. Further analyses would be needed to explore these possibilities.

References

Groves, R and Wissoker, D.W. (1999). "Early Non-response Studies of the 1997 National Survey of America's Families," Report #7, Assessing the New Federalism. Urban Institute: Washington, DC.

Keeter, S., Miller, C., Kohut, A., Groves, R.M., and Presser, S. (2000). "Consequences of Reducing Non-response in a National Telephone Survey." *Public Opinion Quarterly*, 64:125-148.

Cantor, D., Waldfogel, J., Kerwin, J., McKinley Wright, M., Levin, K., Rauch, J., Hagerty, T., and Stapleton Kudela, M., 2001. *Balancing the Needs of Families and Employers: Family and Medical Leave Surveys.* Rockville, MD: Westat.