NSAF Methodology

1997 NSAF Snapshot Survey Weights

Report No. 3

Prepared by:

J. Michael Brick
Gary Shapiro
Ismael Flores-Cervantes
David Ferraro
Teresa Strickler

Assessing the New Federalism
An Urban Institute Program to Assess Changing Social Policies

Preface

Introduction

1997 NSAF Snapshot Survey Weights is the third report in a series describing the methodology of the 1997 National Survey of America=s Families (NSAF). The NSAF has been recently reissued as a companion for the 1999 NSAF Report on the same topic. Related reports for both surveys are on sample design and variance estimation, Report Nos. 2 and 4 in each series.

About the National Survey of America=s Families (NSAF)

As discussed elsewhere (e.g., see especially Report No. 1 in the 1997 NSAF methodology series), NSAF is part of the Assessing the New Federalism Project at the Urban Institute, being done in partnership with Child Trends. Data collection for the NSAF was conducted by Westat.

In each rounds of NSAF, carried out so far, over 40,000 households were interviewed, yielding information on over 100,000 people. NSAF has focused on the economic, health, and social characteristics of children, adults under the age of 65, and their families. The sample is representative of the nation as a whole and of 13 states. Because of its large state sample sizes, NSAF has an unprecedented ability to measure differences between the 13 states it targeted.

About the 1997 and 1999 NSAF Methodology Series

The 1997 and 1999 methodology series of reports have been developed to provide readers with a detailed description of the methods employed to conduct the 1997 NSAF. The two series are nearly parallel, except for the documentation of the public use files, where an on-line system is being used for the 1999 survey and we are planning to reissue the 1997 files on a similar basis.

Report No 1 in the 1997 series introduces NSAF. Report Nos.2 through 4 in both series **B** plus Report No. 14 in the 1997 series -- describe the sample design, how survey results were estimated and how variances were calculated. Report Nos. 5 and 9 in each series describe the interviewing done in for the telephone (RDD) and in-person samples. Report Nos. 6 and 15 in the 1997 series and Report No. 6 in the 1999 series displays and discusses the comparisons we made to surveys that partially overlapped NSAF in content -- including the Current Population Survey and the National Health Interview Survey, among others. Report Nos. 7 and 8 in both series cover what we know about nonresponse rates and nonresponse biases. Report No. 10 in both series covers the details of the survey processing, after the fieldwork was completed, including the imputation done for items that were missing. Report No. 11 in both series introduces the public use files made available.

In the 1997 series, there were additional reports on the public use files available in a PDF format as Report No. 13, 17-22. These will all eventually be superceded by the on-line data file codebook system that we are going to employ for the 1999 survey. The 1997 and 1999 NSAF questionnaires are available respectively as Report No. 12 in the 1997 series and Report No. 1 in the 1999 series. Report No. 16 for the 1997 series, the only report not so far mentioned contains occasional papers of methodological interest given at professional meetings through 1999, regarding the NSAF work as it has progressed over the years since 1996 when the project began.

About this 1997 Report

Report No. 3 focuses on the methods employed to produce estimation weights and the procedures to use these weights to make state and national estimates from the survey data. These weights were used to produce estimates in *Snapshots of America's Families*, the first reports released from the NSAF. see also Report No. 14 which has additional materials on the 1997 weight calculations.

For More Information

For more information about the National Survey of America's Families, contact Assessing the New Federalism, Urban Institute, 2100 M Street, NW, Washington, DC 20037, telephone: (202) 261-5886, fax: (202) 293-1918, Website: http://newfederalism.urban.org. For information about this report, contact by e-mail BrickM1@Westat.com.

Jenny Kenney
and
Fritz Scheuren

Table of Contents

List of Appendices

List of Tables

List of Figures

Chapter 1
Introduction

This report focuses on the methods employed to produce estimation weights and the procedures to use these weights to make state and national estimates from the survey data. These weights were used to produce estimates in the *Snapshots*, the first reports released from the NSAF. Other estimation related activities for the 1997 NSAF are covered in two other reports in this methodology series — on variance estimation, and on other weighting procedures, data editing, and imputation.

The report provides information on the statistical methods used in weighting. We have attempted to give heuristic explanations of the steps to aid readers, especially in the introduction. This chapter is aimed primarily at those interested in learning about the general approaches used in weighting and how the weights are used to produce national estimates. The last section in the chapter provides a "roadmap" of the structure of the remaining chapters of the report.

1.1 Overview of the Survey

The NSAF collected information on the economic, health, and social dimensions of the well-being of children, adults under the age of 65, and their families in 13 states, Milwaukee, and the balance of the nation (see Figure 1-1). In this section we briefly outline the sample design features. More details on the design are given in the *1997 NSAF Sample Design Report*, another report in this series.

Figure 1-1.
Study Areas

Alabama	Massachusetts	New Jersey	Milwaukee County
California	Michigan	New York	Balance of Wisconsin
Colorado	Minnesota	Texas	Balance of Nation
Florida	Mississippi	Washington	

A key goal of the survey was to obtain social and economic information about children in low-income households (income below 200 percent of the poverty threshold). Similar data on children in all households, low-income adults under age 65, and on other adults under age 65 were also required.

The two components of the survey were a random digit dialing (RDD) survey of households with telephones, and an area sample conducted in person for those households without telephones. This dual-frame approach is further described in Waksberg, et al. (1997).

In the RDD sample screener-based subsampling of households was used to sample low-income households at a higher rate than other households. A very short income question was asked during the RDD screening interview, and those that reported an absence of children or reported incomes above 200 percent of the poverty threshold were subsampled. In the area sample,

blocks with very high telephone coverage rates as of the 1990 census were excluded to reduce costs and households containing only adults age 65 and over were screened out.

Within both the RDD and the area samples, household members were subsampled to reduce the respondent burden. If there were multiple children under age 6, one was randomly selected. The same was done for children 6 to 17 years old. Data were collected from the most knowledgeable adult (MKA) in the household for the sampled child. During the MKA interview, data were also collected about the MKA and about his/her spouse/partner. Most questions asked about the MKA were repeated in reference to the spouse/partner; however, some questions on health insurance and health care utilization were asked in reference to only one of the two. The target of these questions was randomly assigned to either the MKA or his/her spouse/partner. There were also some questions that were asked only about the MKA, related to feelings, religious activities, and opinions.

Other adults in households with children were subsampled, as were adults in adult-only households. Adults were eligible only if they would not have been MKAs for other children in the household if those children had been selected. Self-response was required for sampled adults and data were also collected about his/her spouse/partner (if living in the same household). Data were not collected directly from the spouse of a sampled adult. As in the MKA interview, there were also some questions that were asked only about the sampled adult, related to feelings, religious activities, and opinions.

1.2 Overview of Weighting

As noted earlier, weights are applied to sample data to estimate aggregate statistics at the national and state levels. In particular, sample weighting was carried out to accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons;
- Reduce biases occurring because nonrespondents may have different characteristics than respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey; and
- Reduce the variance of the estimates by using auxiliary information.

Weighting is carried out in stages for both the RDD and in-person components to produce household, person, and family weights. Generally speaking, each weight is computed in three stages. The first stage is the computation of the base weight (i.e., the inverse of the probability of selecting the unit). The second stage is an adjustment for nonresponse. In the third stage, the nonresponse adjusted weights are further adjusted to independent controls.

The first stage of weighting is described in chapter 2. The concept is simple — if units are sampled at a rate of 1 in 100, the ones sampled must be weighted by 100 to represent the entire population. The nonresponse and coverage adjustments are also based on this idea, but the

rationale underlying these adjustments is less obvious. For this reason, we describe the philosophy used for these adjustments in the next section.

In NSAF, individual case weights were created and attached to the data records. Separate weights were created for four analysis groups: children, adults 18 to 64 years old (including MKAs), families, and households. Each complete child record has a single weight.[1] For adults, multiple weights are needed because of the way data were collected in the interview, even though the population of interest for both adult weights is all adults ages 18 to 64 years old. The three adult weights are called the adult pair weight (APW), the random adult weight (RAW), and the adult childless weight (ACW). These are described in more detail in the next section. One weight was prepared for households and two weights for families; the two family weights were developed because two definitions of families can be constituted from the NSAF data.

Table 1-1 shows the weights and adjustments for each of the four analysis groups.[2] The last column of the table shows that replicate weights were created for each group. These replicate weights were produced to estimate sampling errors. The methods used to develop the replicate weights and how they can be used to estimate sampling errors of the estimate are described in part 2 of the estimation reports.

Table 1-1.
Weights and Adjustments for the NSAF

	Base Weight	Screener Nonresponse Adjustment	Extended Nonresponse Adjustment	Household Undercoverage Adjustment	Control Total Adjustment	Replicate Weights
Household	X	X		X (area only)	X	X
Child	X		X		X	X
Adult (3 sets)	X		X		X	X
Family (2 sets)	X		X			X

1.3 Weight Adjustment Approach

In an ideal survey, all the units in the inference population are eligible to be selected into the sample and all those that are selected participate in the survey. In practice, neither of these conditions occur. Some units are not eligible for the sample (undercoverage) and some of the sampled units do not respond (nonresponse). If undercoverage and nonresponse are not addressed, then the estimates from the survey will be biased. In the 1997 NSAF, the weights of those that are eligible and respond are adjusted to represent the undercovered persons and nonrespondents. An overview of the approaches used to account for these two sources of missing data is discussed below, beginning with adjusting for nonresponse. In subsequent chapters, the specific adjustments are described in detail.

[1] A second child weight that excludes interviews conducted in the summer is described in part 3 of the estimation reports.

[2] We considered the possibility of having different study areas and national estimation weights, but decided this was impractical.

Nonresponse results in biases in survey estimates when the characteristics of respondents differ from those of nonrespondents. The size of the bias depends on the magnitude of this difference and on the response rate (see Groves 1989). The purpose of adjusting for nonresponse is to reduce the bias. A weighting class adjustment (see Brick and Kalton 1996) method is the type of nonresponse adjustment procedure used in the NSAF. In this procedure, adjustments are computed and applied separately by cell, where a cell is defined using characteristics known for both nonrespondents and respondents. For example, we can identify from the telephone exchange (area code and first three digits of the number) where the household is located (in a central city, outside a central city but in a metropolitan area, or outside a metropolitan area). Thus, this variable can be used to define cells, and weighting adjustments can be computed separately for each of these cells. The adjustment reduces bias if response rates and survey characteristics vary among cells, and the characteristics of respondents are shared by nonrespondents in the same cell. The nature of nonresponse in the 1997 NSAF is explored in Report No. 7, the Nonresponse Studies report in this series.

The drawback to nonresponse adjustment is that it increases the variability of the weights and thus increases the sampling variance (Kish 1992). A nonresponse adjustment is beneficial only when the reduction in bias more than compensates for the increase in variance. When the cells contain sufficient cases and the adjustment factors do not become inordinately large, the effect on variances is often modest. Very large adjustment factors usually occur in cells with small numbers of respondents. To avoid this situation, cells with few cases are "collapsed" or combined to form a new cell with a larger number of cases.

The guiding principle for nonresponse adjustment in the NSAF was to define adjustment cells for which response rates and important survey characteristics vary considerably and to avoid cells with either a small number of cases or a large adjustment factor. Oh and Scheuren (1983) discusses some of the statistical features associated with making these adjustments.

As noted above, nonresponse adjustment classes can be formed only if data are available for both responding and nonresponding units. Since the nonresponse adjustment is done for each stage of data collection, the data available for forming cells is different for each stage. For screening interviews, the nonresponse unit is a household (or more properly a telephone number), and data must be available for all households (telephone numbers). For extended interviews, the nonresponse adjustment is done by type of person (adult or child) or family. At this level, data from the screening interview can be used to define cells.

The approach to adjusting for undercoverage is somewhat different because uncovered units or persons were never even eligible to be sampled. The procedure used to adjust for undercoverage is to use data from external sources (control totals) in a process called poststratification (Holt and Smith 1979). The primary objective of poststratification is to dampen potential biases arising from a combination of response errors, sampling frame undercoverage, and nonresponse. A secondary objective is to reduce sampling errors, which is important because NSAF sample sizes within states are fairly modest for some subclasses. In general, the sample is poststratified to as many independent figures as possible, subject to some constraints. We use the term poststratified loosely here to include raking, a form of multidimensional poststratification (see Brackstone and Rao 1979).

A principle used to determine the poststratification adjustment procedures in the 1997 NSAF was that estimates from other surveys should be used as controls for poststratification only if they have both low bias and significantly lower variances than unadjusted NSAF estimates.[3] Controls at the study area or state level were preferred to national controls, since many analyses will be at the study area level. Control totals had to be based on relatively current data or use variables that are very stable over time. In general, decisions about poststratification were made without examining NSAF estimates to avoid criticisms that the estimates were adjusted to have specific features. These were not strict rules but did provide guidelines. The adjustments are discussed for each analysis unit in chapter 3.

1.4 Using the Weights

As discussed earlier, weights were attached to each responding case so that approximately unbiased estimates could be produced for each study area and for the entire nation. Because households and persons were sampled with very differential probabilities, the use of weights is essential to produce estimates that are representative of the population. This section describes procedures that should be followed to produce estimates from the 1997 NSAF, focusing on which weights should be used in different circumstances. In order to make appropriate statements based on the NSAF data, researchers should also be aware that some questions are asked of subsets of respondents and this must be taken into account to make valid statements from the data.

Approximate unbiased estimates of characteristics of persons and families in the study areas and for the nation can be produced by appropriately weighting the survey responses. The estimates are of the noninstitutionalized[4] population of persons under age 65 in the study areas and in the nation. For families, the estimates are limited to those families with at least one person who is under 65 years old. The weights used to produce study area estimates are the same as those used to produce national estimates.

Five weights are available with the NSAF data, each appropriate for a different set of respondents or group of questions from the survey. The five weights are the child weight (CW), the APW, the RAW, the ACW, and the family weight[5] (FW). These weights are described briefly and examples of using the weights are given.

The **CW** is the weight developed to produce estimates of the number and characteristics of children less than 18 years old. This weight should be used to produce virtually all estimates of children. Children were sampled from households and the MKA answered questions about this child. The CW includes factors that adjust for the probability of selecting the child (including differential factors by reported poverty level and the number of children per household), and nonresponse at the household and person level. Furthermore, the weights were adjusted to be

[3] Preference was given to control totals based on well-defined and consistently reported variables such as age, race, and sex. Key survey outcomes and conceptually complex variables were avoided whenever possible.

[4] Most persons living in group quarters (housing with many unrelated persons such as boarding houses) were also excluded.

[5] There are actually two family weights, each based on different definitions of "family". Since the procedures used to create both family weights and the methods of applying the weights to make estimates are the same, we only discuss one family weight here.

consistent with known totals of the number of children by race, Hispanic ethnicity, age, sex, and tenure (rent or own the home) for each study area and the nation.

The **APW** is the weight developed to produce estimates of adults 18 to 64 years old, for most of the questions relevant for adults. When an adult was sampled from a household, most of the questions about adults were asked about both the sampled adult and their spouse/partner, if the spouse/partner lived in the household. Because both the adult and the spouse/partner were effectively sampled for these questions, the weight is called the APW. The APW includes factors for the probability of selection (of the pair of adults), nonresponse, and adjustments to known totals of the number of adults by race, Hispanic ethnicity, age, sex, tenure, and educational attainment for each study area and the nation.

The **RAW** is the weight developed to produce estimates of 18- to 64-year-old adults for the few questions asked only for a specific adult (not for both the respondent and the spouse/partner). When a child was sampled from a household, the MKA was asked a few questions that were specifically about either themselves or their spouse/partner, but not both. For households without children, these questions were asked about both members of the pair. The probability of selection of the adult for the RAW is the same as for the APW, except for MKAs with a spouse/partner in the household. In this case, the probability of selection for the adult randomly selected is generally twice the pair selection probability. The weighting procedures and control totals are the same for the RAW as for the APW.

The third of the adult weights is the **ACW,** which is used to produce estimates of adults without children in the household for a few questions. These questions were asked only of the 18- to 64-year-old adult respondent, without any random selection of the adult and that adult's spouse/partner. As a result of this procedure, the spouse/partner of the MKA (often the male spouse/partner) had no chance of being sampled and thus no estimates for adults with children are possible for these questions. On the other hand, adults without children were randomly selected to be the respondent, so they all had a chance of being asked the questions. The ACW is appropriate for these adults, for the questions asked only of respondents. The ACW was produced by modifying the APW for the probability of selecting the particular member of the pair, without further adjustment to control totals.

The **FW** is appropriate for making estimates of the number and characteristics of families that have at least one member who is under 65 years old. Families in which all members are over 65 are not included in the 1997 NSAF. Families were defined using two sets of rules. One is essentially the same definition as used in the Current Population (CPS). The other is based on relationships and is broader than the CPS definition. One important difference is that the broader definition includes unmarried partners as members of a family, whereas the CPS definition excludes such persons. Another difference is that all members of the extended family are considered to be in the same family in the broader definition.

The family weight is derived from the APW, with adjustments for the probability of sampling the family. No family level control totals are applied, but the weights using the control totals from the APW are the basis for this weight. In most cases, the APW for the MKA and the family weight are identical. However, in some families more than one person is an MKA and in these cases the FW may not be equal to the APW for either of the sampled MKAs. Even in common

situations, such as a family with one parent and a child who is over 17 years old, the APW weight and the FW will not be identical. The FW should only be used for estimates of the number and characteristics of families.

Even though most households have no more than one family, households and families are not equivalent. For example, households of exactly one person are not family households by definition. Thus, the number of families is substantially less than the number of households. No household weight is available for analysis, although one was created as a part of the weighting process. Rather, for the few questions that are asked at the household level, the data can best be analyzed by changing the unit of analysis to a different population. For example, rather than estimate the percent of households with at least one person born outside the United States, estimate the number of persons living in households in which at least one person was born outside the United States.

Simple Examples

The examples here illustrate relatively simple ways of using the weights to produce estimates. Each of the estimates could be produced at the national or the study area (by restricting the data to the specific area). More complex examples that involve questions asked for different groups of respondents are given later, along with some of the principles that guide these procedures.

Example 1. Characteristics of Children

The CW is used for virtually all of these estimates. Several examples are the number of children (less than 18 years old) who are male, the percent of children who are in a specific grade in school, the number of children who live in two-parent families, the percent of children who live in a family that owns a car, the percent of 14- to 17-year-olds who work, and the percent of children who have an MKA who reports the family has problems paying for food. Notice that some of the examples were of subgroups of children, and no special consideration is needed for these types of estimates.

For some statistics it is possible to estimate either the number of children who live in a family that is below poverty or the number of families with children that are below poverty. In many situations, the former is the preferred statistic because it gives information about the number of children irrespective of the number of children per family. If the researcher chooses to present the child estimate, the CW weight is appropriate rather than the FW.

A related situation arises for estimates about characteristics related to the MKA of a child. Again, estimates can be made of either children or MKAs, but in many situations the child estimate is preferred. For example, the percent of children who have an MKA that reports having problems paying for food (using CW) may be more appropriate than estimates about the percent of adults who responded that they were the most knowledgeable adult about a sampled child.

Example 2. Characteristics of All Adults

For producing simple estimates of adults (18- to 64-years-old), the specific questions must be considered to determine the appropriate weight for analysis. For questions asked of both the respondent and the spouse/partner, the APW is used. Examples are the percent of adults with a high school education, the number of adults born outside the United States, and the percent of adults who live in households with children. Subgroup estimates such as the percent of Hispanic adults who are currently insured can also be made with this weight.

For questions asked randomly of either the respondent or the spouse/partner, the RAW is used. Some examples of such estimates are the percent of adults who did not get or postponed getting care in the last 12 months, the number of visits to the doctor in the last 12 months, and the percent of adults who had health insurance continuously for the last 12 months. Once again, subgroup analysis presents no special problems, but the RAW and not the APW should be used whenever any of the questions are asked only of the randomly selected adult.

No estimates of all adults based on questions asked of the respondent only can be produced. These questions include opinion questions about welfare and questions such as, "During the past 12 months, how often have you felt calm and peaceful?"

Example 3. Characteristics of MKAs

Questions asked of all MKAs can be analyzed using the APW. This includes questions asked of the MKA and the spouse/partner of the MKA and questions asked only of the MKA. Two examples in which the APW is the correct weight are the percent of MKAs that reported arguing with their children, and the percent of MKAs that never attended religious services in the last 12 months. Estimates of subgroups such as MKAs that are under 40 years old can also be made using this weight. Note that MKAs outside the 18- to 64-year-old age range are not represented in these estimates.

If the characteristic of the MKA is asked only for the randomly selected adult, then the RAW should be used. For example, the number of MKAs that had more than one visit to the emergency room in the last 12 months is estimated using the RAW.

Example 4. Characteristics of Adults Without Children

For most questions, adults without children can be handled as a subgroup and the APW or RAW weight is appropriate, as discussed in example 2. However, a few questions were asked only of respondents and estimates of adults without children in the household (not all adults as mentioned in example 2 can be produced for these using ACW). For example, the percent of adults without children in the household who never attended religious services in the last 12 months can be estimated using ACW. For estimating characteristics of subgroups that are based on these questions,

even if other questions involved in the estimates are asked of all adults, the ACW must be used.

Example 5. Characteristics of Families

The number and characteristics of families are estimated using the FW. As described in earlier examples, sometimes the preferred method is to make estimates of children (CW) in families with a characteristic or adults (APW or RAW) in families with a characteristic instead of making family estimates directly. Some estimates of families using the FW are: percent of families with income below the poverty level, the percent of families with children, the percent of families with children and two parents. Subgroup analysis of families can also be conducted. Caution is needed when making family estimates so that characteristics of a specific individual are not presumed to hold for the entire family. For example, even if the sampled adult did not have any emergency room visits in the last year, this does not mean that no one in the family did.

Principles for Handling More Complex Estimates

In the examples above, some general principles for producing estimates were applied, but not completely explained. Now, these general principles are stated explicitly and extended to cover more complex examples.

One of the first tasks is to *determine the appropriate unit of analysis*. In the examples above, this often involved making a choice between the child and the MKA, the child and the family, or the adult and the family. Once the unit of analysis is determined, the choice of the appropriate weight is relatively simple. If the child is the unit of analysis, the CW is appropriate; if the family is the unit, then the FW is used; and, if the adult (or MKA) is the unit, then the APW, the RAW, or the ACW is used.

A second principle involves choosing between the APW and the RAW. In example 2, the choice was dependent on whether the question was asked of all adults or only randomly selected adults. The principle is that if the characteristic can only be obtained from a subset of the respondents, then the weight for that subset must be used in the analysis. An example demonstrates this principle in a more complex situation.

Example 6. Characteristics of Adults from Merged Files

An example of a more complex nature is the percent of adults who are currently employed and had continuous health coverage for the past 12 months. Current employment is available for all adults and could be estimated using the APW, but continuous health coverage over the past 12 months is only available for randomly sampled adults and requires the use of the RAW. Since the smallest subset is the randomly sampled adults, the RAW must be used in this analysis.

The same principle holds for doing subgroup analysis. If any of the characteristics that define the subgroup or are involved in the estimates within the subgroup are from the randomly sampled adult, then the RAW must be used.

The hierarchy in the use of the weights for adults is that the APW weight should be used unless questions asked only of the randomly sampled adult or of the respondent are involved. If any of the questions in the analysis are only asked of the randomly sampled adult, then the RAW is used. Finally, if any question that is only asked of the respondent is in the analysis, then the ACW must be used. As noted in example 2, estimates for all adults cannot be made for these questions.

Example 7. Characteristics of Persons Under 65 from Merged Files

Another example is estimating the number of persons under 65 (including children and adults) with some characteristic. To do this, concatenate the child and appropriate adult files, rename the weights so that they are the same, and then carry out the analysis. For example, to estimate the number of persons under 65 who live in a family with income below 200 percent of the poverty level, merge together the child and adult pair files and rename the CW and the APW to be, say, TOTW. The analysis can then proceed with TOTW. If the analysis was to be of persons under 65 who were uninsured at some time in the last year, then the child (with the CW) and the random adult (with the RAW) files would have to be merged. This follows because the insurance question is only asked of random adults.

The other principle in determining the appropriate weight depends on the subgroups the questions are about, in other words, the questionnaire skip pattern. The skip pattern has been stated explicitly in the examples above. These cover major subgroups, such as the randomly selected adult and the respondent.

An example that brings this point out clearly involves the few questions that are asked only about the male member of a pair or male respondent.[6] Estimates of adult males can generally be considered as a simple subset, and the rules regarding the use of the APW or the RAW can be applied as described earlier. However, the next example involves questions asked only about the male in a spouse/partner situation.

Example 8. Characteristics of Males Under 65

Only adult males are asked if they have any children under 18 years old who live outside the adult's household. To estimate the number of males who have children under 18 years old living outside the household, the APW is used. This easily extends to questions only asked of the randomly selected adult. For example, to estimate the percent of males who have children under 18 years old living outside the household who have had insurance continuously for the past 12 months, the RAW is used.

[6] In 37 cases both the MKA and spouse/partner were male and these questions were only asked about the respondent. This results in a very small bias in these estimates.

These principles and examples should guide researchers in using the weights developed in the 1997 NSAF. Once the principles are understood, applications will become relatively straightforward with practice.

1.5 Structure of the Report

So far, we have presented a general discussion about the weights, the principles for their construction, and how they can be used to produce estimates from the 1997 NSAF data. Chapter 2 provides a more detailed and technical description of the development of the base weights and the adjustment of these weights for nonresponse. Each of the weights described above is included in the description. Chapter 3 focuses on the adjustments made to the weights to align the estimates with external control totals, essentially the poststratification step mentioned above.

To be able to fully understand the weighting procedures described here, the details of the sample design are very important. Reading the *1997 NSAF Sample Design Report* -- report No 2 in this methodology series -- is the best way to become familiar with those details. Several other reports in the 1997 methodology series relate to 1997 NSAF estimation issues. These are reports Nos. 4, 10, 14 and 16. Report No.4, for example, focuses on variance estimation and discusses how to make estimates of the sampling precision from the 1997 NSAF. Report No. 10 covers estimation topics such as data editing, imputation, and other survey steps taken to reduce nonsampling errors. Report No. 14 is of particular value since it takes the estimation steps described here further -- so that researchers can produce undercount adjusted weights. Finally, see Report No. 16 for related technical papers that cast more light on the approaches to estimation taken, connecting them up with best practice elsewhere.

Chapter 2
Base Weights and Nonresponse Adjustments

In this chapter, the base weights and adjustments of these weights due to nonresponse are described. We begin by presenting the steps involved in preparing the household weights. These steps include the base weights and nonresponse adjustments for both the RDD and area samples, as well as special adjustments for subsampling RDD households and for excluding some block groups (BGs) in the area sample. The poststratification of the household weights is discussed in the next chapter.

The three subsequent sections of this chapter deal with creating weights for the three main analysis groups — children, adults, and families. Each of these weights is based on the household weight but has further adjustments for subsampling and nonresponse. Chapter three then describes how these weights are later adjusted to external control totals to form the final weights for analysis.

To simplify presentation of these materials, the results of the weighting for one study area, Michigan, are used to illustrate the procedures. Detailed tables with the same information for the other study areas are given in appendix A.

2.1 Household Weights

Since the sampling procedures for the telephone and area samples were different, the weighting of these components is somewhat different. The discussion that follows reflects this. For example, the method of forming base weights for the RDD sample is given first, followed by the method for the area sample.

2.1.1 Base Weights

For the RDD telephone sample, the base weight of a telephone number is the ratio of the total number of telephone numbers in 100-banks with at least one listed telephone number to the number of telephone numbers sampled for screening by site. This can be written as

$$BW_{RDD} = \frac{N \cdot 100}{n},$$

where

N = Total number of listed 100-banks, and
n = Number of telephone numbers in the sample.

Table 2-1 shows the base weight for the RDD sample, by study area.[7] It should be noted that the study areas are the ones used at the time of initial sample selection, prior to eliminating Ohio and Pennsylvania as separate study areas (see the *1997 NSAF Sample Design Report* for a discussion of the sampling). The original balance of the United States also included Colorado because that study area was added later. These original sites are shown because all the telephone numbers sampled within them had the same chance of being selected and the same base weight.

Table 2-1.
Released Sample of Telephone Numbers and Number of Working Banks

Study Area	Number of 100-Banks	Sample Numbers	Base Weight
Alabama	28,228	20,000	141.14
California	237,385	27,000	879.20
Colorado	30,683	33,270	92.22
Florida	117,308	34,000	345.02
Massachusetts	47,132	44,000	107.12
Michigan	70,397	30,000	234.66
Minnesota	35,106	32,000	109.71
Mississippi	15,465	16,000	96.66
New Jersey	63,865	52,000	122.82
New York	145,356	34,000	427.52
Texas	139,601	27,000	517.04
Washington	43,507	32,000	135.96
Milwaukee	7,542	30,000	25.14
Balance of Wisconsin	30,431	30,000	101.44
Balance of U.S. (Original Sample)	813,000	33,730	2,410.32
Balance of U.S. (Ohio)	90,115	3,886	2,318.97
Balance of U.S. (Pennsylvania)	88,471	4,374	2,022.66

Table 2-1 shows that the weight for each sampled telephone number in Michigan is 234.66. Since the number of telephone numbers selected from Michigan was 30,000, the sum of the weights for all the sampled telephone numbers in that state, prior to any other adjustment, is 7,039,700.

For the area sample, the household base weight is the product of reciprocals of the overall probability of selection of the PSU in the site, the probability of selecting the cluster within the PSU, and the probability of selecting the household within the cluster. Some of the sampled clusters (or segments) were large and these were subsampled to achieve the target sample size. The subsampling is described in Report No. 2. The household base weight that reflects this subsampling is given by

[7] We examined the possibility of adjusting these base weights because some households may have had telephone service for part of the data collection period but not all the time. After reviewing several options, we decided that no modification would be made.

$$BW_{area} = \frac{1}{POS_{psu}} \cdot \frac{1}{POS_{c1}} \cdot \frac{1}{K_{c1}},$$

where

POS_{psu} = probability of selection of the PSU;
POS_{c1} = probability of selection of the cluster; and
K_{c1} = within cluster subsampling rate.

Three segments were not completely listed during data collection due to unavoidable circumstances. The base weights in these three sites were adjusted to account for the segments not listed. In the sites without an unlisted segment, the adjustment was equal to one, otherwise the adjustment is

$$UL_s = \frac{\sum_L w_{si} + \sum_{UL} w_{si}}{\sum_L w_{si}},$$

where *s* is the site or study area and

w_{si} = BW_{area} for segment *i* in site *s*;
L = listed segments; and
UL = unlisted segments.

The final area base weight, FBW_{area}, adjusted for unlisted segment is

$$FBW_{area} = BW_{area} \cdot UL_s .$$

The base weights for a particular site are not simple to illustrate because the probabilities of selection for each primary sampling unit (PSU) and cluster within the PSU, as well as the subsampling rates are different. In Michigan, a total of 64 segments were sampled and listed, resulting in 2,263 households being listed. Table A-2 row 5 gives these numbers for other sites, along with any adjustments for unlisted segments.

After reviewing the data, it was observed that, in a few states, one segment in the area sample accounted for a large proportion of the total nontelephone sample. For example, in Florida one segment had 24 nontelephone households out of the total of 90 nontelephone households in the state. The next largest segment size was six nontelephone households. This variation in segment size had two effects. First, the estimates of nontelephone households in the study area were heavily influenced by the characteristics of households in the large segment. Second, the variances of the estimates were inflated in very much the same way that variation in weights affect variances (Kish 1992). To ameliorate these effects, the weights of households in the segments with a very large number of nontelephone households were reduced and the weights of the nontelephone households in the other segments in the state were increased a corresponding

amount.[8] The result was the total number of nontelephone households in the state was unaffected, but the sample from the large clusters had a reduced effect on the study area estimates and variances. The goal was to reduce both the conditional bias and variance of the estimates,[9] primarily for state level estimates. The revisions were done in segments in each of the following study areas: Alabama, Colorado, Florida, Minnesota, Mississippi, New Jersey, Texas, Washington, and the balance of the United States. This step was done prior to nonresponse adjustment of the area sample.

2.1.2 Nonresponse Adjustment

Nonresponse adjustments at the household level were done using a weighting class adjustment procedure, as discussed in chapter 1. The base weight for a completed screening interview was multiplied by the nonresponse adjustment to produce the nonresponse adjusted weight. If the adjustment is effective, then the resulting estimates should be approximately unbiased. The nonresponse adjustments for the RDD and area sample are described below.

RDD Nonresponse Adjustment

The first step, prior to adjusting the RDD base weights for screening interview nonresponse, was to estimate the number of eligible residential telephone numbers in the sample. This is necessary because some telephone numbers, despite being dialed many times over a period of time, cannot be classified as residential or not. The telephone numbers that did not complete the screener interview fall into one of the following three categories:

a. Unknown residential status: All telephone numbers with undetermined residential status (such as ring with no answer on repeated calls);

b. Residential, unknown eligibility status: All telephone numbers determined to be households, but with no data on the presence of persons under 65 years old; or

c. Residential, known eligible: All telephone numbers that are known to be eligible households (with at least one person under 65 years old) for which the screening interview was not completed.

The first step was to estimate the number of residential telephone numbers from those classified with an unknown residential status (a). This was done by multiplying the number of telephone numbers (actually the sum of the base weights of these cases) that were never answered (NA) by 0.27 and the number of answering machine (NM) cases (again, the sum of the base weights) by 0.63. The proportion for NA cases, 0.27, was estimated from the National Immunization Survey (NIS) from calls made to the telephone company business office. (This proportion is shown in

[8] The distributions of segment sizes were inspected for each site and a trimming type of rule (see Potter 1988) was used to set the threshold for that site. For example, in Florida the large segment had its sum of weights cut in half and the weight distributed to the other nontelephone households in the site proportional to their original weight.

[9] The conditional approach, described by Holt and Smith (1979), was used frequently to help resolve issues of this nature. The effect on the variances of this adjustment are discussed in the methodology report, *1997 NSAF Variance Estimation*.

table 3 of an unpublished version of Shapiro, et al. (1995).) The proportion for NM cases, 0.63, was estimated from the 1995 National Household Education Survey (NHES:1995) (see Brick and Broene 1997).

The sum of the weights for numbers with unknown residential status in Michigan, after being multiplied by the appropriate factors, is 140,315. These values for the other sites are shown in table A-1, row 3.1.c. All the weights are distributed to the other telephone numbers in the category, residential numbers with unknown eligibility. For Michigan, this means the 140,315 was added to the previous sum of weights for residential unknown eligibility (542,057) to bring the total in this category to 682,372 (see row 3.2.b in table A-1).

The next step was to estimate the number of eligible households from those classified as having a residential number with unknown eligibility status (this includes those in (b) plus those estimated to be residential from (a)). This is computed by multiplying the number of residential numbers for which the eligibility status is unknown by the percentage eligible based on those numbers for which eligibility status is known. Since the weights for the numbers with unknown eligibility are set equal to zero, this process can be written as adjustment that is applied to the base weights of those numbers with known eligibility status. The adjustment is

$$NR1 = \frac{\sum_U w_i + \sum_I w_i + \sum_{ER} w_i + \sum_{ENR} w_i}{\sum_I w_i + \sum_{ER} w_i + \sum_{ENR} w_i},$$

where

w_i = BW_{RDD} for household i;
U = households with unknown eligibility status;
I = ineligible households;
ER = eligible respondent households; and
ENR = eligible nonrespondent households.

This adjustment was created for each study area. For example, in Michigan the adjustment factor is

$$NR1_{MI} = \frac{682{,}372 + 562{,}003 + 2{,}049{,}726 + 56{,}318}{562{,}003 + 2{,}049{,}726 + 56{,}318} = 1.256,$$

where the sums of the weights are given in table A.1, rows 4.1 a-d. When the adjustment is applied, the weights of the unknown eligibles (682,372) are distributed to the other three categories. So, the adjustment of 1.256 multiplied by the number of residential eligible respondents (2,049,726) gives the estimated number of eligible residential respondents of 2,583,046. The corresponding adjusted values for the other sites are given in table A-1, rows 4.2 a-c.

It is now possible to adjust for nonresponse because the residential and eligibility issues have been taken into account. The nonresponse adjustment is just the ratio of the sampled eligible

households to those that completed the interview. The weight for respondents (all of whom are eligible) is multiplied by the following factor:

$$NR2_c = \frac{\sum_{ER}\left(NR1_c \cdot w_{ci}\right) + \sum_{ENR}\left(NR1_c \cdot w_{ci}\right)}{\sum_{ER}\left(NR1_c \cdot w_{ci}\right)},$$

where *c* is an adjustment class, as defined below. The weights for nonrespondents were set equal to zero.

For example, the weights of the nonrespondents in Michigan (whose sum equals 71,685 before the adjustment) are set equal to zero. The weights of the respondents are multiplied by NR_c (which depends on which class the respondent belongs) and the adjusted sum of the weights of the respondents is increased by 71,685 to 2,654,731 (see table A-1, row 5.1.a, for the adjusted sum of weights for the other sites).

The classes or cells used to form the weighting class adjustments[10] are important because nonresponse bias may be reduced if the cells are correlated with response propensity (Little 1986). One important correlate is whether there was an address for the telephone number (i.e., it was listed and the letter could be mailed to the address). A higher response rate is typical for listed numbers and is documented for the 1997 NSAF in the Response Rates and Methods Evaluation Report in this series. Response rates also vary by other variables such as metropolitan statistical area (MSA) status. MSA status is one of several variables available at the exchange level for the 1997 NSAF; others include percent in race/ethnicity groups, percent of households in specified income ranges, percent renters, percent with children, and percent college graduates.

A series of univariate profiles of screener nonresponse rates by these exchange-level variables provided some indication of response propensity, but it was difficult to interpret these because there were many characteristics to consider. Furthermore, some of the characteristics were correlated and univariate profiles did not explore these relationships. A multivariate analysis was more appropriate for examining the relationship of the characteristics and the response rates.

A categorical search algorithm called CHAID (Kass 1980) was used to divide the data set into cells with the greatest discrimination with respect to the response rates. In other words, CHAID attempts to divide the data set into groups so that the response rates within cells is as constant as possible, and the response rates across cells are as different as possible.

If a nonresponse adjustment cell has less than 30 cases or has a very large adjustment factor, it may cause a large increase in the variance of the estimates. CHAID has parameters that control the cell sizes. We set this so that no cells had fewer than 30 cases. We also examined the cells with large adjustments formed in CHAID. These cells were combined or collapsed with other similar cells to form new cells with smaller adjustments. For example, if the original cell for the

[10] Other weighting methods, such as raking to the sample counts (see Kalton and Kaspryzk 1986) could have been used instead of the weighting class approach.

central city of an MSA had a large adjustment factor, then a new combined cell would have been formed by collapsing together the central city and in the MSA but not in a central city.

The exchange level variables used in the CHAID analysis are shown in table 2-2.[11] The results of the CHAID analysis are shown in table 2-3 by study area. This shows that the variable, CMAILED (the variable that indicated if a valid address was available for the number) was the most powerful predictor in most sites. The variables in the table with numbers assigned were used to form the adjustment cells.

The nonresponse adjusted RDD household weight, ARW_{RDD}, can be written as

$$ARW_{RDD,i} = BW_{RDD,i} \cdot NR1 \cdot NR2_c \,,$$

where the BW_{RDD} is as defined above, except it also includes the allocation for the unknown residential telephone numbers described earlier.

[11] We explored the possibility of using Donnelley data at the individual household level for creating cells but rejected it; our analysis showed the data were not accurate or complete and in addition, they were costly.

Table 2-2.
Variables Used in the CHAID Analysis for RDD Nonresponse

Variable Name	Description	Levels	Description
CPERHISP	Percent of Hispanic in the exchange	1 2 3 4 5 6	Less than 2.5% 2.5%– less than 5% 5%– less than 10% 10%– less than 20% 20%– less than 30% 30% or more
CPERBLCK	Percent of black in the exchange	1 2 3 4 5 6 7	Less than 2.5% 2.5%– less than 5% 5%– less than 10% 10%– less than 20% 20%– less than 30% 30%– less than 40% 40% or more
CPERCHLD	Percent of persons under 18 in the exchange	1 2 3 4	Less than 20% 20%– less than 25% 25%– less than 30% 30% or more
CPERENT	Percent of renters in the exchange	1 2 3 4 5	Less than 20% 20%– less than 30% 30%– less than 40% 40%– less than 50% 50% or more
CMEDINC	Median income in the exchange	1 2 3 4	Less than $20,00 $20,000– less than $35,000 $35,000– less than $45,000 $45,000 or more
CAVGRNT	Average rent in the exchange	1 2 3 4 5	Less than $300 $300– less than $400 $400– less than $500 $500– less than $600 $600 or more
METSTAT	Metro status where the exchange is located	1 2 3 4 5	Inside central city of an MSA Outside central city but inside the county containing the central city Inside a suburban county of an MSA In an MSA that has no central city Not in an MSA
CI0_25	Percent of families in exchange with income less than $30,000	1 2 3 4 5	Less than 20% 20%– less than 30% 30%– less than 40% 40%– less than 50% 50% or more
CLISTED	Listed telephone flag	1 2	Listed Unlisted
CMAILED	Mailed flag (based on valid address)	1 2	Had a valid address Did not have a valid address

Table 2-3.
Variables That Were Predictive* of Nonresponse in the RDD Sample, by Study Area

Site	CMAILED	CPERHISP	CPERBLCK	CPERCHLD	CPERENT	CMEDINC	CAVGRNT	METSTAT	CI0_25
Alabama	2						1		
California	2	3	2	1		4	3	3	4
Florida	1		3	3	3	3	2		4
Massachusetts	1			3		2	2	3	
Michigan	1			3			2	2	
Minnesota	1				4			2	3
New Jersey	1			2		3	2	4	
New York	1	4	3	2		3	2	3	
Texas	2		2	3		1	4	3	
Washington	1	2		3		3		3	2
ML	1			2					2
BW	1			3	4		2		
Mississippi	2					3		1	
Colorado	2			3	2	4	1	3	
Balance of U.S.	1		3		3	3	2	3	3

* The number indicates the variable was found to be predictive of the response rate. A lower number indicates the more powerful predictor.

Area Sample Nonresponse Adjustment

The procedures adjust mean nonresponse for the area sample are simpler than for the RDD sample. The first step accounts for unknown eligibility status and the second step adjusts for nonresponse in the eligible nontelephone households.

For the area sample, nonresponding households were classified into one of the following categories:

a. Unknown eligibility status: All households where the presence of a telephone or the presence of persons under 65 years old was not determined. The few incompletely listed segments were accounted for earlier. As discussed in *1997 NSAF In-Person Survey Methods*, reports from neighbors were not accepted in the survey so the information on eligibility had to come from the household itself.

b. Known eligible nontelephone household: All nontelephone households with eligible persons (with at least one person under 65 years old) for which the screening and prescreening interviews were not completed.

The first step was to estimate the proportion of households that were eligible (i.e., those without a telephone and with someone under 65 years old). This proportion is needed to estimate the number of eligible households for those whose eligibility was not determined. The proportion is

$$P = \frac{\sum_{C} w_i + \sum_{R} w_i}{\sum_{I} w_i + \sum_{C} w_i + \sum_{R} w_i}$$

w_i = FBW_{area} for household i;
A = households with unknown eligibility status (used below);
I = ineligible households (telephone or all members over 64 years);
C = eligible nonrespondent nontelephone households; and
R = eligible respondent nontelephone households.

In Michigan the proportion, based on the number in table A-2, rows 4.1 a to d, is

$$P = \frac{4{,}243 + 50{,}863}{1{,}423{,}293 + 4{,}243 + 50{,}863} = .037 .$$

The next steps are similar to the process used in the RDD sample. The first adjustment is for unknown eligibility status and is given by

$$NR1 = \frac{P\sum_{A} w_i + \sum_{C} w_i + \sum_{R} w_i}{\sum_{C} w_i + \sum_{R} w_i} .$$

For Michigan, this factor is

$$NR1 = \frac{.037(3{,}551) + 4{,}243 + 50{,}863}{4{,}243 + 50{,}863} = 1.002 .$$

The factors for the other sites can be reproduced using the values in table A-2, rows 4.1 a to d.

The nonresponse adjustment is the ratio of the sampled eligible respondents to those that completed the interview. The weight of a respondent (all of whom are eligible) is multiplied by the following factor

$$NR2_c = \frac{\sum_{C} \left(NR1_c \cdot w_{ci}\right) + \sum_{R} \left(NR1_c \cdot w_{ci}\right)}{\sum_{R} \left(NR1_c \cdot w_{ci}\right)},$$

where c is an adjustment class based on the geography of the sampled household (segments), discussed below. For Michigan, table A-2 shows that the sum of the weights for the nonrespondents (4,310) were distributed to the respondents (50,929) to bring the sum of the

nonresponse adjusted weights of the respondents to a total of 55,239, as shown in the table (row 4.3). The weights for the nonrespondents were set to zero.

The nonresponse cells were defined by segment. If a segment had fewer than 10 sampled households or an adjustment factor over 2.0, then that segment was collapsed with other segments in the same PSU. The rule used in collapsing was to combine those segments with the most similar values of the percentage of households without telephone numbers in the BG (based on the 1990 Census data).

The nonresponse adjusted nontelephone household weight, NRW_{area}, is

$$NRW_{area,i} = FBW_{area,i} \cdot NR1 \cdot NR2_c .$$

The weights for all but respondents are zero.

2.1.3 RDD Subsampling Adjustment

The next adjustment in the RDD weights accounts for the subsampling of households, both by income and presence of children. All households identified as being low income in the screening were asked to participate in the survey, while only a subsample of the households with high and unknown income were included. The weight adjustment for the subsampling was done after the nonresponse adjustment because the income category was obtained only for respondents. The adjustment for subsampled households is the ratio of the sum of the weights for the households screened with incomes over 200 percent of poverty (nonpoor) over the sum of the weights for the subsampled households. The same type of adjustment was done for those screened with unknown poverty status at the screener, and both were done separately by whether or not the household had children.

We write this adjustment factor as S_l, where l denotes one of the four categories (households with or without children crossed with poverty status, nonpoor, or unknown). For example, table A-1, rows 9.1 b and c, show that in Michigan there were 268,477 weighted nonpoor households screened and subsampled, while 589,876 weighted nonpoor households were screened but not subsampled. Thus,

$$S_1 = \frac{589{,}876 + 268{,}477}{268{,}477} = 3.20$$

and consequently the weight for each subsampled nonpoor household with children is multiplied by this factor. Thus, the sum of the weights for each category is retained (see table A-1, row 9.2 b). This table also shows the weights for the households in the other categories by site.

The next adjustment accounts for subsampling households without children. When a telephone number was sampled originally, it was randomly assigned to either the "option A" or the "option B" path (see the *1997 NSAF Sample Design Report* for more details on the subsampling procedures). If it was assigned to the "B" path, then all adults in the household were eligible to

be subsampled for an interview. For households with children, this means adults who were MKAs of children in the households as well as other adults ("straggler B" adults) were eligible for sampling. If, on the other hand, the household was assigned to the "A" path, the only adults sampled were MKAs (and the spouse/partners of MKAs) who were sampled with the focal child.

The weighting factor applied to account for this subsampling is relatively simple for households with no children. The adjustment is the estimate of the numbers of households in the site divided by the estimated number of households subsampled for the "option B" path. This adjustment is

$$A_{nc} = \frac{\sum_{i \in nokids} w_i}{\sum_{i \in nokids} w_i \cdot d_i(B)},$$

where the summation is over all households with no children, w_i is the nonresponse adjusted weight, and $\delta(B)$ is an indicator variable equal to one if the household is sampled for the "option B" path and zero otherwise. The adjustment is applied to the weights of households without children in the "option B" path, so $A_{nc} = 1$ otherwise.

For Michigan, the values to produce this adjustment are given in table A-1. Rows 10.1.a and 10.2.a show that the sum of the weights for households with children are not affected, as noted above. However, the adjustment for households with no children is derived from the values in rows 10.1.b and 10.1.c, and the adjustment is

$$A_{nc,MI} = \frac{882{,}193 + 457{,}341}{882{,}193} = 1.52 .$$

For households with children, the same principle was used to account for the subsampling, but the subsampling was only done for "straggler B" adults (the MKAs were sampled along with the focal child in both paths). This adjustment is addressed in a later section of the chapter with the other adult-level weighting procedures since the subsampling in this case was not at the household level.

A supplemental subsample of adults was selected when it was determined that too few adults, especially low-income adults, were being interviewed. To increase the sample size, an additional sample of RDD households in the sites with too few adults was selected. The supplement of "refielded" households met all the following conditions: they were respondents to the original screener interview, they did not have children, and they were not subsampled for "option B." Table A-1 shows that of the 3,350,419 households in Michigan (row 7.1.a), 3,104,703 (row 7.1.f) were not included in the supplement. The refielded cases are called the "IK" supplemental sample because the households had the disposition code "IK" before they were resampled.

The supplemental sample affects only the weights of households without children. In theory, the only change required to account for the supplemental sample is to change the A_{nc} weighting factor to account for the additional "IK" households being sampled. However, some of the households that responded to the original screener interview did not respond when they were refielded for the second screening interview. Some refused or fell into some other nonresponse

category, and the weights for these households were adjusted for this additional level of nonresponse.

Some households did not participate in the second screener because the status of the telephone number changed in the time between the original and second screening interviews. For example, some telephone numbers were nonworking when called for the second screening interview, even though we obtained responses at the first interview. Operationally, we classified households with a changed status at the time of the second screener (those who no longer had the same residential working number, those in which everyone was now older than 65, and those that now had children) as ineligible. In Michigan, the sum of the weights of the ineligible households was 24,671 (Table A-1, row 7.1.c), about 10 percent of all the households included in the supplemental sample. This means that adjustments for these types of changed status households is included in the poststratification rather than the nonresponse adjustment.

The nonresponse adjustment for the "IK" supplemental sample in households without children is the standard nonresponse adjustment factor at the site. In the sites with no supplemental sampling this factor is unity. The factor in sites with the supplemental sample is

$$NR(IK) = \frac{\sum_{R} ARW_i + \sum_{NR} ARW_i}{\sum_{R} ARW_i},$$

where the summations are over the sample of re-released "IK" households. This formulation is a simplification because the adjustment actually includes the allocation of the households with unknown eligibility, as was done in the previous RDD nonresponse adjustment. For example, in Michigan the 32,853 households with unknown eligibility (Table A-1, row 7.1.d) are proportionally allocated to the households with known eligibility. This results in the sum of the weights being 184,778 for "IK" respondent households (row 7.2.a), 3,692 for the nonrespondents (row 7.2.b), and 28,894 for the ineligible households (row 7.2.c). Now, the sum of the weights of the respondents is multiply by the adjustment factor that brings it to the total of the respondents and nonrespondents (188,470 given in row 7.3.a).

The household weight that incorporates the nonresponse adjustment at the screener level, including the special adjustment for the supplemental sample nonresponse, can now be written as a product of the appropriate factors. The weight includes the subsampling of households by poverty status and the subsampling of households without children. The nonresponse adjusted weight, NRW_{RDD}, is

$$NRW_{RDD,i} = ARW_{RDD,i} \cdot S_{l,i} \cdot A_{nc} \cdot NR(IK),$$

where the terms were defined earlier. Note that the additional factors on the right are equal to 1.0 if the household was not subject to the subsampling or nonresponse. In other words, $S_{l,i} = 1$ if the household was screened as below 200 percent poverty, $NR(IK) = 1$ if the household was not in the supplemental sample, and $A_{nc} = 1$ if the household had children.

2.1.4 Area Undercoverage Adjustment

The RDD and area weights (NRW_{RDD} and NRW_{area}) include adjustments for all the stages of sampling and nonresponse at the household level. Despite this, estimates using these weights underestimate population totals because of undercoverage. Both the telephone and area sampling frames are incomplete by design. The telephone sample excludes banks of telephone numbers that have no listed numbers; the area sample excludes households in BGs that had a high percentage of telephone households according to the 1990 Census of Population. To reduce the undercoverage bias in the estimates from the survey, some adjustments in the weights are necessary.

The RDD sampling frame excludes less than 3 percent of all telephone households. Furthermore, studies of this source of undercoverage indicate little difference in the characteristics of the telephone households in listed 100-banks (the covered population) and telephone households in the 100-banks with no listed numbers (the noncovered population) (see Brick et al. 1995, and Giesbrecht et al. 1996). Because the percentage undercovered and the differences in the characteristics of the covered and noncovered are both small, and the telephone households account for more than 97 percent of all households, household poststratification is adequate to adjust the weights of the telephone sample for the undercoverage. Household poststratification is discussed in the next chapter.

On the other hand, a special adjustment before poststratification is needed to account for the undercoverage of nontelephone households from the area sample because the percentage excluded is larger in this component.[12] About 8 percent of nontelephone households were in the excluded BGs, as computed from the 1990 Decennial Census data file. The coverage rate varies somewhat by site. Furthermore, the actual percentage of nontelephone households in the excluded BGs in 1997 is greater than the 8 percent computed from the 1990 Census as discussed in the *1997 NSAF In-Person Survey Methods*. Special CPS tabulations obtained from the Census Bureau indicate that much more than 8 percent of nontelephone households are in excluded BGs in 1997 for most states. See Waksberg, et al. (1998) for more on this issue.

The benefit from doing a special coverage adjustment is that it increases the weights of nontelephone households for the undercoverage of nontelephone households before poststratification. Since the poststratification adjustment is not done separately for telephone and nontelephone households, the undercoverage adjustment should be done first. Furthermore, the adjustment allows the use of BG characteristics, sample size permitting. This more detailed level of geography may produce adjustments that reduce undercoverage bias.

Several methods could be used to account for the exclusion of BGs from the sampling frame, most of which can be expressed as an adjustment factor applied to the weights (NRW_{area}) of the responding nontelephone households. For example, we could simply multiply the weights of the responding nontelephone households by the ratio of the number of BGs in the population to the number of BGs in the sampling frame. This method is not advisable because it treats all BGs

[12] Since the telephone undercoverage due to zero banks is being ignored, it could be argued that the area sample should only be adjusted to this same level of undercoverage. However, as discussed later in this section, the undercoverage adjustment procedure is not so precise that it is feasible to fine-tune it to this level.

equally, and the percent of nontelephone households in the excluded BGs was designed to be lower than in the included BGs (about 55 percent of the BGs were excluded while only 8 percent of the nontelephone households in 1990 were excluded). This method of adjusting would overcompensate for the undercoverage and estimate many more nontelephone households than actually exist.

Another way of expressing the undercoverage problem is in terms of a model of the relationship between the units included and excluded from the sampling frame. Consider a model in which a covered BG could be "exchanged" or substituted for a noncovered BG, so that the distribution of the estimates is the distribution that would have been obtained if no BGs were excluded. In this model, the BG adjustment mentioned above would be appropriate. The problem is that a key model assumption does not hold; a covered BG is not exchangeable with a noncovered one because it generally contains too many nontelephone households. An alternative model that assumes households from the covered frame are exchangeable with households from the noncovered frame is inappropriate for the same reason.

A model with more appeal posits that a nontelephone household from the covered frame can be exchanged for a nontelephone household in the noncovered frame. This model can be implemented as an adjustment factor applied to all responding covered households; the factor is the ratio of the number of nontelephone households in the population to the number of nontelephone households in BGs in the sampling frame. Both of these counts can be obtained from the 1990 Census tabulations. The ratio is

$$R_1 = \frac{T_\cdot}{T_f},$$

where T_f is the number of nontelephone households in the sampling frame, and $T_\cdot$ is the number of nontelephone households in the entire population.

An improvement in the model is to define classes (using variables in addition to site) in which the nontelephone households are more likely to be exchangeable. Characteristics including race, poverty, low income, family type, region, metropolitan status, and size of household were considered by Thornberry and Massey (1988) as sources of variation in the percent of households without telephones. Since poverty status and family type are so highly correlated with many of the analysis outcomes of the survey, these variables were considered in defining classes by examining special tabulations from the March 1997 CPS. As a result of this analysis, classes were defined based on the 1990 Census percentage of households below poverty in the BG.

When the value of R_1 was very large or based on a small cell size, the cells were collapsed. The rules for collapsing were similar to those discussed earlier for nonresponse weighting classes. Since R_1 was calculated using Census data rather than survey data, the main concern was the number of NSAF sample segments with each class. Cells were collapsed when there were less than two sample segments in a cell.

Another slight improvement in the model accounts for sampling error in the particular sample of BGs selected for the NSAF sample. Let $R_2 = \frac{T_f}{\hat{T}_f}$, where the denominator is the estimated number of nontelephone households from the 1990 Census data using the NSAF sample BGs. Since R_2 is based on CPS data, sample sizes were too small to define classes within states for the adjustment. One such adjustment was made for each site.

The model as formulated thus far is static; it assumes that the composition of nontelephone households in the BGs in 1990 is identical to that in 1997. Other analysis suggests this may not be valid. To address this, the model was extended to account for changes between 1990 and 1997 by assuming that one covered nontelephone household from 1990 can be exchanged with r uncovered nontelephone households in 1997.

The special 1997 CPS tabulations, when combined with estimates from the 1990 Census files, were used to estimate r. For each site, the 1997 CPS estimates of the percentage of nontelephone households in the BGs in the sampling frame were obtained and the estimate of r for a site was formulated as

$$r = \frac{\frac{T_f}{T_.}}{\frac{\hat{T}_{f,97}}{\hat{T}_.,97}},$$

where the denominator of r is the percentage of nontelephone households in the BGs in the sampling frame based on the 1997 CPS estimates. For example, suppose that in a particular state 92 percent of the nontelephone households were in the sampling frame in 1990, but the CPS estimate for this percent was only 88 percent for 1997. The value of r would be 1.045 (92/88). Since the CPS estimates of the percent in the sampling frame are based on a limited sample size,[13] r is estimated only at the overall site level.

The area undercoverage adjustment is then the product of R_{1k}, R_2, and r, where the subscript k indicates that this factor is computed separately for each poverty class k with the site. The adjusted weight for the area sample respondents is

$$CAW_{area,i,k} = NRW_{area,i,k} \cdot R_{1k} \cdot R_2 \cdot r .$$

The product of the last three factors is the undercoverage adjustment factor that indicates the size of the adjustment associated with the planned undercoverage of the area sample of nontelephone households. In Michigan, the average undercoverage adjustment factor (the last three factors on the right-hand side above) was 1.46. Thus, the sum of the weights of the residential eligible population (55,239 in table A-2, row 4.3) was multiplied by 1.46 (which is an average since R_{1k}

[13] We examined the coefficients of variation (CV) for the overall state estimates of the denominator of r (the numerator is based on Census data with no sampling error) and found that the CVs are less than 10 percent, even in the smallest states. Thus, the estimates of r should be reliable at the state level.

varies by poverty class) and resulted in an undercoverage adjusted sum of weights of 80,484 (row 5.2).

The next step in the weighting was the poststratification of the household weights to known population control totals. This step is discussed in the next chapter with the other adjustments of this type. Since the weights for the child and adult extended interviews are based on the poststratified household weight, we will call the poststratified household weight FHW and proceed with the development of the weights for extended interviews.

2.2 Extended Interview Initial and Nonresponse Adjusted Weights

As mentioned above, the poststratified household weight is the starting point in the process of producing weights for children, adults, and families. The initial extended level weights (referred to as the initial weights rather than base weights because they include all the adjustments made at the household level) are the product of the poststratified household weight and the probability of selecting the unit from the household for each group. These weights are then adjusted for nonresponse by site, separately for the RDD and area samples. The final step is the poststratification of these weights to external control totals, and this is discussed in the next chapter. Since the formulas are the same for each site, the subscripts representing sites are not shown. The procedures for creating the initial and nonresponse adjusted extended interview weights are described below.

2.2.1 Child Weights

The initial weight for a child interview is the product of the household weight and the probability of selecting the child within the household. The child weight depends on the age of the child; samples were selected independently for children under six and for children ages six to seventeen. If only one child in a household was under six, the child was selected with certainty. The same is true when there was only one child ages six to seventeen. In general, the probability of selecting a child within a household is the reciprocal of the number of children under six (or ages six to seventeen). The initial weight, IW_i, for child i is

$$IW_i = FHW_i \cdot f_i ,$$

where

FHW_i = the poststratified household weight of the household with child i; and
f_i = number of children under six (or six to seventeen) in the household with child i.

If data collected in the extended interview showed the sampled child was ineligible (e.g., the age reported in the screening interview was incorrect), the child weight was set to zero.

Table A-3 shows the initial weights and some of the sample sizes for each site. For example, in Michigan there were 2,583 sampled children and the sum of the initial weights for these children

was 2,649,303. The table also shows these numbers by whether they were a part of the RDD or area sample.

Since the eligibility is known for all sampled children, the extended level nonresponse adjustment using a weighting class adjustment is

$$NR_c = \frac{\sum_{NR} w_{ci} + \sum_{R} w_{ci}}{\sum_{R} w_{ci}},$$

where

w_{ci} = the initial child weight assigned for child i in adjustment cell c;
NR = eligible nonrespondents; and
R = eligible respondents.

In Michigan, the average adjustment factor was

$$NR_{MI} = \frac{416{,}732 + 2{,}232{,}571}{2{,}232{,}571} = 1.19,$$

where these numbers are given in rows 2.1.a and 2.1.b of table A-3. This is just the overall adjustment factor, since the adjustments were actually done on a cell by cell bases using the cells defined below.

To form the cells for the child interview nonresponse adjustment, variables available from the screening interview were used. The cells formed were

- Sample type (two cells — area and RDD);
- Age (two cells — under six years old and older than six);
- Poverty level from the screener (three cells — under 200 percent, above 200 percent, and unknown); and
- Metro status (three cells — in a central city, outside a central city but in a metropolitan area, and outside a metropolitan area).

When cells were too small or resulted in large adjustments, they were collapsed to form larger cells using the guidelines discussed in chapter 1. The collapsing was across categories of metro status first, then across poverty level and age. Cells were never collapsed across sample type.

The child weight was then adjusted for multiple residential telephone lines in a household. If a household had more than one telephone number used for residential purposes, the household could be sampled on any of these numbers. To account for the multiple chances of selecting the household, the weight was divided by two if the household had more than one residential

telephone number.[14] The nonresponse adjusted child weight that includes the adjustment for multiple telephone lines is

$$CIW_{ci} = IW_{ci} \cdot NR_c \cdot l_i ,$$

where l_i is 0.5 if the household has more than one residential line and is one otherwise. The further adjustment of these weights is discussed in the next chapter.

2.2.2 Adult Weights

As discussed in Chapter 1, there are three adult weights. One is used to estimate characteristics based on questions about the respondent and their spouse/partner and is called the APW. The second is used for questions asked only about a randomly selected adult in the pair and is called the RAW. The third set of adult weights is the ACW and is used for a small number of questions asked of only "Option B" adults without children. Each of these is discussed below.

The initial weights for the extended adult interview are developed in the same manner as those of the children (i.e., by multiplying the final household weight by the reciprocal of the probability of selecting the adult within the household). However, there are some additional complexities for the extended adult interviews. All the children in a household were enumerated during the screening interview and the probability of selection of each child could be determined for all households that completed the screening interview. For adults, this was not possible because the adults in the household were not enumerated until the end of the first extended interview in the household. As a result, in households with children, the number of "straggler B" adults in the household was not known unless an extended interview was completed in the household. Furthermore, for adult only households, the relationships of persons within the household needed to assign the probability of selection were not known unless an extended interview was completed. To compensate for this, a household nonresponse adjustment was done prior to the assignment of the probabilities of selection.

In addition, a second household adjustment was needed for "straggler B" adults, as mentioned in Section 2.2. Both adjustments are described below.

The nonresponse adjustment for households that were sampled for at least one extended interview, but did not complete any, uses the same procedures described above for the child nonresponse adjustment. The screening level variables were used because all the households completed the screening interview. The variables used to form the nonresponse cells were:

- Sample type (2 cells — area and RDD);
- Age (2 cells — under 6 years old and older than 6);

[14] The weight could be modified by the reciprocal of the number of residential telephone lines in the household, but the adjustment by two is thought to reduce the mean square error. Table A-3 shows that about 90 percent of the household had only one telephone (see row 3.4). Of those with multiple lines, about 80 percent had only one additional line. See Massey and Botman (1988) for a discussion of this issue.

- Poverty level from the screener (3 cells — under 200%, above 200%, and unknown); and
- Metro status (3 cells — in a central city, outside a central city but in a metropolitan area, and outside a metropolitan area).

When cells were too small or resulted in large adjustments, they were collapsed to form larger cells using the guidelines discussed in chapter 1. The collapsing was across categories of metro status first, then across poverty level, age. Cells were never collapsed across sample type.

This household level nonresponse adjustment can be written as

$$NR_c = \frac{\sum_{NR} w_{ci} + \sum_{R} w_{ci}}{\sum_{R} w_{ci}},$$

where

w_{ci} = the poststratified weight for household i in adjustment cell c;
NR = eligible nonrespondents (completed no extended interviews); and
R = eligible respondents (at least one extended interview completed in household).

Table A-4 shows that in Michigan the sum of the weights of the nonrespondent households (511,465 in row 1.3) was redistributed to the respondents (2,535,429 in row 1.2), so that the nonresponse adjusted number of households was 3,046,894 (row 2).

The second household adjustment is for subsampling "straggler B" adults for the "option B" path. These adults were only sampled in households with children if the telephone number was assigned to the "B" path. The adjustment is

$$A_s = \frac{\sum_{i \in hhkids} w_i}{\sum_{i \in hhkids} w_i \cdot d_i(B)}$$

and is at the household level but applied only to "straggler B" adults. The weights of the MKA adults in these households were not adjusted since these were not subject to the subsampling.

Table A-4 shows the components of this adjustment for Michigan and the other sites. In Michigan, the sum of the weights for households with children that were subsampled for path "B" was 732,171 (row 2.1) and for the households with children that were not subsampled for path "B" it was 587,745 (row 2.2). Thus, the adjustment for Michigan was

$$A_{s,MI} = \frac{732{,}171 + 587{,}745}{732{,}171} = 1.80\,.$$

The last step in creating the adult initial weight was to reflect the probability of selecting adults within the household. The probability of selection of MKAs and their spouse/partners is determined by the probability of selection of their children. If all children in a household have

the same MKA/spouse, then the probability of selecting the MKA/spouse is 1.0. For other adults under 65 in households with children, ("straggler B" adults), the probability is determined by the number of such adults in the household. In households without children, the probability of selection is determined in a somewhat more complex fashion, depending on the number of adults in the household.

The initial weight for the APW can be written as a product of the factors described above. For sampled MKAs in households with children and for adults sampled from households without children, the initial weight is

$$IAW_{ij} = FHW_i \cdot NR_c \cdot F_{adlt} ,$$

where the subscripts refer to household i and adult j in household i, NR_c is the nonresponse adjustment for households without a completed extended interview, and F_{adlt} is the reciprocal of the within-household probability of sampling the adult.[15] For sampled "straggler B" adults in households with children, the initial weight have the additional household level nonresponse adjustment. Their weights are

$$IAW_{ij} = FHW_i \cdot NR_c \cdot A_s \cdot F_{adlt} .$$

The RAW initial weight is similar to the APW in that it must account for selecting one of the adults within a pair to respond to questions about the specific adult. This only occurred in households with children. The RAW initial weight is

$$IRW_{ij} = IW_{ij} \cdot d_{ij} \cdot B_{ij}$$

where d_{ij} is an indicator variable that is 1 if the items were asked for adult j in household i and is zero otherwise, and B_{ij} is the ratio of the number of adults that could have been sampled for these questions (for couples in households with children this is 2 and for other situations it is 1).

For both the APW and the RAW, the nonresponse adjustment follows exactly the same procedure as used for the child weights. The cells for the weighting class adjustment were the same as for children, except one variable was added. This variable classified adults in households with children as either the MKA or a "straggler B." Collapsing cells was done so that the metro status cells were combined first, then poverty level, MKA/"straggler B" category, type of household, and sample type.

In Michigan, the sum of the weights of the nonrespondents for the APW (199,043 shown in table A-4, row 4.1b) was distributed to the 5,725,656 responding adults (row 4.1.a) so that the nonresponse adjusted total number of adult respondents was 5,924,699 (row 4.2.a).

The nonresponse adjusted APW, denoted *NAW*, is

[15] When an adult was sampled, the spouse/partner was automatically sampled. Thus, if there are two adults in the household and they are married, then $F_{adlt} = 1$; if they are not related or partners, then $F_{adlt} = 2$.

$$NAW_{cij} = IAW_{cij} \cdot NR_c \cdot l_i ,$$

where the nonresponse adjustment was computed using data for all sampled records with positive initial weights (IAW_{cij}). Similarly, the nonresponse adjusted RAW, denoted *NRW*, is

$$NRW_{cij} = IRW_{cij} \cdot NR_c \cdot l_i ,$$

where the nonresponse adjustment was computed using records with positive initial weights ($IARW_{cij}$) and l_i is the adjustment for multiple telephone lines. The adjustment for multiple telephone lines in Michigan decreased the sum of the weights for the APW by about 6 percent (rows 5.1 and 5.3 in table A-4). For both the APW and the RAW, the nonresponse adjusted weight was zero if the adult was not eligible (e.g., not between 18 and 64 years old).

The ACW is used to produce estimates of adults without children in the household for the few questions asked only of the 18- to 64-year-old adult respondent. These adults were not randomly selected when there were children in the household. As a result, in households with children the spouse/partner of the MKA (most often the male spouse/partner) had no chance of being asked these items. Thus, for these items, estimates for adults with children cannot be produced from the survey. On the other hand, adults without children were randomly selected to be the respondent, so they all had a chance of being asked the questions. The ACW is used to produce estimates from these adults for those items.

Rather than beginning at the household level and going through the whole weighting process again, the ACW was derived from the final poststratified APW. The procedures for creating the poststratified APW are discussed in the next chapter, so here we will simply refer to this weight as FACW. The data set with APW weights was subset to retain only adults in households without children and "straggler B" adults. The final ACW was set to twice the final APW if the person had a spouse/partner. If the adult did not have a spouse/partner, then the ACW was set to FACW. The final ACW is

$$FACW_j = FACW_j \cdot AC_j ,$$

where AC_j is 2 if the person had a spouse/partner and one otherwise.[16]

2.2.3 Family Weights

The last extended interview analysis group is the family. The composition of the family for each sampled household member was determined in the first extended interview when the household composition questions were asked. The definition determines how family weights were prepared, since the weights must be consistent with the way questions were asked of persons in the interviews. In fact, the data collected about relationships of persons in the households allow two different definitions of families to be constructed. These are discussed briefly below.

[16] The ACW was not further poststratified to control totals because the final APW was already subject to this adjustment.

The definition of a family used during data collection to determine which persons were asked about their family is called the "social" family. It is similar to what CPS defines as a family group (e.g., see Rawlings (1994) for definitions of CPS terms), but there are two major differences. One difference is that unmarried partners are in the same family in the "social" family definition; the CPS does not recognize these partners in their definition of a family or a family group. The second major difference is the treatment of related individuals who live in the same household. For example, consider a couple, their unmarried daughter, and her child, all living in the same household. The "social" family includes all these household members as one family, while the CPS defines two family groups in this household.

The terminology is confusing. CPS uses the term family, but this is not equivalent to the "social" family definition in the NSAF. The CPS definition of a family group is closer to the NSAF definition of a family. In the CPS, a household can have only one primary family even though it can have multiple family groups. For example, suppose a household consists of two married couples, where members of one couple are not related to members of the other couple. In the CPS, the couple containing the householder is a family, and it is also a family group. The other couple is not a family, but is a family group. The "social" family defines these two couples as two families.

The flags were created on the data set to identify the "social" family definition for each household with a completed extended interview. Flags were also created to identify families using the CPS definition. The weighting of the family units using the CPS definition is virtually identical to that used for the "social" family so it is not discussed further.

The procedure used to create the family weights used the poststratified person-level weights as an initial weight and then adjusted them to account for the possibility that a household could have more than one family.[17] Since the person weights were already adjusted to control totals at the site level, this ensured a high degree of consistency between the person and family weights.

To implement the procedure, three types of families were defined: families in households with children where the sampled family is created around the sampled child; families in households with children where the sampled family is created around a sampled "straggler B" adult; and families in households without children. The procedures for weighting all three types of families are similar.

The family weights were computed by first determining the within household probability of sampling the family. The reciprocal of this probability is the within-household family weight. In the most common situation, all members of a household were in the same family, and the within-household family weight was one. However, if a household contained more than one family or a household without children contained a family and some unrelated individuals, then the family might not have been sampled with certainty. In this case, the within-household family

[17] We originally planned a different method of weighting families. The original method had three steps: (1) compute a probability of sampling the family within household, (2) compute a family nonresponse adjustment, and (3) poststratify the estimates to CPS-based control totals for families. This procedure was not implemented because the CPS sample sizes by site were not large enough to use as controls and those control totals were adjusted for the 1990 Census undercount, making the family controls inconsistent with the person control.

weight was computed based on whether or not the sampled members of the household were in the family.

The initial family weight, APSW, is

$$APSW_j = \frac{FAPW_j}{F_{adlt}},$$

where FAPW is the poststratified APW and F_{adlt} is the reciprocal of the within-household probability of sampling the adult. When more than one sampled adult was a member of the same family (e.g., a married couple), the weights of the sampled members were averaged to produce the family weight.

There were two exceptions to these rules for households with children in which the sampled family was created around the sampled children rather than the adults. If an MKA and a "straggler B" were in the same family, then only the weight of the MKA was used. If no eligible MKA was in the family (emancipated minors living without any persons over 17 years old), then the final child weight was used.

The final family weight, *FFW*, adjusts the initial family weight by the inverse of the probability of sampling the family. It is

$$FFW_i = APSW_i \cdot F_{fam},$$

where F_{fam} is the within-household family weight.

2.3 Summary

The initial steps of weighting for each of the analysis groups were given above. The first step was the creation of household weights that reflected the probability of selection of the household from either the area or RDD samples. These weights were then adjusted for nonresponse. The RDD household weights were further adjusted for subsampling by income and presence of children, while the area sample weights were adjusted for the exclusion of BGs from the area sampling frame. The extended interview initial weights and the nonresponse adjustment were then presented. The extended interview weights for children, adults, and families were described.

The next chapter describes the final steps of weighting, which is the adjustment of the weights to external control totals. It also discusses how the weights were modified if they took on unusual values.

Chapter 3
Adjustments to Control Totals

In this chapter, further adjustments of the nonresponse weights to population control totals and adjustments for outlying weights are described. We begin by discussing the poststratification of the household weights. The subsequent section deals with adjusting the child and adult weights to control totals and handling outlying weights. The result of these procedures was the final weights used to prepare the estimates in the *Snapshots of America's Families*.

3.1 Household Poststratification

The last step in the weighting process modifies the nonresponse adjusted weights so that they are equal to known population control totals. Deville and Särndal (1992) gives the overall theory of adjusting to known totals from auxiliary sources. In theory, the estimates should differ from the population values only as a result of sampling error. In practice, other error sources such as residual nonresponse and coverage errors still may have an important effect on the accuracy of the estimates. Thus, often the primary goal of poststratification is to reduce these biases and reduction of sampling errors is a secondary goal. These were the main reasons for introducing poststratification in the 1997 NSAF.

Since control totals are available at both the household and at the extended interview level, some choices are available. Greater importance was assigned to poststratification at the extended interview level than at the household level for two reasons. First, some variables are just more suitable for poststratification of persons than of households. For example, age and sex are person-level variables and not suitable at the household level. Second, variables that could be used for either household or person poststratification were generally reserved for the person level because we believed that the person estimates would benefit more from this. For example, home ownership could be used at either the household or person level. Person estimates benefit more from applying poststratification at the end of the weighting process at the person level because additional adjustments for sampling and nonresponse at the extended level would dilute any household adjustments.

The household weights were poststratified at the study-area level to the number of households by type of household (households with children and adult-only households with at least one person under age 65).[18] The poststratification adjustment (see chapter 7 in Särndal, Swensson, and Wretman (1992) for a discussion of poststratification and the related topic of the use of conditional inference) was

[18]We considered poststratification by other variables including MSA status, poverty status, and household size but decided against these for various reasons. MSA status was not used because it was not known precisely for each residential household in the sample. Poverty status was rejected because it was subject to different measurement methods that could affect its value. Household size was not used, but it was observed afterwards that the average household size for the survey was quite close to that of the CPS. (Report No. 1 *An Overview of the NSAF Sample Design, Data Collection Techniques, and Estimation Methods*)

We had planned to poststratify to the number of households by telephone status and household type, but decided to drop the telephone status cells after some examination of the outcomes of the poststratification. We found large discrepancies between the control totals and the estimates of the percentage of nontelephone households in the survey. As discussed in the *1997 NSAF In-Person Survey Methods* report, we concluded that the control total estimates of the percentage of households with telephones were not consistent with the experience in the NSAF and poststratification to this total would result in biases in the estimates.

$$PS_c = \frac{CT_c}{\sum w_{ci}},$$

where c is an adjustment class defined by household type,

CT_c = poststratification control totals for cell c;
w_{ci} = weight (NRW_{RDD} or CAW_{area}) for household i.

The final household-level weight, FHW, is

$$FHW_{i(c)} = PS_c \cdot w_{ci},$$

where the terms are defined above.

The computations of the poststratification adjustment for Michigan should help clarify these operations. Table A-5 shows the sum of the weights for households with children (1,305,993 in row 1.3.b) and with adults only (1,371,977 in row 1.3.a). The control totals for Michigan are 1,370,324 for households with children (row 4.3.b) and 1,676,570 for households with adults only (row 4.3.a). In this case, the poststratification adjustment for households with children in Michigan was

$$PS_{MI,chld} = \frac{1{,}370{,}324}{1{,}305{,}993} = 1.05.$$

The adjustment for adults was done the same way.

The control totals for the number of households were constructed based on data from two different sources. The primary source was the Census Bureau's official published estimates of the total number of households at the state level for July 1996. These estimates are decennial census level and are not adjusted for census undercount (see part 3 of this series for a discussion of issues related to estimation from the 1997 NSAF at census undercount adjusted level). The 1996 number was projected to a 1997 estimate of the total number of households. Once we estimated the total number of households, the totals by type of household were created by estimating the percentage of households in each type and multiplying that percentage by the grand total. Household type is very stable over time, as shown in the CPS estimates in table 3-1. Since these estimates exhibit no general or consistent trends over the 1991–1996 time period, the 1990 Census as opposed to a more current CPS estimate was chosen to be the source of the percentage distribution by household type.

The poststratified household estimates play a key role in the estimation strategy for the 1997 NSAF, even though these weights are not used to produce direct estimates of households from the survey. The poststratified weights are so important because they are included as a component in the initial weights for all of the extended interview weights. This was discussed in chapter 2. In the next section, the development of the final extended interview weights is described.

Table 3-1.
Estimated Number and Percent of Households by Type and Region, 1991–1996

Characteristic	1991	1992	1993	1994	1995	1996
Total households	94,312	95,669	96,391	97,107	98,990	99,627
Northeast	19,271	19,314	19,437	19,470	19,593	19,695
Midwest	23,223	23,327	23,307	23,385	23,683	23,707
South	32,312	33,073	33,392	33,904	34,766	35,143
West	19,506	19,955	20,255	20,347	20,948	21,082
Percent of households with householder 65+	21.76%	21.87%	21.68%	21.43%	21.58%	21.57%
Percent of households with related children under 18	36.58%	36.44%	36.82%	37.53%	37.15%	36.85%
Percent of households without related children under 18	63.42%	63.56%	63.18%	62.47%	62.85%	63.15%
Percent of households with members 65+						
Northeast	13.34%	13.16%	13.43%	13.15%	13.23%	13.22%
Midwest	12.28%	12.04%	11.78%	11.61%	11.95%	12.22%
South	11.96%	12.45%	12.24%	11.94%	11.64%	12.17%
West	10.78%	10.89%	11.20%	10.77%	10.91%	10.33%
Percent of households with married couple families						
Northeast	54.42%	53.84%	53.46%	53.05%	52.75%	51.77%
Midwest	55.59%	54.57%	55.03%	54.21%	54.20%	54.54%
South	56.03%	55.47%	56.01%	55.35%	55.23%	54.36%
West	54.57%	55.03%	55.54%	56.03%	54.83%	53.79%
Percent of households with other families						
Northeast	15.38%	15.68%	15.88%	16.06%	16.06%	16.82%
Midwest	13.82%	14.75%	15.01%	15.10%	14.47%	14.44%
South	16.01%	16.12%	15.98%	16.71%	16.31%	17.06%
West	14.50%	14.61%	15.06%	14.72%	15.29%	15.64%
Percent of households with nonfamilies						
Northeast	30.20%	30.48%	30.65%	30.90%	31.19%	31.41%
Midwest	30.59%	30.68%	29.96%	30.69%	31.32%	31.02%
South	27.96%	28.40%	28.01%	27.94%	28.46%	28.58%
West	30.92%	30.36%	29.41%	29.25%	29.89%	30.58%

Source: Current Population Survey

3.2 Child and Adult Adjustments to Control Totals

In this section, the adjustment of the child and adult interview weights to control totals for the 1997 NSAF are described. We begin by discussing the source of the person-level control totals. Next, we describe the raking procedure used in more detail than given in the first chapter. We then present the child and adult procedures, including any adjustments that had to be made to reduce the contributions due to weights that were very large.

3.2.1 Control Totals

Two sources of control totals were considered for child and adult adjustments: Census Bureau independent controls by age/race/ethnicity/sex and CPS estimates of the number of persons by various characteristics. The Census Bureau produces estimates of the total population and

institutional population at the county level by single year of age, by race/ethnicity, and by sex http://www.census.gov/population/www/estimates/countypop.html. At the time of processing the data, the latest estimates of this type were for July 1, 1996. By subtracting the institutional population from the total, we produced control totals at the site level for both the child and adult poststratification. Since the population estimates were for 1996, the totals were extrapolated to March 1997. The race/ethnicity categories are Hispanic, black/non-Hispanic, and other. Age was categorized into five year intervals. These control totals were suitable for use as control totals because they were not subject to sampling errors, were current, were well-defined, and were at the state level. For a discussion of the quality of these estimates, see Long (1993).

The second source of control totals considered was estimates from CPS. The characteristics used as controls were home ownership (own or rent) and educational attainment. These data items are collected the first time households are interviewed in CPS. As we discussed previously, other CPS variables such as poverty were rejected for this type of adjustment. Home ownership and educational attainment are better suited to be control totals because they are not key outcome variables, they are stable over time, and they can be averaged over several months of CPS data collection within the same year. The averaging is possible because these data are obtained monthly.

The reliability of the CPS estimates was improved by creating averages of CPS estimates across three months (July 1996, November 1996, and March 1997). Table 3-2 shows approximate CPS and 1997 NSAF household sample sizes by state, where the CPS sample sizes are for three months of data collection. The table shows that CPS sample sizes are considerably larger than those for the NSAF in a few of the larger states. For the four largest states (California, Florida, New York, and Texas) the CPS sample sizes were large enough to be used at the site level. For the other sites, the CPS sample size were too small to be used as controls, so the estimates were aggregated to the balance of each of the four Census regions. Table 3-3 shows these sample sizes.

Educational attainment was used as a control total only for adults. Educational attainment for adults is reported the same way for each adult in both the CPS and the 1997 NSAF. Thus, the variable is appropriate for use as a control total for the 1997 NSAF. On the other hand, the same classification of educational attainment cannot be developed for children in both the 1997 NSAF and CPS because comparable data are not obtained.[19]

Since the age, race/ethnicity, and sex control totals and the CPS estimates do not sum to the same number of adults or children, the CPS estimates had to be converted to the same basis to be used in the adjustment. The percentage of persons in each category (home ownership and educational attainment) were estimated from the CPS and multiplied by the control total for each group (children and adults, separately) to produce the totals used in the adjustment. Thus, the aggregate of the age, race/ethnicity, and sex control totals and the aggregate of the modified CPS estimates were made to be the same.

[19] For the child level, we considered using the highest level of education of all adults in a household since that can be obtained from CPS. However, this variable is not available from the 1997 NSAF because education is not collected for all adults in the household. Conversely, the MKA and the MKA's education for each sampled child can be obtained in the NSAF, but cannot be obtained from the CPS.

Table 3-2.
Approximate CPS and NSAF Household Sample Size, by State

State	Total		Less Than 200% Poverty		Household with Children		Less Than 200% Poverty with Children	
	CPS	NSAF	CPS	NSAF	CPS	NSAF	CPS	NSAF
Total U.S.	139,860	40,978	48,255	17,661	51,567	25,116	19,179	12,551
Alabama	2,001	2,429	903	1,203	774	1,549	333	878
California	10,875	2,308	3,870	1,154	4,356	1,401	1,821	827
Colorado	1,872	2,730	465	1,124	711	1,606	195	756
Florida	7,011	2,172	2,640	1,037	2,337	1,469	957	801
Massachusetts	3,339	2,844	978	1,043	1,098	1,776	336	747
Michigan	4,881	2,549	1,467	1,030	1,806	1,585	561	728
Minnesota	2,007	3,130	591	1,197	732	1,770	189	775
Mississippi	1,773	2,677	843	1,132	738	1,361	423	824
New Jersey	4,161	3,036	1,155	1,027	1,569	1,827	363	702
New York	9,207	2,416	3,429	1,150	3,300	1,611	1,350	869
Texas	6,564	2,303	2,424	1,211	2,721	1,609	1,125	986
Washington	1,827	3,219	510	1,302	669	1,857	204	877
Wisconsin	2,109	2,978	513	1,154	792	1,821	219	793

Source: The estimated number of households in the CPS was computed as the number from the March 1996 file multiplied by three (after excluding Hispanic oversample).

Table 3-3.
Approximate CPS and NSAF Household Sample Size, for Large States and Balance of Regions

CPS Sample Households	Total		Less Than 200% Poverty		Household with Children		Less Than 200% Poverty with Children	
	CPS	NSAF	CPS	NSAF	CPS	NSAF	CPS	NSAF
Total U.S.	139,860	40,978	48,255	17,661	51,567	25,116	19,179	12,551
California	10,875	2,308	3,870	1,154	4,356	1,401	1,821	827
Florida	7,011	2,172	2,640	1,037	2,337	1,469	957	801
New York	9,207	2,416	3,429	1,150	3,300	1,611	1,350	869
Texas	6,564	2,303	2,424	1,211	2,721	1,609	1,125	986
Northeast (except NY)	21,177	5,880	6,423	2,070	7,496	3,603	2,196	1,449
Midwest	33,351	8,567	10,497	3,381	12,021	5,176	3,882	2,296
South (except FL and TX)	30,093	5,106	11,778	2,335	10,923	2,910	4,695	1,702
West (except CA)	21,582	5,949	7,194	2,426	8,412	3,463	3,153	1,633

Source: The estimated number of households in the CPS was computed as the number from the March 1996 file multiplied by three (after excluding Hispanic oversample).

3.2.2 Raking Procedure

The adjustment of the household weights to control totals was accomplished using poststratification. At the person level, a raking procedure was used so that the constraints from more auxiliary information could be imposed. In particular, both age/race/sex totals and CPS homeownership and educational attainment (for adults) totals can be controls with raking but not with poststratification. This was briefly mentioned in chapter 1, but is more thoroughly discussed here.

Raking is a commonly used estimation procedure in which estimates are controlled to marginal population totals. It can be thought of as a multidimensional poststratification procedure, because the weights are basically poststratified to one set of control totals (a dimension), then these adjusted weights are poststratified to another dimension. The procedure continues until all dimensions are adjusted. The process is then iterated so that the control totals for all the dimensions are simultaneously satisfied (at least within a specified tolerance).

Oh and Scheuren (1987), Brackstone and Rao (1979), and Deville and Särndal (1992) discuss raking. The raked weights can be expressed as a product of factors that depend only on the marginal values of the control variables. Suppose, for example, weights are raked to control totals with two dimensions, say, variable *A* and variable *B*. At the conclusion of the iterations, the raked weight can be written as a product of the original weight and two factors — one depending on the level of *A* and the other depending on the level of *B*.

In this sense, the weights are determined by the marginal distributions of the control variables. As a result, the sample sizes of the marginal distributions are the important determinants of the stability of the weighting procedure, not the cells formed by the crossing of the variables. This means that deficient cells (cells with small sample sizes) are defined by looking at the sample sizes of the margins. Furthermore, this permits the use of more variables or control totals with raking than is possible with poststratification.

3.2.3 Child Interview Raking and Trimming

For children, two dimensions were used in the raking. The first dimension has categories defined by the cross-classification of age, sex, and race. The second dimension has categories of home ownership. In this case, the marginal distribution by home ownership had more than adequate sample sizes. However, the sample sizes by categories of age/sex/race had to be examined and collapsed if they were too small (for raking 20 was the minimum sample size). The cells were combined so that no combining across age was ever done. Collapsing to form larger cells of age/sex/race was done by first collapsing over sex and then over race, as needed.

As mentioned above, the raking factor can be represented as a multiplicative factor that depends only on the levels of the two dimensions. This raking factor is RF_{cd}, where the two subscripts denote the levels of the two raking dimensions. Using this notation, the final child extended weight, *FCW*, is

$$FCW_{i,cd} = CIW_{i,cd} \cdot RF_{cd},$$

where the *CIW* is the nonresponse adjusted child weight described in the previous chapter, *c* denotes the age/sex/race category, and *d* denotes the home ownership category. Only completed extended child interviews are assigned positive weights.

The results of the raking procedure are given in table A-3. For example, in Michigan the sum of the weights prior to poststratification was 246,0936 (row 3.3) and after the adjustment it was 2,542,812 (row 4.1). The average raking adjustment across all children was 1.03 (row 4.2.a), with a minimum raking adjustment of 0.70 (row 4.2.b) and a maximum factor of 1.70 (row 4.2.c). The mean raking adjustment factors for cells defined by race, sex, and home ownership are given in rows 4.3 to 4.5.

After raking, the weights were examined to determine if there were any outliers. Outliers are weights that are much larger than typical. Such weights can have an important effect of the estimates, and especially on the variances of the estimates (see Kish (1992) for a discussion of the effect of large weights on the estimates and their variances). The plan was to reduce the effect of these outliers by trimming the weights (see Potter (1988) and Potter (1990) for a discussion of trimming procedures). Trimming is generally done by reducing the weights of the outliers, and sometimes the weight is redistributed to other cases to keep the total sum of weights constant.

The procedure used to identify outliers was to examine any weights that were more than four times the mean of the weights. This examination was done separately within site by class, where the classes were the sample type (area or RDD) and poverty level from the screener interview (under 200 percent poverty, above 200 percent poverty, and unknown). The classes were used because of the large differences in the probabilities of selection by these variables. If an outlier was identified and its weight was much larger than the average, the weight was reduced; often it was trimmed to equal four times the mean weight in the class.

Since no outliers were identified in Michigan, we use the balance of the U.S. as an example of this procedure. Table A-3, row 5.4 shows that one outlier was trimmed in the balance of the U.S. It was reduced by 11,878 (row 5.5) as a result of the trimming. As can be seen from table A-3, weights across the sites were trimmed only if they were very large.[20]

After the outlying weights were trimmed, the sum of the weights were less than the control totals. The raking procedure described above was applied again, but this time the trimmed weights were used instead of the nonresponse adjusted weights. This restored the estimates to the proper totals. Table A-3, row 6.1, shows the restored weights. In Florida, where one weight was trimmed, the sum of the weights after this second raking is shown to be equal to the sum of the weights after the first raking. These are the final weights used to produce child estimates in the *Snapshots.*

[20] A conservative approach with minimal trimming was appropriate because many of the large weights were the result of having a large number of eligible children in the household. More aggressive trimming would have reduced the effect of children in large households in the estimates.

3.2.4 Adult Interview Raking and Trimming

The raking adjustment for adults is very similar to that of children. The first dimension was defined by the cross-classification of age, sex, and race/ethnicity. The second dimension was by home ownership. For adults, a third dimension was defined by educational attainment (less than 12^{th} grade, high school diploma but not college graduate, college graduate or higher). Since the population of adults is identical for the APW and the RAW, the control totals were identical for these two weights.

As with the child weighting procedure, the sample sizes by categories of age/sex/race were examined and collapsed if they were too small. Cells were never combined across age. Collapsing to form larger cells of age/sex/race was done by first collapsing over sex and then over race, as needed. The sample size for the other two dimensions were large enough so that no collapsing across levels of these variables was needed.

The raked weight is the product of the extended nonresponse adjustment and the extended raking adjustment. Using the notation given before for the raking factor, the raked APW is

$$FAPW_{i,cde} = NAW_{i,cde} \cdot RF_{cde} ,$$

where *NAW* is the nonresponse adjusted APW, and the additional subscript, *e*, denotes the educational attainment category. Positive raked weights are assigned for all completed adult interviews (one record for each adult and each spouse/partner with a completed interview).

Table A-4 shows the effect of the raking for the APW. In Michigan, the sum of the weights before raking is 5,573,790 (row 5.3) and after raking it is 5,813,743 (row 6.1). The mean raking adjustment is 1.04 (row 6.2.a), and the other raking factors are given in rows 6.2.b through 6.6.c.

The final *RAW* is written in the same fashion

$$FRAW_{i,cde} = NRW_{i,cde} \cdot RF_{cde} ,$$

where *NRW* is the nonresponse adjusted random pair weight. Again, only completed adult interviews (one record for each adult asked the questions about the randomly selected adult) have positive weights.

The procedure used to identify outliers and trim their weights was the same for adults and children. The general guideline used for trimming was to trim if a weight was more than four times the mean of the weights in the cell. The cells used to examine the weights for trimming were the same as for children, with the addition of the variable adult group (adult-only household, MKA, or "straggler B"). Once located, the same procedures were used to trim the weights as were used for the child weights.

No outliers for adults were found in Michigan, so we use California as an example. In California, one outlying weight for an adult was identified and trimmed (table A-4, row 7.2). The weight was reduced a total of 10,576 (row 7.3).

After the outlying weights were trimmed, the raking procedure was used again, but with the trimmed weights instead of the nonresponse adjusted weights. This procedure was done for both the APW and the RAW. These final adult weights were used to produce estimates of adults for the *Snapshots.*

3.3 Final Weights

The procedures in this chapter described how the nonresponse adjusted weights were further modified to produce weights that are consistent with control totals from auxiliary sources. The purpose of these weight adjustments was primarily to reduce any residual biases in the estimates and, at least to some extent, to reduce the sampling errors of the estimates. The estimation of the sampling errors is covered in part 2 of the estimation reports in this series.

References

Brackstone, G. J., and Rao, J. N. K. 1979. "An Investigation of Raking Ratio Estimation." *Sankhya* C (41). 97–114.

Brick, J. M., and Broene, P. 1997. "Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1995 National Household Education Survey." Working Paper 97–06, U.S. Dept. of Education, Office of Educational Research and Improvement.

Brick, J. M., and Kalton, G. 1996. "Handling Missing Data in Survey Research." *Statistical Methods in Medical Research* 5: 215–238.

Brick, J. M., Waksberg, J., Kulp, D., and Starer, A. 1995. "Bias in List-assisted Telephone Surveys." *Public Opinion Quarterly*. 59 (2): 218–235.

Deville, J. C. and Särndal, C. E. 1992. "Calibration Estimators in Survey Sampling." *Journal of the American Statistical Association* 87: 376–382.

Giesbrecht, L., Kulp, D., and Starer, A. 1996. "Estimating Coverage Bias in RDD Samples With Current Population Survey Data." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 503–508.

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

Holt, D., and Smith, T. 1979. "Poststratification." *Journal of Royal Statistical Society* A (142) 33–46.

Kalton, G., and Kaspryzk, D. 1986. "The Treatment of Missing Survey Data." *Survey Methodology* 12: 1–16.

Kass, G. 1980. "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Applied Statistics* 29: 119–127.

Kish, L. 1992. "Weighting for Unequal Pi." *Journal of Official Statistics* 8: 183–200.

Little, R. J. A. 1986. "Survey Nonresponse Adjustments." *International Statistical Review* 54: 139–17.

Long, J. F. 1993. "Postcensal Population Estimates: States, Counties, and Places." Statistical Policy Office Statistical Policy Working Paper 21. Washington, DC: Office of Management and Budget.

Maklan, D., and Waksberg, J. 1988. "Within-household Coverage in RDD Surveys." In *Telephone Survey Methodology*, edited by R. Groves et al. New York: John Wiley and Sons. 73–98.

Massey, J. T. and Botman, S. L. 1988. "Weighting Adjustments for Random Digit Dialed Surveys." In *Telephone Survey Methods*, edited by R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg. New York: John Wiley and Sons, 143-160.

Oh, H. L., and Scheuren, F. J. 1983. "Weighting Adjustments for Unit Nonresponse." In *Incomplete Data in Sample Surveys* 2, edited by W. G. Madow, I. Olkin, and D.B. Rubin. Academic Press, 143–184.

Oh, H. L., and Scheuren, F. J. 1987. "Modified Raking Ratio Estimation." *Survey Methodology* 13: 209–219.

Potter, F. 1988. "Survey of Procedures to Control Extreme Sampling Weights." *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 453–458.

Potter, F. 1990. "A Study of Procedures to Identify and Trim Extreme Sampling Weights" *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 225–230.

Rawlings, S. W. 1994. "Household and Family Characteristics: March 1993." *Current Population Reports*, U.S. Bureau of the Census, P20-477.

Särndal, C. E., Swensson, B., and Wretman, J. 1992. *Model Assisted Survey Sampling*, Springer-Verlag.

Shapiro, G., Battaglia, M., Camburn, D., Massey, J., and Tompkins, L. 1995. "Calling Local Telephone Company Business Offices to Determine the Residential Status of a Wide Class of Unresolved Telephoned Numbers in Random-digit Dialing Sample." Unpublished long version of paper in *Proceeding of the Survey Research Methods Section of the American Statistical Association*.

Thornberry, O. T. and Massey, J. T. 1988. "Trends in United States Telephone Coverage Across Time and Subgroups." In *Telephone Survey Methods*, edited by R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg, 51-72. New York: John Wiley and Sons.

Waksberg, J., Brick, J. M., Shapiro, G., Flores-Cervantes, I., and Bell, B. 1997. "Dual-frame RDD and Area Sample for Household Survey with Particular Focus on Low-income Population." *Proceedings of the Survey Research Methods Section of the American Statistical Association*. 713–718.

Appendix A

Detailed Weighting Steps by Site

Table A-1.
Household Weighting for the RDD Sample, by Study Area

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
1	Sample										
	1.1	Total telephone numbers in sample		20,000	30,000	27,000	33,270	34,000	44,000	30,000	30,000
		a.	Residential telephone numbers	10,244	14,508	11,972	14,760	15,398	21,270	13,680	12,850
		b.	Nonresidential telephone numbers	8,962	14,050	12,965	16,454	16,503	19,631	14,416	15,730
		c.	Unknown	794	1,442	2,063	2,056	2,099	3,099	1,904	1,420
2	Base weight										
	2.1	Sum of weights		2,822,800	3,043,100	23,738,500	3,068,300	11,730,800	4,713,200	7,039,700	754,200
3	Allocating unknown residential										
	3.1	Sum of weights before adjustment		1,480,770	1,517,125	11,103,859	1,423,846	5,549,693	2,383,959	3,350,419	334,507
		a.	Residential known eligibility	1,272,236	1,285,811	8,132,634	1,096,085	4,248,275	1,773,127	2,668,046	275,937
		b.	Residential unknown eligibility	173,602	185,832	2,393,192	265,145	1,064,398	505,276	542,057	47,112
		c.	Unknown residential allocated	34,962	45,482	578,032	62,617	237,021	105,555	140,315	11,458
	3.2	Sum of weights after adjustment		1,480,770	1,517,125	11,103,859	1,423,846	5,549,693	2,383,959	3,350,419	334,507
		a.	Residential known eligibility	1,272,236	1,285,811	8,132,634	1,096,085	4,248,275	1,773,127	2,668,046	275,937
		b.	Residential unknown eligibility	208,534	231,314	2,971,225	327,761	1,301,418	610,832	682,372	58,570
4	Adjusting for unknown eligibles										
	4.1	Sum of weights before adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,383,959	3,350,419	334,507
		a.	Res. elig. respondents	972,878	996,920	6,390,932	900,546	3,015,161	1,376,362	2,049,726	209,341
		b.	Res. elig. nonrespondents	26,252	19,273	234,747	13,232	75,905	37,491	56,318	4,626
		c.	Res. ineligible	273,106	269,619	1,506,955	181,605	1,157,209	359,274	562,003	61,970
		d.	Res. unknown (nonrespondents)	208,534	231,314	2,971,225	328,333	1,301,418	610,832	682,372	58,570
	4.2	Sum of weights after adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,383,959	3,350,419	334,507
		a.	Res. elig. respondents	1,134,198	1,178,895	8,756,598	1,174,364	3,957,580	1,855,182	2,583,046	254,394
		b.	Res. elig. nonrespondents	30,699	23,106	323,429	17,409	101,160	51,186	71,685	5,697
		c.	Res. ineligible	315,874	315,124	2,023,832	231,943	1,490,954	477,591	695,688	74,416
		d.	Res. unknown (nonrespondents)	0	0	0	0	0	0	0	0

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

					AL	BW	CA	CO	FL	MA	MI	ML
					Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
5	Adjusting for eligible nonrespondents											
	5.1	Sum of weights after adjustment			1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,383,959	3,350,419	334,507
		a.	Eligible respondents		1,164,897	1,202,001	9,080,027	1,191,773	4,058,739	1,906,368	2,654,731	260,091
		b.	Elig. nonrespondents		0	0	0	0	0	0	0	0
		c.	Ineligible		315,874	315,124	2,023,832	231,943	1,490,954	477,591	695,688	74,416
		d.	Unknown		0	0	0	0	0	0	0	0
6	Type of HHs before IK refielded											
	6.1	Sum of weights			1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,383,959	3,350,419	334,507
		a.	HH with children		548,996	562,253	4,374,274	542,708	1,745,383	811,879	1,257,952	111,955
		b.	HH with no children		615,900	639,748	4,705,753	649,066	2,313,356	1,094,489	1,396,779	148,136
		c.	Elderly HHs		315,874	315,124	2,023,832	231,943	1,490,954	477,591	695,688	74,416
7	Adjusting for refielded IKs											
	7.1	Sum of weights before adjustment								2,383,959	3,350,419	
		a.	IK respondent		N/A	N/A	N/A	N/A	N/A	161,262	156,758	N/A
		b.	IK nonrespondent		N/A	N/A	N/A	N/A	N/A	2,710	3,082	N/A
		c.	IK ineligible		N/A	N/A	N/A	N/A	N/A	23,521	24,671	N/A
		d.	IK residential unknown elig.		N/A	N/A	N/A	N/A	N/A	33,087	32,853	N/A
		e.	IK nonresidential		N/A	N/A	N/A	N/A	N/A	31,579	28,351	N/A
		f.	Other		N/A	N/A	N/A	N/A	N/A	2,131,800	3,104,703	N/A
	7.2	Adjusting for unknown eligibility								2,352,380	3,322,067	
		a.	IK respondent		N/A	N/A	N/A	N/A	N/A	189,409	184,778	N/A
		b.	IK nonrespondent		N/A	N/A	N/A	N/A	N/A	3,226	3,692	N/A
		c.	IK ineligible		N/A	N/A	N/A	N/A	N/A	27,945	28,894	N/A
		d.	IK residential unknown elig.		N/A	N/A	N/A	N/A	N/A	0	0	N/A
		e.	IK nonresidential		N/A	N/A	N/A	N/A	N/A	0	0	N/A
		f.	Other		N/A	N/A	N/A	N/A	N/A	2,131,800	3,104,703	N/A

Note: N/A – not available.

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
	7.3	Adjusting for eligible nonrespondents							2,324,435	3,293,173	
		a.	IK respondent	N/A	N/A	N/A	N/A	N/A	192,634	188,470	N/A
		b.	IK nonrespondent	N/A	N/A	N/A	N/A	N/A	0	0	N/A
		c.	IK ineligible	N/A	N/A	N/A	N/A	N/A	0	0	N/A
		d.	IK residential unknown elig.	N/A	N/A	N/A	N/A	N/A	0	0	N/A
		e.	IK nonresidential	N/A	N/A	N/A	N/A	N/A	0	0	N/A
		f.	Other	N/A	N/A	N/A	N/A	N/A	2,131,800	3,104,703	N/A
		g.	Losses after refield (among re-fielded IKs)	N/A	N/A	N/A	N/A	N/A	59,524	57,245	N/A
8	Type of HHs after IK refield										
	8.1	Sum of weights after 1K adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,324,435	3,293,173	334,507
		a.	HH with children	548,996	562,253	4,374,274	542,708	1,745,383	811,879	1,257,952	111,955
		b.	HH with no children	615,900	639,748	4,705,753	649,066	2,313,356	1,034,964	1,339,534	148,136
		c.	Elderly HHs	315,874	315,124	2,023,832	231,943	1,490,954	477,591	695,688	74,416
		d.	Losses after refield (among no chld HH)	N/A	N/A	N/A	N/A	N/A	59,524	57,245	N/A
9	Reflecting poverty subsampling										
	9.1	Sum of weights before adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,324,435	3,293,173	334,506
		a.	Chld. scrnd. as poor HH	195,785	135,617	1,515,568	139,586	538,072	173,324	325,701	31,125
		b.	Chld scrd as nonpoor HH retained	121,086	113,680	818,481	117,557	337,683	163,095	268,477	20,340
		c.	Chld scrnd as nonpoor HH not retained	203,601	288,546	1,749,965	260,121	758,628	428,606	589,876	54,308
		d.	Total child screened as nonpoor	324,687	402,226	2,568,446	377,677	1,096,311	591,701	858,352	74,648
		n.	Adlt scrnd. as poor HH	70,976	64,933	567,172	66,325	166,551	83,158	137,291	11,546
		o.	Adlt scrd as nonpoor HH retained	73,306	92,718	821,746	108,269	201,429	152,354	214,995	15,314
		p.	Adlt scrnd as nonpoor HH not retained	129,223	242,339	1,982,542	232,522	436,729	383,380	463,542	38,513
		q.	Total adult screened as nonpoor	202,529	335,057	2,804,288	340,791	638,158	535,733	678,537	53,828

Note: N/A – not available.

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
	9.2	Sum of weights after adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,324,435	3,293,173	334,506
		a.	Chld. scrnd. as poor HH	195,785	135,617	1,515,568	139,586	538,072	173,324	325,701	31,125
		b.	Chld scrd as nonpoor HH retained	324,687	402,226	2,568,446	377,677	1,096,311	591,701	858,352	74,648
		c.	Chld scrnd as nonpoor HH not retained	0	0	0	0	0	0	0	0
		d.	Total child screened as nonpoor	324,687	402,226	2,568,446	377,677	1,096,311	591,701	858,352	74,648
		h.	Chld scrnd as unknown HH retained	28,524	24,410	290,259	25,444	111,000	46,854	73,898	6,183
		i.	Chld scrnd as unknwn HH not retained	0	0	0	0	0	0	0	0
		j.	Total child screened as unknown	28,524	24,410	290,259	25,444	111,000	46,854	73,898	6,183
		k.	Adlt scrnd. as poor HH	70,976	64,933	567,172	66,325	166,551	83,158	137,291	11,546
		l.	Adlt scrd as nonpoor HH retained	202,529	335,057	2,804,288	340,791	638,158	535,733	678,537	53,828
		m.	Adlt scrnd as nonpoor HH not retained	0	0	0	0	0	0	0	0
		n.	Total adult screened as nonpoor	202,529	335,057	2,804,288	340,791	638,158	535,733	678,537	53,828
		o.	Adlt scrnd as unknown HH retained	19,481	25,128	251,207	23,754	67,868	45,841	66,364	4,192
		p.	Adlt scrnd as unknwn HH not retained	0	0	0	0	0	0	0	0
		q.	Total adult screened as unknown	19,481	25,128	251,207	23,754	67,868	45,841	66,364	4,192
		r.	No income screened	638,788	529,755	3,106,918	450,139	2,931,733	847,824	1,153,029	152,987
	9.3	Sum of weights after adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,324,435	3,293,173	334,507
		a.	HH with children	548,996	562,253	4,374,274	542,708	1,745,383	811,879	1,257,952	111,955
		b.	HH with no children	615,901	639,748	4,705,753	649,066	2,313,356	1,034,964	1,339,534	148,136
		c.	Elderly HHs	315,874	315,124	2,023,832	231,943	1,490,954	477,591	695,688	74,416
10	Reflecting path AB subsampling										
	10.1	Sum of weights before adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,324,435	3,293,173	334,507
		a.	HH with children	548,996	562,253	4,374,274	542,708	1,745,383	811,879	1,257,952	111,955
		b.	HH with no children retained	292,986	425,117	3,622,668	430,870	872,577	664,731	882,193	69,565
		c.	HH with no children not retained	322,915	214,630	1,083,086	218,196	1,440,780	370,233	457,341	78,571
		d.	Elderly HH	315,874	315,124	2,023,832	231,943	1,490,954	477,591	695,688	74,416

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
	10.2	Sum of weights after adjustment		1,480,770	1,517,125	11,103,859	1,423,716	5,549,693	2,324,435	3,293,173	334,507
		a.	HH with children	548,996	562,253	4,374,274	542,708	1,745,383	811,879	1,257,952	111,955
		b.	HH with no children retained	615,900	639,748	4,705,753	649,066	2,313,356	1,034,964	1,339,534	148,136
		c.	HH with no children not retained	0	0	0	0	0	0	0	0
		d.	Elderly HH	315,874	315,124	2,023,832	231,943	1,490,954	477,591	695,688	74,416

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
1	Sample										
	1.1	Total telephone numbers in sample		32,000	16,000	52,000	34,000	27,000	32,000	43,260	484,530
		a.	Residential telephone numbers	15,087	8,170	24,204	15,499	11,737	14,500	20,090	223,969
		b.	Nonresidential telephone numbers	15,440	7,274	23,124	16,098	13,842	15,772	20,985	231,246
		c.	Unknown	1,473	556	4,672	2,403	1,421	1,728	2,185	29,315
2	Base weight										
	2.1	Sum of weights		3,510,600	1,546,500	6,386,500	14,535,600	13,960,100	4,350,700	102,230,000	203,430,600
3	Allocating unknown residential										
	3.1	Sum of weights before adjustment		1,706,787	805,792	3,153,993	6,949,800	6,295,917	2,049,870	49,113,409	97,219,746
		a.	Residential known eligibility	1,445,709	698,825	2,248,171	5,072,069	5,014,261	1,650,819	40,056,592	76,938,597
		b.	Residential unknown eligibility	209,429	90,857	724,499	1,554,027	1,054,246	320,592	7,417,712	16,547,976
		c.	Unknown residential allocated	51,649	16,111	181,323	323,704	227,410	78,459	1,639,104	3,733,202
	3.2	Sum of weights after adjustment		1,706,787	805,792	3,153,993	6,949,800	6,295,917	2,049,870	49,113,409	97,219,746
		a.	Residential known eligibility	1,445,709	698,825	2,248,171	5,072,069	5,014,261	1,650,819	40,056,592	76,938,597
		b.	Residential unknown eligibility	261,078	106,968	905,822	187,730	1,281,656	399,052	9,056,817	18,591,149
4	Adjusting for unknown eligibles										
	4.1	Sum of weights before adjustment		1,706,787	805,792	3,153,993	6,949,800	6,295,917	2,049,870	49,113,409	97,219,616
		a.	Res. elig. respondents	1,136,886	533,156	1,771,394	3,887,418	3,990,003	1,322,885	30,930,372	59,483,980
		b.	Res. elig. nonrespondents	23,477	13,532	47,776	134,241	103,408	36,573	855,435	1,682,286
		c.	Res. ineligible	285,346	152,137	429,001	1,050,411	920,333	291,361	8,270,785	15,771,115
		d.	Res. unknown (nonrespondents)	261,078	106,968	905,822	1,877,730	1,282,173	399,052	9,056,817	20,282,238
	4.2	Sum of weights after adjustment		1,706,787	805,792	3,153,993	6,949,800	6,295,917	2,049,870	49,113,409	97,219,616
		a.	Res. elig. respondents	1,345,220	615,526	2,495,600	5,340,033	5,038,409	1,647,945	38,089,146	75,466,136
		b.	Res. elig. nonrespondents	28,118	15,753	68,930	188,295	132,628	46,128	1,068,289	2,172,512
		c.	Res. ineligible	333,449	174,513	589,463	1,421,472	1,124,880	355,797	9,955,974	19,580,970
		d.	Res. unknown (nonrespondents)	0	0	0	0	0	0	0	0

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

					MN	MS	NJ	NY	TX	WA	U.S.	
					Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
5	Adjusting for eligible nonrespondents											
	5.1	Sum of weights after adjustment			1,706,787	805,792	3,153,993	6,949,800	6,295,917	2,049,870	49,113,409	97,219,616
		a.	Eligible respondents		1,373,338	631,279	2,564,529	5,528,328	5,171,038	1,694,073	39,157,435	77,638,647
		b.	Elig. nonrespondents		0	0	0	0	0	0	0	0
		c.	Ineligible		333,449	174,513	589,463	1,421,472	1,124,880	355,797	9,955,974	19,580,970
		d.	Unknown		0	0	0	0	0	0	0	0
6	Type of HHs before IK refielded											
	6.1	Sum of weights			1,706,787	805,792	3,153,993	6,949,800	6,295,917	2,049,870	49,113,409	97,219,616
		a.	HH with children		622,822	307,415	1,180,106	2,476,183	2,581,991	779,423	18,290,714	36,194,055
		b.	HH with no children		750,516	323,864	1,384,424	3,052,145	2,589,562	914,650	20,866,721	41,445,109
		c.	Elderly HHs		333,449	174,513	589,463	1,421,472	1,124,364	355,797	9,955,974	19,580,452
7	Adjusting for refielded IKs											
	7.1	Sum of weights before adjustment					3,153,993				49,113,409	
		a.	IK respondent		N/A	N/A	N/A	360,168	N/A	N/A	N/A	3,454,568
		b.	IK nonrespondent		N/A	N/A	N/A	6,607	N/A	N/A	N/A	45,251
		c.	IK ineligible		N/A	N/A	N/A	46,354	N/A	N/A	N/A	465,250
		d.	IK residential unknown elig.		N/A	N/A	N/A	69,368	N/A	N/A	N/A	551,204
		e.	IK nonresidential		N/A	N/A	N/A	63,904	N/A	N/A	N/A	684,473
		f.	Other		N/A	N/A	N/A	2,607,591	N/A	N/A	N/A	43,912,663
	7.2	Adjusting for unknown eligibility					3,090,089				48,428,936	
		a.	IK respondent		N/A	N/A	N/A	420,775	N/A	N/A	N/A	3,931,039
		b.	IK nonrespondent		N/A	N/A	N/A	7,867	N/A	N/A	N/A	51,927
		c.	IK ineligible		N/A	N/A	N/A	53,855	N/A	N/A	N/A	533,307
		d.	IK residential unknown elig.		N/A	N/A	N/A	0	N/A	N/A	N/A	0
		e.	IK nonresidential		N/A	N/A	N/A	0	N/A	N/A	N/A	0
		f.	Other		N/A	N/A	N/A	2,607,591	N/A	N/A	N/A	43,912,663

Note: N/A – not available.

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
	7.3	Adjusting for eligible nonrespondents				3,036,233				47,895,629	
		a.	IK respondent	N/A	N/A	N/A	428,642	N/A	N/A	N/A	3,982,966
		b.	IK nonrespondent	N/A	N/A	N/A	0	N/A	N/A	N/A	
		c.	IK ineligible	N/A	N/A	N/A	0	N/A	N/A	N/A	
		d.	IK residential unknown elig.	N/A	N/A	N/A	0	N/A	N/A	N/A	
		e.	IK nonresidential	N/A	N/A	N/A	0	N/A	N/A	N/A	
		f.	Other	N/A	N/A	N/A	2,607,591	N/A	N/A	N/A	43,912,663
		g.	Losses after refield (among re-fielded IKs)	N/A	N/A	N/A	117,759	N/A	N/A	N/A	1,217,780
8	Type of HHs after IK refield										
	8.1	Sum of weights after 1K adjustment		1,480,770	1,706,787	805,792	3,036,233	6,949,800	6,295,917	2,049,870	47,895,629
		a.	HH with children	548,996	622,822	307,415	1,180,106	2,476,183	2,581,991	779,423	18,290,714
		b.	HH with no children	615,900	750,516	323,864	1,266,664	3,052,145	2,589,562	914,650	19,648,941
		c.	Elderly HHs	315,874	333,449	174,513	589,463	1,421,472	1,124,364	355,797	9,955,974
		d.	Losses after refield (among no chld HH)	N/A	N/A	N/A	117,759	N/A	N/A	N/A	1,217,780
9	Reflecting poverty subsampling										
	9.1	Sum of weights before adjustment		1,480,770	1,706,787	805,792	3,036,233	6,949,800	6,295,917	2,049,870	47,895,629
		a.	Chld. scrnd. as poor HH	195,785	150,582	126,363	211,010	780,883	866,024	210,074	5,327,695
		b.	Chld scrd as nonpoor HH retained	121,086	118,463	61,348	213,301	481,894	456,682	171,059	3,669,454
		c.	Chld scrnd as nonpoor HH not retained	203,601	330,534	101,310	674,922	1,052,120	1,091,394	356,608	8,240,165
		d.	Total child screened as nonpoor	324,687	448,997	162,658	888,224	1,534,014	1,548,076	527,667	11,909,619
		n.	Adlt scrnd. as poor HH	70,976	75,673	48,174	123,352	256,995	222,814	109,944	3,048,181
		o.	Adlt scrd as nonpoor HH retained	73,306	122,991	43,201	211,546	323,710	256,342	160,192	3,797,323
		p.	Adlt scrnd as nonpoor HH not retained	129,223	317,786	60,703	706,296	787,804	594,422	353,047	8,826,753
		q.	Total adult screened as nonpoor	202,529	440,776	103,905	917,842	1,111,514	850,764	513,239	12,624,076

Note: N/A – not available.

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
	9.2	Sum of weights after adjustment		1,706,787	805,792	3,036,233	6,949,800	6,295,917	2,049,870	47,895,629	0
		a.	Chld. scrnd. as poor HH	150,582	126,363	211,010	780,883	866,024	210,074	5,327,695	10,727,410
		b.	Chld scrd as nonpoor HH retained	448,997	162,658	888,224	1,534,014	1,548,076	527,667	11,909,619	23,313,305
		c.	Chld scrnd as nonpoor HH not retained	0	0	0	0	0	0	0	0
		d.	Total child screened as nonpoor	448,997	162,658	888,224	1,534,014	1,548,076	527,667	11,909,619	23,313,305
		h.	Chld scrnd as unknown HH retained	23,243	18,394	80,871	161,285	167,891	41,682	1,053,400	2,153,340
		i.	Chld scrnd as unknwn HH not retained	0	0	0	0	0	0	0	0
		j.	Total child screened as unknown	23,243	18,394	80,871	161,285	167,891	41,682	1,053,400	2,153,340
		k.	Adlt scrnd. as poor HH	75,673	48,174	123,352	256,995	222,814	109,944	3,048,181	5,053,083
		l.	Adlt scrd as nonpoor HH retained	440,776	103,905	917,842	1,111,514	850,764	513,239	12,624,076	22,151,037
		m.	Adlt scrnd as nonpoor HH not retained	0	0	0	0	0	0	0	0
		n.	Total adult screened as nonpoor	440,776	103,905	917,842	1,111,514	850,764	513,239	12,624,076	22,151,037
		o.	Adlt scrnd as unknown HH retained	31,596	12,935	81,235	123,575	78,170	41,299	1,092,967	
		p.	Adlt scrnd as unknwn HH not retained	0	0	0	0	0	0	0	0
		q.	Total adult screened as unkn	31,596	12,935	81,235	123,575	78,170	41,299	1,092,967	0
		r.	No income screened	535,919	333,364	733,700	2,981,532	2,562,179	605,966	12,839,691	30,403,522
	9.3	Sum of weights after adjustment		1,706,787	805,792	3,036,233	6,949,800	6,295,917	2,049,870	47,895,629	95,767,307
		a.	HH with children	622,822	307,415	1,180,106	2,476,183	2,581,991	779,423	18,290,714	36,194,055
		b.	HH with no children	750,516	323,864	1,266,664	3,052,145	2,589,562	914,650	19,648,941	39,992,799
		c.	Elderly HHs	333,449	174,513	589,463	1,421,472	1,124,364	355,797	9,955,974	19,580,454
10	Reflecting path AB subsampling										
	10.1	Sum of weights before adjustment		1,706,787	805,792	3,036,233	6,949,800	6,295,917	2,049,870	47,895,629	95,767,307
		a.	HH with children	622,822	307,415	1,180,106	2,476,183	2,581,991	779,423	18,290,714	36,194,055
		b.	HH with no children retained	548,046	165,013	1,122,428	1,492,085	1,151,747	664,482	16,765,224	29,169,730
		c.	HH with no children not retained	202,470	158,851	144,236	1,560,060	1,437,815	250,168	2,883,717	10,823,070
		d.	Elderly HH	333,449	174,513	589,463	1,421,472	1,124,364	355,797	9,955,974	19,580,452

Table A-1.
Household Weighting for the RDD Sample, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
	10.2		Sum of weights after adjustment	1,706,787	805,792	3,036,233	6,949,800	6,295,917	2,049,870	47,895,629	95,767,307
		a.	HH with children	622,822	307,415	1,180,106	2,476,183	2,581,991	779,423	18,290,714	36,194,055
		b.	HH with no children retained	750,516	323,864	1,266,664	3,052,145	2,589,562	914,650	19,648,941	39,992,800
		c.	HH with no children not retained	0	0	0	0	0	0	0	0
		d.	Elderly HH	333,449	174,513	589,463	1,421,473	1,124,364	355,797	9,955,974	19,580,453

Table A-2.
Household Weighting for the Area Sample, by Study Area

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc.	California	Colorado	Florida	Mass.	Michigan	Milwaukee
1	Baseweights										
	1.1	Number of segments		106	70	71	120	106	76	64	50
	1.2	Total households listed		3,587	1,987	2,166	3,513	3,453	2,638	2,263	1,877
2	Adjusting unlisted segments										
	2.1	Number of segments		106	69	71	120	106	76	64	50
	2.2	Listed segments sum of weights		33,622	24,341	194,221	28,373	100,740	20,318	48,265	3,751
	2.3	Unlisted segments sum of weights		0	358	0	0	0	0	0	0
	2.4	Adjusted segments sum of weights		33,622	24,699	194,221	28,373	100,740	20,318	48,265	3,751
3	Adjusting weights for large segments										
	3.1	Sum of weights before adjustment		1,123,826	710,103	5,433,014	826,707	3,238,297	710,885	1,702,614	136,889
	3.2	Sum of weights after adjustment		1,140,502	710,103	5,433,014	850,023	3,711,128	710,885	1,702,614	136,889
	3.3	Difference		16,677	0	0	23,316	472,832	0	0	0
4	Nonresponse adjustment										
	4.1	Adjusting for unknown eligibles									
		a.	Residential eligible respondents	53,306	12,206	146,115	23,258	82,034	9,264	50,863	6,583
		b.	Residential eligible nonrespondents	8,075	1,797	30,251	2,661	10,976	3,919	4,243	1,062
		c.	Residential ineligible	952,047	575,356	4,656,117	671,690	3,102,419	648,238	1,423,293	115,796
		d.	Residential unknown eligibility	2,979	7,831	104,097	26,288	30,494	5,109	3,551	1,699
		e.	Nonresidential	124,096	112,913	496,435	126,126	485,205	44,356	220,664	11,749
	4.2	Adjusting for eligible nonrespondents									
		a.	Residential eligible respondents	53,739	12,436	149,851	24,194	82,684	9,432	50,929	6,725
		b.	Residential eligible nonrespondents	8,158	1,814	30,251	2,693	11,043	3,958	4,310	1,077
		c.	Residential ineligible	954,510	582,940	4,756,478	697,009	3,132,197	653,139	1,426,711	117,337
	4.3	Residential eligible respondents		61,897	14,250	180,101	26,888	93,726	13,390	55,239	7,802
5	Household undercoverage adjustment										
	5.1	Average adjustment		1.28	1.04	1.49	1.74	1.38	1.89	1.46	1.37
	5.2	Undercoverage adjusted weight		78,911	14,868	268,422	46,997	130,811	25,299	80,484	7,875

Table A-2.
Household Weighting for the Area Sample, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
1	Baseweights										
	1.1	Number of segments		69	89	136	118	79	82	152	1,388
	1.2	Total households listed		2,268	3,232	3,943	3,531	2,586	2,740	4,690	44,474
2	Adjusting unlisted segments										
	2.1	Number of segments		69	88	134	118	79	82	152	1,384
	2.2	Listed segments sum of weights		25,292	19,510	33,168	102,332	130,080	33,589	940,809	1,738,413
	2.3	Unlisted segments sum of weights		0	187	451	0	0	0	0	996
	2.4	Adjusted segments sum of weights		25,292	19,698	33,620	102,332	130,080	33,589	940,809	1,739,409
3	Adjusting weights for large segments										
	3.1	Sum of weights before adjustment		854,740	720,237	979,044	2,953,503	4,293,552	1,129,389	28,327,165	53,139,963
	3.2	Sum of weights after adjustment		882,425	730,810	1,001,397	2,953,503	4,464,247	1,211,270	28,913,271	54,552,082
	3.3	Difference		27,685	10,573	22,354	0	170,696	81,881	586,106	1,412,119
4	Nonresponse adjustment										
	4.1	Adjusting for unknown eligibles									
		a.	Residential eligible respondents	22,677	57,501	30,153	42,861	213,463	25,890	1,103,557	1,879,731
		b.	Residential eligible nonrespondents	3,769	7,002	9,322	10,944	38,050	11,770	133,066	276,907
		c.	Residential ineligible	764,563	558,009	847,964	2,533,345	3,768,728	1,023,484	24,998,663	46,639,712
		d.	Residential unknown eligibility	6,305	8,253	16,787	70,821	14,108	40,959	50,646	389,927
		e.	Nonresidential	85,111	100,044	97,172	295,532	429,898	109,167	2,627,340	5,365,808
	4.2	Adjusting for eligible nonrespondents									
		a.	Residential eligible respondents	22,727	57,818	30,320	43,860	213,689	26,195	1,105,842	1,890,441
		b.	Residential eligible nonrespondents	3,769	7,030	9,861	11,007	38,074	12,206	134,194	279,445
		c.	Residential ineligible	770,818	565,917	864,045	2,603,104	3,782,586	1,063,703	25,045,896	47,016,390
	4.3	Residential eligible respondents		26,496	64,848	40,181	54,867	251,763	38,400	1,240,036	2,169,884
5	Household undercoverage adjustment										
	5.1	Average adjustment		1.69	1.24	1.54	1.34	1.37	1.37	1.28	1.33
	5.2	Undercoverage adjusted weight		44,804	80,747	62,021	73,677	345,916	52,780	1,586,157	2,899,770

Table A-3.
Extended Interview Weighting for Child Interview, by Study Area

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
1	Base weight										
	1.1	Total no. of children w/comp. screeners		2,529	2,599	2,680	2,798	2,540	2,942	2,583	2,219
	1.2	Sum of weights		1,153,528	1,136,637	8,841,795	1,059,581	3,259,641	1,475,028	2,649,303	237,266
	1.3	Area – sample size		121	25	55	65	76	27	61	61
	1.4	Area – sum of weights		110,353	16,946	516,105	40,256	153,980	35,572	117,760	9,688
	1.5	Telephone – sample size		2,408	2,574	2,625	2,733	2,464	2,915	2,522	2,158
	1.6	Telephone – sum of weights		1,043,175	1,119,691	8,325,690	1,019,365	3,105,661	1,439,455	2,531,544	227,577
2	Adjusting for eligible nonrespondents										
	2.1	Sum of weights before adjustment		1,153,528	1,136,637	8,841,795	1,059,620	3,259,641	1,475,028	2,649,303	237,265
		a.	Eligible respondents	975,530	1,006,300	6,888,834	872,499	2,658,293	1,212,619	2,232,571	196,099
		b.	Eligible nonrespondents	174,803	130,203	1,933,160	186,874	597,910	260,990	416,732	41,166
		c.	Ineligible	3,195	134	19,801	247	3,437	1,418	0	0
	2.2	Sum of weights after adjustment		1,153,528	1,136,637	8,841,795	1,059,620	3,259,641	1,475,028	2,649,303	237,265
		a.	Eligible respondents	1,150,333	1,136,503	8,821,995	1,059,373	3,256,203	1,473,609	2,649,303	237,265
		b.	Eligible nonrespondents	0	0	0	0	0	0	0	0
		c.	Ineligible	3,195	134	19,801	247	3,437	1,418	0	0
		d.	Sample size	2,098	2,320	2,060	2,298	2,063	2,381	2,143	1,804
3	Adjusting for multiple telephone households										
	3.1	Sum of weights before adjustment		1,150,333	1,136,503	8,821,995	1,059,373	3,256,203	1,473,609	2,649,303	237,265
	3.2	Mean factor (weighted)		0.93	0.95	0.93	0.94	0.90	0.92	0.93	0.91
	3.3	Sum of weights after adjustment		1,075,259	1,079,733	8,178,503	995,623	2,936,046	1,359,856	2,460,936	216,396
	3.4	% records with more than 1 phone		10.6%	8.3%	11.2%	9.3%	16.3%	12.1%	11.3%	14.2%

Table A-3.
Extended Interview Weighting for Child Interview by Study Area

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
4	First Raking										
	4.1	Sum of weights after adjustment		1,081,790	1,116,360	8,942,851	1,016,077	3,461,253	1,426,366	2,542,812	240,797
	4.2	a.	Mean factor (weighted)	1.01	1.03	1.09	1.02	1.18	1.05	1.03	1.11
		b.	Min factor	0.64	0.67	0.76	0.79	0.08	0.65	0.70	0.84
		c.	Max factor	1.41	1.43	1.50	1.55	2.32	1.50	1.70	1.79
	4.3	Factors race/ethnicity									
		a.	Hispanic	0.68	0.89	1.03	1.04	1.06	0.97	0.82	0.96
		b.	Black, non-Hispanic	1.14	1.00	1.21	1.21	1.38	1.14	1.15	1.38
		c.	Other	0.96	1.04	1.13	1.01	1.15	1.05	1.02	1.02
	4.4	Factors sex									
		a.	Female	1.03	1.03	1.12	1.02	1.17	1.04	1.05	1.12
		b.	Male	0.98	1.04	1.07	1.02	1.19	1.06	1.02	1.10
	4.5	Factors home ownership									
		a.	Owner	0.97	1.01	1.04	0.98	1.07	1.02	1.01	1.02
		b.	Renter	1.11	1.12	1.15	1.11	1.42	1.10	1.13	1.24
5	Trimming										
	5.1	Sum of weights after adjustment		1,075,911	1,116,360	8,942,851	1,016,077	3,461,253	1,426,366	2,542,812	240,797
	5.2	Max area weight after trimming		3,995	1,699	35,977	2,698	10,830	3,165	5,138	610
	5.3	Max RDD weight after trimming		1,994	2,766	23,604	2,533	11,497	3,646	6,565	845
	5.4	Number of records trimmed		2	0	0	0	0	0	0	0
	5.5	Difference		5,879	0	0	0	0	0	0	0
6	Second raking										
	6.1	Sum of weights after adjustment		1,081,790	1,116,360	8,942,851	1,016,077	3,461,253	1,426,366	2,542,812	240,797

Table A-3.
Extended Interview Weighting for Child Interview, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
1	Base weight										
	1.1	Total no. of children w/comp. screeners		2,663	2,342	3,254	2,841	2,785	2,892	3,932	41,599
	1.2	Sum of weights		1,291,980	801,762	1,938,775	4,469,676	5,818,269	1,521,959	35,362,526	71,017,725
	1.3	Area – sample size		57	202	104	44	119	32	154	1,203
	1.4	Area – sum of weights		66,026	132,417	70,910	90,492	568,784	48,856	2,160,367	4,138,511
	1.5	Telephone – sample size		2,606	2,140	3,150	2,797	2,666	2,860	3,778	40,396
	1.6	Telephone – sum of weights		1,225,954	669,345	1,867,865	4,379,184	5,249,484	1,473,103	33,202,159	66,879,252
2	Adjusting for eligible nonrespondents										
	2.1	Sum of weights before adjustment		1,291,980	801,762	1,938,775	4,469,676	5,818,269	1,521,959	35,362,526	71,017,764
		a.	Eligible respondents	1,153,847	689,172	1,538,512	3,584,460	4,739,204	1,315,429	30,607,499	59,670,868
		b.	Eligible nonrespondents	138,134	111,857	398,415	881,601	1,072,989	206,374	4,747,971	130,688,632
		c.	Ineligible	0	732	1,848	3,615	6,076	156	7,056	47,715
	2.2	Sum of weights after adjustment		1,291,980	801,762	1,938,775	4,469,676	5,818,269	1,521,959	35,362,526	71,017,764
		a.	Eligible respondents	1,291,980	801,029	1,936,927	4,466,061	5,812,193	1,521,803	35,355,470	70,970,047
		b.	Eligible nonrespondents	0	0	0	0	0	0	0	0
		c.	Ineligible	0	732	1,848	3,615	6,076	156	7,056	47,715
		d.	Sample size	2,360	1,984	2,566	2,252	2,249	2,469	3,392	34,439
3	Adjusting for multiple telephone households										
	3.1	Sum of weights before adjustment		1,291,980	801,029	1,936,927	4,466,061	5,812,193	1,521,803	35,355,470	70,970,047
	3.2	Mean factor (weighted)		0.95	0.96	0.86	0.94	0.91	0.92	0.93	0.93
	3.3	Sum of weights after adjustment		1,224,288	771,350	1,664,714	4,182,459	5,288,144	1,403,822	33,034,157	65,871,287
	3.4	% records with more than 1 phone		8.1%	6.9%	22.7%	8.6%	14.4%	12.6%	10.3%	

Table A-3.
Extended Interview Weighting for Child Interview, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
4	First Raking										
	4.1	Sum of weights after adjustment		1,265,667	761,592	1,987,303	4,536,723	5,555,259	1,460,250	34,223,123	69,618,223
	4	a.	Mean factor (weighted)	1.03	0.99	1.19	1.08	1.05	1.04	1.04	1.06
		b.	Min factor	0.69	0.41	0.92	0.71	0.64	0.55	0.85	
		c.	Max factor	1.49	1.83	1.77	1.83	1.61	1.50	1.92	
	4.3	Factors race/ethnicity									
		a.	Hispanic	0.86	0.42	1.23	1.07	1.03	0.88	1.03	
		b.	Black, non-Hispanic	0.94	1.08	1.37	1.18	1.15	1.12	1.45	
		c.	Other	1.04	0.94	1.15	1.07	1.04	1.06	0.98	
	4.4	Factors sex									
		a.	Female	1.08	1.01	1.25	1.14	1.10	1.04	1.05	
		b.	Male	1.00	0.96	1.15	1.04	1.01	1.04	1.03	
	4.5	Factors home ownership									
		a.	Owner	1.02	0.94	1.14	1.00	1.00	1.01	0.98	
		b.	Renter	1.08	1.10	1.32	1.19	1.15	1.10	1.17	
5	Trimming										
	5.1	Sum of weights after adjustment		1,265,667	761,592	1,987,303	4,531,937	5,555,259	1,460,250	34,211,245	69,595,680
	5.2	Max area weight after trimming		3,290	2,581	3,242	7,935	21,864	3,161	55,759	
	5.3	Max RDD weight after trimming		3,163	2,652	4,629	9,652	17,302	2,951	62,282	
	5.4	Number of records trimmed		0	0	0	2	0	0	1	5
	5.5	Difference		0	0	0	4,786	0	0	11,878	22,543
6	Second raking										
	6.1	Sum of weights after adjustment		1,265,667	761,592	1,987,303	4,536,723	5,555,259	1,460,250	34,223,122	69,618,222

Table A-4.
Extended Interview Weighting for Adults, by Study Area

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
1	HH Adjustment for completed interview			1,366,915	1,314,078	9,731,175	1,342,244	4,452,744	1,941,276	3,046,894	300,568
	1.1	Unweighted respondents		2,553	3,086	2,543	3,175	2,386	3,238	2,776	2,269
	1.2	Complete extended		1,156,556	1,145,076	7,627,632	1,101,451	3,592,178	1,558,091	2,535,429	251,521
	1.3	Nonrespondents		210,359	169,002	2,103,543	240,793	860,566	383,185	511,465	49,047
2	HH Adjustment for A/B path			1,366,915	1,314,077	9,731,175	1,342,244	4,452,744	1,941,276	3,046,894	300,569
	2.1	HH with children retained		289,396	389,869	3,383,214	370,151	697,245	354,578	732,171	59,373
	2.2	HH with children not retained		294,586	195,698	754,023	166,708	1,019,958	410,009	587,745	59,599
	2.3	Other		782,933	728,510	5,593,938	805,385	2,735,541	1,176,689	1,726,978	181,597
3	Initial base weight										
	3.1	Total no. of age eligible adults		4,254	5,456	4,459	5,453	4,008	5,404	4,780	3,571
	3.2	Sum of weights		2,523,067	2,595,874	19,336,768	2,532,006	8,373,631	3,761,813	5,960,220	547,648
4	Adjusting for eligible nonrespondents										
	4.1	Sum of weights before adjustment		2,523,067	2,595,874	19,336,768	2,532,006	8,373,631	3,761,813	5,960,220	547,648
		a.	Eligible respondents	2,452,034	2,518,363	18,322,449	2,464,485	8,067,581	3,625,612	5,725,656	529,747
		b.	Eligible nonrespondents	66,380	59,962	957,814	58,727	272,945	120,876	199,043	16,997
		c.	Ineligible	4,653	17,549	56,505	8,794	33,105	15,325	35,522	904
	4.2	Sum of weights after adjustment		2,523,067	2,595,874	19,336,768	2,532,006	8,373,631	3,761,813	5,960,220	547,648
		a.	Eligible respondents	2,518,414	2,578,325	19,280,263	2,523,212	8,340,526	3,746,488	5,924,699	546,744
		b.	Eligible nonrespondents	0	0	0	0	0	0	0	0
		c.	Ineligible	4,653	17,549	56,505	8,794	33,105	15,325	35,522	904
		d.	Unweighted number	4,182	5,362	4,282	5,344	3,922	5,297	4,677	3,497
5	Adjusting for multiple telephone households										
	5.1	Sum of weights before adjustment		2,518,414	2,578,325	19,280,263	2,523,212	8,340,526	3,746,488	5,924,699	546,744
	5.2	Mean factor (weighted)		0.95	0.95	0.91	0.94	0.90	0.92	0.94	0.92
	5.3	Sum of weights after adjustment		2,382,363	2,454,583	17,491,399	2,372,121	7,545,550	3,456,224	5,573,790	505,028
	5.4	% with more than 1 phone		10.1%	8.4%	13.8%	10.1%	16.4%	12.8%	10.7%	14.0%

Table A-4.
Extended Interview Weighting for Adults, by Study Area (Continued)

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
6	First raking										
	6.1	Sum of weights after adjustment		2,622,693	2,571,365	19,342,093	2,458,623	8,288,729	3,786,843	5,813,743	558,606
	6.2	a.	Mean factor (weighted)	1.10	1.05	1.11	1.04	1.10	1.10	1.04	1.11
		b.	Min factor	0.54	0.63	0.64	0.67	0.81	0.65	0.57	0.51
		c.	Max factor	2.01	1.77	2.44	1.97	2.41	1.94	2.13	2.28
	6.3	Factors race/ethnicity									
		a.	Hispanic	0.67	0.86	1.15	1.05	1.22	1.06	1.01	0.94
		b.	Black, non-Hispanic	1.32	1.34	1.47	1.32	1.45	1.22	1.32	1.39
		c.	Other	1.05	1.05	1.06	1.02	1.03	1.09	1.01	1.06
	6.4	Factors sex									
		a.	Female	1.09	1.03	1.05	1.03	1.05	1.04	1.03	1.10
		b.	Male	1.12	1.06	1.16	1.04	1.16	1.16	1.06	1.11
	6.5	Factors home ownership									
		a.	Owner	1.06	1.11	1.09	1.01	1.02	1.06	1.01	1.04
		b.	Renter	1.24	1.15	1.13	1.10	1.30	1.16	1.17	1.21
	6.6	Factors Education									
		a.	Some high school	1.37	1.30	1.31	1.29	1.57	1.36	1.33	1.37
		b.	Finished high school	1.09	1.06	1.12	1.06	1.04	1.13	1.05	1.12
		c.	Some college	0.94	0.93	0.96	0.93	1.04	0.99	0.92	0.96
7	Trimming										
	7.1	Sum of weights after adjustment		2,621,638	2,571,365	19,331,517	2,458,623	8,288,729	3,786,843	5,813,743	558,606
	7.2	Number of records trimmed		1	0	1	0	0	0	0	0
	7.3	Difference		1,055	0	10,576	0	0	0	0	0
8	Second raking										
	8.1	Sum of weights after adjustment		2,622,693	2,571,365	19,342,093	2,458,623	8,288,729	3,786,843	5,813,743	558,606

Table A-4.
Extended Interview Weighting for Adults, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
1	HH Adjustment for completed interview			1,482,306	822,316	2,424,712	5,625,363	6,119,990	1,845,297	42,264,584	84,080,462
	1.1	Unweighted respondents		3,285	2,390	3,567	2,632	2,452	3,393	4,716	44,461
	1.2	Complete extended		1,292,667	694,691	1,878,659	4,456,737	4,943,707	1,572,256	35,659,271	69,465,922
	1.3	Nonrespondents		189,639	127,625	546,053	1,168,626	1,176,283	273,041	6,605,313	14,614,540
2	HH Adjustment for A/B path			1,482,306	822,316	2,424,711	5,625,364	6,119,990	1,845,296	42,264,583	84,080,460
	2.1	HH with children retained		464,628	205,758	570,212	1,178,133	1,232,214	555,568	11,732,002	22,214,512
	2.2	HH with children not retained		157,351	159,878	425,796	1,104,109	1,449,369	204,273	5,761,934	12,751,036
	2.3	Other		860,327	456,680	1,428,703	3,343,122	3,438,407	1,085,455	24,770,647	49,114,912
3	Initial base weight										
	3.1	Total no. of age eligible adults		5,630	4,018	6,139	4,419	4,303	5,907	8,072	75,873
	3.2	Sum of weights		2,806,372	1,566,629	4,956,803	10,872,579	12,106,092	3,505,357	81,917,505	163,362,363
4	Adjusting for eligible nonrespondents										
	4.1	Sum of weights before adjustment		2,806,372	1,566,629	4,956,803	10,872,579	12,106,092	3,505,357	81,917,505	163,362,366
		a.	Eligible respondents	2,751,235	1,528,060	4,662,405	10,332,458	11,513,069	3,409,977	79,276,571	157,179,702
		b.	Eligible nonrespondents	50,024	34,326	269,961	483,687	476,506	80,360	2,332,963	5,480,571
		c.	Ineligible	5,113	4,243	24,437	56,434	116,517	15,020	307,972	702,093
	4.2	Sum of weights after adjustment		2,806,372	1,566,629	4,956,803	10,872,579	12,106,092	3,505,357	81,917,505	163,362,366
		a.	Eligible respondents	2,801,259	1,562,386	4,932,366	10,816,145	11,989,575	3,490,337	81,609,533	162,660,273
		b.	Eligible nonrespondents	0	0	0	0	0	0	0	0
		c.	Ineligible	5,113	4,243	24,437	56,434	116,517	15,020	307,972	702,093
		d.	Unweighted number	5,557	3,941	5,940	4,291	4,181	5,790	7,905	74,168
5	Adjusting for multiple telephone households										
	5.1	Sum of weights before adjustment		2,801,259	1,562,386	4,932,366	10,816,145	11,989,575	3,490,337	81,609,533	162,660,273
	5.2	Mean factor (weighted)		0.95	0.96	0.84	0.93	0.91	0.93	0.94	0.93
	5.3	Sum of weights after adjustment		2,668,376	1,507,598	4,166,994	10,101,477	10,957,506	3,245,619	76,578,984	151,007,612
	5.4	% with more than 1 phone		8.1%	6.6%	24.9%	9.6%	14.7%	12.5%	10.0%	

Table A-4.
Extended Interview Weighting for Adults, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
6	First raking										
	6.1	Sum of weights after adjustment		2,849,258	1,620,127	4,852,837	11,043,448	11,795,966	3,468,579	80,582,935	161,655,845
	6.2	a.	Mean factor (weighted)	1.07	1.07	1.16	1.09	1.08	1.07	1.05	
		b.	Min factor	0.49	0.33	0.84	0.58	0.63	0.54	0.58	
		c.	Max factor	1.80	2.21	2.09	2.73	2.01	2.15	1.70	
	6.3	Factors race/ethnicity									
		a.	Hispanic	0.79	0.42	1.19	1.19	1.20	0.83	1.06	
		b.	Black, non-Hispanic	1.31	1.23	1.50	1.39	1.17	1.56	1.26	
		c.	Other	1.07	1.02	1.12	1.03	1.01	1.07	1.03	
	6.4	Factors sex									
		a.	Female	1.05	1.05	1.14	1.06	1.04	1.05	1.03	
		b.	Male	1.09	1.10	1.20	1.13	1.11	1.09	1.07	
	6.5	Factors home ownership									
		a.	Owner	1.04	1.04	1.11	1.06	1.06	1.03	1.02	
		b.	Renter	1.18	1.20	1.29	1.14	1.11	1.15	1.16	
	6.6	Factors Education									
		a.	Some high school	1.36	1.31	1.48	1.82	1.43	1.23	1.32	
		b.	Finished high school	1.09	1.06	1.19	1.11	1.08	1.10	1.06	
		c.	Some college	0.96	0.90	1.05	0.86	0.86	0.95	0.92	
7	Trimming										
	7.1	Sum of weights after adjustment		2,849,258	1,619,622	4,852,837	11,043,448	11,795,966	3,468,579	80,563,639	161,624,413
	7.2	Number of records trimmed		0	1	0	0	0	0	1	4
	7.3	Difference		0	505	0	0	0	0	19,296	31,432
8	Second raking										
	8.1	Sum of weights after adjustment		2,849,258	1,620,127	4,852,837	11,043,448	11,795,966	3,468,579	80,582,935	161,655,847

Table A-5.
Household Poststratification, by Study Area

				AL	BW	CA	CO	FL	MA	MI	ML
				Alabama	Bal. Wisc	California	Colorado	Florida	Mass.	Michigan	Milwaukee
1	Sum of weights before poststratification										
	1.1	Area sample									
		a.	Adult	32,589	6,231	88,042	24,006	61,836	9,884	32,443	4,129
		b.	Child	46,322	8,636	180,381	22,991	68,975	15,415	48,041	3,746
	1.2	RDD sample									
		a.	Adult	615,900	639,748	4,705,753	649,584	2,313,356	1,034,964	1,339,534	148,136
		b.	Child	548,996	562,253	4,374,274	54,174	1,745,383	811,879	1,257,952	111,955
	1.3	Combined sample									
		a.	Adult	648,490	645,979	4,793,795	673,590	2,375,193	1,044,848	1,371,977	152,265
		b.	Child	595,318	570,890	4,554,654	77,165	1,814,358	827,294	1,305,993	115,701
2	Sum of weights before trimming										
	2.1	Area sample									
		a.	Adult	32,589	6,231	88,042	24,006	61,836	9,884	32,443	4,129
		b.	Child	46,322	8,636	180,381	22,991	68,975	15,415	48,041	3,746
	2.2	Number of records trimmed		0	0	2	0	2	0	0	0
3	Sum of weights after trimming										
	3.1	Area sample									
		a.	Adult	32,589	6,231	88,042	23,666	60,366	9,884	32,443	4,129
		b.	Child	46,322	8,636	171,339	22,991	68,975	15,415	48,041	3,746
	3.2	Difference									
		a.	Adult	0	0	0	340	1,470	0	0	0
		b.	Child	0	0	9,041	0	0	0	0	0
4	Sum of weights after poststratification										
	4.1	Area sample									
		a.	Adult	36,869	6,941	99,760	27,510	67,842	10,993	39,645	4,817
		b.	Child	49,274	8,994	162,055	22,785	67,861	14,517	50,408	3,981
	4.2	RDD sample									
		a.	Adult	696,790	712,575	5,332,123	755,090	2,599,838	1,151,179	1,636,925	172,799
		b.	Child	583,982	585,568	4,137,237	536,859	1,717,203	764,587	1,319,916	118,971
	4.3	Combined sample									
		a.	Adult	733,659	719,516	5,431,883	782,600	2,667,680	1,162,172	1,676,570	177,616
		b.	Child	633,256	594,562	4,299,292	559,644	1,785,064	779,104	1,370,324	122,952

Table A-5.
Household Poststratification, by Study Area (Continued)

				MN	MS	NJ	NY	TX	WA	U.S.	
				Minnesota	Miss.	New Jersey	New York	Texas	Wash.	Bal. U.S.	Total U.S.
1	Sum of weights before poststratification										
	1.1	Area sample									
		a.	Adult	16,648	35,396	22,934	24,301	121,407	30,058	579,198	1,089,102
		b.	Child	28,155	45,352	39,087	49,376	224,509	22,722	1,006,959	1,810,668
	1.2	RDD sample									
		a.	Adult	750,516	323,864	1,266,664	3,052,145	2,589,562	914,650	18,883,574	39,227,951
		b.	Child	622,822	307,415	1,180,106	2,476,183	2,581,991	779,423	17,705,798	35,120,605
	1.3	Combined sample									
		a.	Adult	767,164	359,260	1,289,599	3,076,446	2,710,969	944,708	19,462,772	40,317,053
		b.	Child	650,978	352,767	1,219,193	2,525,559	2,806,501	802,145	18,712,757	36,931,272
2	Sum of weights before trimming										
	2.1	Area sample									
		a.	Adult	16,648	35,396	22,934	24,301	121,407	30,058	579,198	1,089,102
		b.	Child	28,155	45,352	39,087	49,376	224,509	22,722	1,006,959	1,810,668
	2.2	Number of records trimmed		0	0	0	0	1		0	5
3	Sum of weights after trimming										
	3.1	Area sample									
		a.	Adult	16,648	35,396	22,934	24,301	121,407	30,058	579,198	1,087,292
		b.	Child	28,155	45,352	39,087	49,376	219,453	22,722	1,006,959	1,796,570
	3.2	Difference									
		a.	Adult	0	0	0	0	0	0	0	1,810
		b.	Child	0	0	0	0	5,056	0	0	14,098
4	Sum of weights after poststratification										
	4.1	Area sample									
		a.	Adult	18,060	39,679	24,822	26,048	143,777	33,831	707,549	1,288,144
		b.	Child	28,117	53,941	32,989	45,509	227,918	22,151	994,910	1,785,410
	4.2	RDD sample									
		a.	Adult	814,150	363,060	1,370,892	3,271,565	3,066,712	1,029,473	23,068,188	46,041,358
		b.	Child	621,979	365,636	996,009	2,282,241	2,681,583	759,842	17,493,937	34,965,550
	4.3	Combined sample									
		a.	Adult	832,210	402,739	1,395,714	3,297,613	3,210,489	1,063,304	23,775,737	47,329,502
		b.	Child	650,096	419,577	1,028,998	2,327,750	2,909,501	781,993	18,488,847	36,750,960