U.S. Department of Health and Human Services
Administration for Children and Families
Administration on Children, Youth, and Families

National Survey on Child and Adolescent Well-Being
(NSCAW)

Office of Management and Budget
Clearance Package Supporting Statement
and Data Collection Instruments

Supporting Statement
Volume I

PART B: COLLECTION OF INFORMATION

EMPLOYING STATISTICAL METHODS

February 28, 2008

B.1 Respondent Universe and Sampling Methods

The following section describes the sample design proposed for NSCAW II, including the target population, methods to be used to select the two-stage samples, a comparison of the universe and proposed sample allocation across domains, and response rates expected based on NSCAW I experience.

The NSCAW I sample was designed to achieve several important goals related to the research objectives. First, a national probability sample was required in order that the sample represent all components of the child welfare system nationwide. Second, as mandated by the legislation authorizing the study, the NSCAW I sampling design provided for State estimates to the extent possible. Third, during the development of the survey objectives, it was determined that the sample size should be controlled for certain analytic domains in order to ensure adequate precision for certain key subgroups of the target population. These groups were defined by age, type of abuse/neglect, placement outside the home, and receipt of services, if any. Finally, the overall size of the sample needed to be such as to achieve the adequate precision for the key analytic domains while satisfying the budgetary constraints of the project.

The proposed NSCAW II sample design mirrors that used in NSCAW I with several notable exceptions. Assessing uses of the data by the research community, we have eliminated the Longer Term Foster Care sample component, and will retain only the Child Protective Service sample component. Similarly, the ability to calculate state estimates has not been an important NSCAW I feature, and is therefore not retained in the NSCAW II sample design. The allocation of the child sample to domains of analytic interest has been reconsidered and updated to focus more on children receiving services and children in out-of-home placement; the abuse type control has been eliminated. This more streamlined array of domains covers many research objectives and maintains a design effect and uneven weight effect close to that from NSCAW I. Conversely, we are extending the age for inclusion in the NSCAW II sample from 14 to 17 ½, due to increasing interest in children in their teen years in the child welfare system. The sample plan involves selecting children in the same Primary Sampling Units (PSUs) so that trends at the system level can be examined using the NSCAW I and II system-level data.

The fundamental features of the NSCAW II sample design are as follows:

- The NSCAW I and II primary sampling units (PSUs) were defined as geographic areas that encompass the population served by a child welfare agency (CWA). In most cases, these areas are counties or groups of counties. However, in large metropolitan areas, smaller geographic areas were defined.

- At the second stage of sampling, a total of 5,700 children will be interviewed. In most agencies, approximately 40–75 children will be randomly selected over the course of 12 months. In some counties, the number of children selected will be larger in order to reflect growth in the number of child abuse and neglect cases since the original sample was taken in 1999–2000 and to represent the 2008 child welfare population accurately.

- We anticipate a baseline response rate of 67 percent, or nearly 5 points better than in NSCAW I. We will analyze the potential for nonresponse bias using sample frame file data, and communicate any necessary cautions to analysts and data users.

- Longitudinal follow-up interviews will be conducted for the 5,700 children in the cohort (whose baseline interview is obtained). At the present time, an 18-month follow-up is scheduled.

The table below shows the expected number of baseline interviews for the sample domains.

Target Population

The target population for the NSCAW II sample is all children who are subjects of child abuse and neglect investigations (or assessments) conducted by CPS between February 2008 and January 2009. Thus, the target group is restricted to children who are reported to CPS and who are subjects of either an investigation or family assessment for child abuse or neglect. Although a number of these will go on to receive services, the target population also includes cases that are not substantiated and cases that are substantiated but do not subsequently receive services. We also anticipate that the sample eligibility criteria will result in the inclusion of some children in the system for other reasons such as dependency cases, status offenders, children on probation or children in need of supervision (ChINS), and children of families who voluntarily seek child welfare services. Cases that receive only such services such as well-baby visits from a public health nurse, mental health services, or preventive services to teen mothers are not included, except when these services are provided because of a child abuse or neglect report, the child is in child welfare custody, or the services are provided under the child welfare budget.

According to 2005 data from the National Child Abuse and Neglect Data System (NCANDS), over 3 million referrals of abuse or neglect, concerning approximately 6 million children, were received by child protective services (CPS) agencies. Almost 62 percent of those referrals were accepted for investigation or assessment.

Selection of the First-stage Sample Units

The selection of primary sampling units (PSUs) in 1998 involved five steps:

1. Partitioning the target population into PSUs

2. Obtaining size measure data for each PSU

2. Calculating the size measure for each PSU

4. Stratifying the PSU sampling frame

3. Selecting the PSUs

The first four steps constituted the frame construction phase. In this section, we outline the activities for carrying out each of the five steps for the NSCAW I sample, and our assessment of that sample for NSCAW II.

Step 1: Partitioning the Target Population into PSUs. The administrative structure of the child welfare system varies considerably across the states and even within states. Therefore, a single definition of a PSU was not feasible since it depended on the administrative structure of the state system, as well as the jurisdictions of child welfare agencies within the state. For most areas of the country, the best definition of a PSU was the county since it corresponds to a clearly defined political entity and geographic area of manageable size. In other areas, the definition of a PSU was not as straightforward, as in a single child welfare agency that had jurisdiction over several counties, the PSU was defined as a part of or the entire area over which the child welfare agency had jurisdiction, depending how large the area and population were. Extremely large counties or MSAs have child welfare agencies with many branch offices, each with its own data system. Such PSUs were divided into smaller units, such as areas delineated by branch office jurisdictions, to create manageable PSUs. For the purpose of the first-stage sampling discussion, we refer to counties as PSUs, for simplicity's sake.

Step 2: Obtaining Size Measure Data for Each PSU. The second-stage sampling units were stratified into 10 domains of interest. The primary reason was to control the second-stage sample allocation to ensure the domain of interest receive sufficient sample sizes. The second-stage NSCAW I domains and the allocation of achieved sample sizes are shown in *Exhibit B.1-2.*

Step 3: Calculating the Size Measures for Each PSU. The simplest size measure for PSUs in NSCAW would be simply the Census counts of children 14 and younger. This simple size measure gives counties with a higher number of children in the eligible age group a higher probability of being selected. However, such a design would be inefficient because we want to oversample children with certain characteristics in the PSU. One of the consequences of using simple size measures when the design calls for second-stage over-sampling of certain domains is that the variability of the sampling weights will be increased and, thus, so will the sampling variances. The composite size measure method, described in Folsom, Potter, and Williams (1987), provided a method for controlling domain sample sizes while maximizing the efficiency of the design. The composite size measure reflects the size of the sample that would fall into the PSU if a national random sample of children were selected with the desired sampling rates for all domains but without PSU clustering.

After the composite size measures were computed, we checked each of the 3,141 counties to determine whether it was large enough to support the planned selection of 67 CPS children per PSU during the twelve-month data collection period. We found 710 counties with fewer than 67 CPS children and deleted them from the first-stage sampling frame. Together, these counties accounted for approximately 22,362 CPS children or 0.9 percent of the target population. The estimated total size of the NSCAW I target population was 2,584,573 eligible CPS children in 2,431 counties nationwide.

Step 4: Stratifying the First-Stage Frame. As mentioned, the PSU frame was stratified into nine major strata. The eight key states where we made state-level estimates constituted the first eight strata. The PSUs in the remaining 42 states and D.C. were grouped into the ninth stratum. Within the nine strata, the PSUs were implicitly stratified by urbanicity for those in the eight-state strata, and by the four Census regions and then by urbanicity for the PSUs in the remaining states. The urbanicity of a PSU were defined by whether the county was part of an MSA. Stratifying PSUs by urbanicity and region allowed

for controlled allocation of sample PSUs in these strata. However, the number of explicit strata we defined was limited since we required that at least two PSUs be selected from each explicit stratum for variance estimation purposes.

Step 5: Selecting the PSUs. Given the first-stage stratification and the size measure S_{hk} , the selection frequency of the kth PSU in the hth first-stage stratum was calculated as

$$\pi_{hk} = n_{1h} \frac{S_{hk}}{S_{h+}}, \quad \text{for } k = 1,2,...,n_{1h} \text{ and } h = 1,2...,9, \tag{2}$$

where n_{1i} is the number of PSUs selected from the hth first-stage stratum and,
$S_{h+} = \sum_{k} S_{hk}$ is the total size measure of all PSUs in the hth first-stage stratum.

PSUs whose selection frequencies were equal to or larger than 1 were included as "certainty" PSUs and separate first-stage strata were created for them. An independent sample then was drawn from each first-stage stratum with probability proportional to size (PPS) using systematic sampling. Implicit stratification was achieved by sorting the first-stage frame before the sample selection.

Design of Second Stage Unit Sample and Overview of the Process

After selection of the 100 PSUs for the study, the process of recruiting the child welfare agencies associated with the PSUs began. As these agencies were recruited, we worked with them individually to refine our projections of the expected sizes of the domains of analysis for sampling in 1999-2000. From these projected domain sizes, the initial sampling rates by domain were specified. Software was developed to apply these sampling rates to the domains during the 15-month second stage sampling period.

The second stage sampling period began in October, 1999 and continued until December, 2000. This sampling period for the NSCAW was desirable for a number of reasons. First, the 15-month sampling period mitigated the effect of any cyclic patterns of abuse and neglect related to the seasons of the year, holidays, and school attendance. Secondly, it provided for a steady, level flow of cases to the field which was most efficient for the field work. Finally, it was sufficiently long to provide an adequate number of cases entered the system for all PSUs in the sample.

During the sampling period, agencies provided frame files, or, in some sites, interviewers visited each child welfare agency on a monthly basis for the purpose of selecting the sample of CPS children. Agency lists contained all investigations or assessments completed in the prior month, and the month prior to that. (For a more detailed description of the NSCAW I within-agency sampling frame data requirements, refer to **Volume II**, NSCAW Sampling Frame Data Request Specifications.) Sampling closed investigations and assessments avoided the problems of contacting caregivers and children during the time when the investigation was still active and their propensity to cooperate with the survey request quite low. Waiting until the investigations were completed also allowed us to use information on the type

of abuse or neglect, sometimes not known with accuracy until after the investigation is completed. We used the list of investigations closed two months prior to identify any cases that were delayed in being entered into the administrative database.   These lists of investigated children constituted the second stage sampling frame. Care was taken to ensure that each child eligible for the study was listed once and only once on the frame and that only children who were eligible for the study were listed. There was considerable duplication among lists and, thus, an important task was to un-duplicate lists so cases on multiple lists did not have higher probabilities of selection than others.

Once the relevant data on each eligible child were in the data base, the new entries were grouped by domain and sampled at a rate to yield the required number of sampled children in each age group. Initially, the number of children to sample in each PSU was set at 1/12 the annual sample size (i.e., approximately 6); however, the sampling rates by domain varied from PSU to PSU. The sampling rates required adjustment as the sampling process continued for several reasons:

- the sampling rates were based on previous year target population sizes and are, therefore, were subject to variation and error,

- the PSU composite size measures were subject to variation and error, thus, sampling rates needed revision to reduce unequal weighting effects,

- the number of children in the target population fluctuated from month to month, while the number of field interviewers in most PSUs remained constant, and

- response rates were different from what was anticipated, necessitating sample size adjustments to achieve the required number of interviews.

Each month, project statisticians reviewed the sample yields by domain for each PSU and determined whether the sampling rates required modification. If adjustments were necessary, modifications to the software were implemented and the field systems updated.

NSCAW II Sample Design

The NSCAW II sample design makes maximum use of the 92 primary sampling units (PSUs) used in the NSCAW I. In addition, we are recontacting agencies eliminated from the original NSCAW sample because of state laws that required the agency contact sampled families first; we have reason to believe that changes in laws and regulations will allow their participation in NSCAW II. Thus, the design specifies a total of 92-100 PSUs, depending on the outcome of discussions with the agency-first-contact states. *Volume II* contains the Sample Design Recommendation memo discussed with the Consultant Group in late May.

Using the NSCAW I PSUs is inarguably the most expedient approach to fielding the survey since, except for the seven agency first contact PSUs, the project has worked in these PSUs for many years, has an established and very experienced workforce in each, and has built a solid cooperative relationship with the child welfare agencies in the sites. Drawing a completely new sample of agencies would lengthen the schedule for getting into the field and would require the recruitment of a large number of new interviewers.

In addition, agency-level nonresponse as well as the incidence of other data-quality-related problems might be expected to be higher with a new sample of PSUs.

A number of technical issues were investigated before the decision to use the NSCAW I PSUs was finalized. For example, the precision of the NSCAW estimates could suffer from the use of PSUs selected using outdated size measures. The reduction in precision might be substantial if the distribution of the child welfare population has changed dramatically since 1998. Therefore, our evaluation examined the change in county-level child welfare populations from 1998 to the present and the effect these changes could have on the precision of NSCAW II results.

We have also conducted a cost-error optimization exercise aimed at determining the optimal number of PSUs to sample when costs and errors are simultaneously considered. As noted, we will use 92-100 PSUs, each having approximately 60 completed child interviews to achieve a total sample size of 5700 completed interviews. The within PSU sample size is essentially the same one used for the NSCAW I sample. It was primarily based upon the optimization of interviewer workloads since, in 1998; there were little or no data on variance components to do a proper cost-error optimization. Now, however, there is an abundance of data to revisit the question of sample size allocation and consider how the proposed design compares to an optimal one. For example, it is possible that fewer PSUs and a larger within PSU sample size could yield greater precision for essentially the same cost. If so, then it would be possible to reduce the number of PSUs while increasing the overall sample size for the study.

A third area of concern is frame coverage. The NSCAW I first stage sampling frame consisted of a list of all counties in the U.S. as well as data on the number of investigated children (by sampling domain). A total of 710 counties that, for NSCAW I, had fewer than 67 investigations for the year, were deleted from the county frame for cost and fieldwork efficiency reasons. This had a very slight effect on frame coverage of the child welfare population. However, for NSCAW II, it is possible that these 710 counties now have substantial numbers of investigations and account for a much greater fraction of the current child welfare population. If so, it would be important for the NSCAW II sample to represent these counties.

The issue of within PSU sampling is also an important area of investigation. The analysis domains of interest in NSCAW II are somewhat different than they were for NSCAW I. It is important that sufficient sample is achieved in the various domains of interest in order to provide sufficient power for analysis. However, as domains are sampled more and more disproportionately, the precision of a general population estimator can be substantially reduced as a result of the unequal weighting effect. In NSCAW I, a composite size measure was used to reduce this effect. In NSCAW II, this will not be possible if the NSCAW I PSUs are retained since the composite size measures are no longer calibrated for the current population. Oversampling of child domains is a crucial issue for the relevance of the NSCAW II results; but it can also cause substantial reductions in precision for analyses that cut across these domains. Therefore, it is important to achieve a balance between the objectives of subgroup analysis and general population analysis

A key issue for this work has been to predict how alternative NSCAW II sample design decisions will affect the precision of estimates. Having good data to predict the distributions of characteristics of the child welfare population in 2008 when the sample will be drawn is essential for optimal sample design. The nature of the NCANDS data, which is currently our best source of information for the sample design work,

has made this task particularly challenging. Unfortunately, the most recent NCANDS data available is from 2005 and state data quality is very poor in major regions of the country.

In addition, the statistical methodology for specifying many aspects of the design is both complex and nonstandard. Since textbook formulas do not apply, new statistical methodology had to be developed. Despite these limitations and complexities, we are reasonably confident in the sample design. The following considers four issues related to the NSCAW II design.

1. Effect on NSCAW II precision of carrying forward the 92 NSCAW I PSUs for use in NSCAW II.
2. The optimality of the proposed NSCAW II design in terms of costs and sampling error, assuming the NSCAW I PSUs are carried forward
3. Implications for the frame coverage of the NSCAW II target population using the NSCAW I PSUs without supplementing for non-represented extremely small counties.
4. Within PSU over-sampling and its implications for general population estimation precision.

Regarding (1), the main advantage of drawing a new sample is the potential for updating the selection probability size measures for changes in the child welfare population that have occurred since the NSCAW I sample was drawn. Therefore, we predicted the potential loss in statistical precision from using the NSCAW I PSUs in the NSCAW II rather than selecting a completely new PSU sample and estimated the changes in the child welfare population using 2005 NCANDS data. In this analysis, two alternative scenarios were considered.

For the first scenario, we assumed that the actual 2007 child welfare population sizes were known for every county in the U.S. The second scenario assumed that these data would be estimated with error. To obtain the magnitude of the error in projecting the 2007 size measures, we used our experience from NSCAW I. Our analysis assumed that the error in our 2007 county child welfare population projections would be essentially equal to the error in the projections we made in 1999 for selecting the NSCSAW I PSU sample. The NSCAW I projection error was then estimated from the 92 PSUs in the NSCAW I sample.

Under the first scenario we concluded that estimator precision could be substantially improved by selecting a new PSU sample. However, under the second scenario, we determined that these precision gains would not be realized due to the error in the population size projections. This analysis showed no significant gain in precision by selecting a new sample of PSUs.

In summary, assuming the frame we would use for selecting a new PSU sample would be of approximately the same quality as the one used for selecting the NSCAW I sample, there would be substantial errors in predicting child welfare population sizes at the county level. This is due, in part, to the most recent data available (i.e., the 2005 NCANDS) being two or more years out of date. In addition, the data contain errors and are missing for many counties. Consequently, a new sample could not be optimized and precision would deteriorate to the point the gains predicted under Scenario 1 would be eliminated. We concluded that there is no precision advantage for selecting a new sample of PSUs. Further, the costs and potential logistical problems of recruiting and maintaining a new sample of child welfare agencies would be substantial. It was our decision, therefore, that the 92 NSCAW I PSUs be retained for NSCAW II.

(Details of the analysis supporting this decision are provided in the Sample Design Recommendation Memo in **Volume II**.)

Regarding (2), we found that a sample of 92 PSUs is "near optimal" in the sense that it will yield estimates having precision close to that of the optimal design. Although gains in precision could be realized if the number of PSUs was increased to, say, 130 we estimate that the gains would be small (about 13%). Adding this number of PSUs to the design would increase agency recruitment and other costs considerably, beyond those with adding sample within existing PSUs. As a result, the overall achieved sample size (5700 interviews completed with children and caregivers) would have to be reduced substantially in order to stay within the current budget. This would have important adverse effects on subgroup analysis. Therefore, we decided on no change in the number of PSUs.

For (3), we calculated the proportion of the current child welfare population in the smallest counties that were deleted from the NSCAW I PSU sampling frame. We determined that the child welfare populations in these counties have grown since 1996. However, their growth has not reached the point where excluding them would have an important effect on population coverage. We estimate that adding a PSU to represent these small counties would only increase coverage about 1.3 percent. Since the current coverage of the target population is 98%, we see no advantage to further increasing frame coverage. Therefore it was decided that no additional PSUs will be selected from this group of counties for the NSCAW II sample.

Finally, with regard to (4), we developed the required formulas for optimizing the allocation of sample to the various domains of interest. The derivations of these formulas are provided in **Volume II**, as is an example of an allocation scheme that provides for oversampling of children in foster care, children receiving other services, infants, and teens. The sample size targets for these domains are more or less in line with those used in NSCAW I, except that in NSCAW I, teens were not oversampled and cases of sexual abuse were oversampled. In discussions with the Consultant Group it was decided to eliminate the oversample of teens.

In determining the domains to sample and how to optimally allocate the total sample, there were three considerations. First, in order to maximize sampling efficiency, it is preferable to select domains defined by variables that are available on the NCANDS data file. These data are important for projecting the sizes of the domains in each PSU which is essential for controlling the design effects. Using these data, we can project the total number of cases available within each domain which will form the basis of our initial allocation. As sampling progresses though the sampling year, these domain size estimates will be updated monthly. The domain allocations will be adjusted each month as well so that the design effects due to oversampling can be minimized.

Second, to reduce the effect of the oversampling on estimates which cut across domains (for example, estimates defined for the entire CW population), the number of domains as well as the extent of the oversampling for each domain should be minimized. Third, in allocating sample to the PSUs, it is important to consider the effect of the PSU allocations on interviewer workloads. For example, an optimal allocation may assign only 20 interviews to some PSUs which is considerably below the 60 interviews per PSU needed to optimize field costs. Therefore, the sample size allocation algorithm is designed so that the

minimum sample size in any PSU is 50 interviews. Likewise, the maximum number of interviews in any PSU is constrained to be less than 200 interviews. As a consequence of these constraints, the design effects were increased somewhat but not as much as they would be if every PSU were required to have exactly 60 interviews (as in NSCAW I).

Table C.1 in the May 2007 Sample Design Recommendation Memo (**Volume II**) provides an allocation scheme that was used in our analysis foe the NSCAW II Consultant Group to illustrate the effect of oversampling on precision. The estimated design effects associated with this scheme are provided in Table C.2. The overall design effect is estimated at 4.6, although this may be somewhat of an underestimate. Changing the allocation to domains can also change the design effect. Nonresponse, frame undercoverage and other weighting adjustments can inflate this number. In addition, it will not be possible to achieve exactly the sample sizes specified by the optimization algorithm. Much depends on the accuracy of projections of the domain sizes in each PSU. Work is currently underway to evaluate the proposed NSCAW II design in light of our expectations and realizations for NSCAW I, in order to provide better projections of the NSCAW II design effects.

Statistical Power. It is impossible to specify the statistical power of NSCAW analysis overall because it will include so many different research questions, variables and subpopulations. The proposed cohort size of 5,700 child interviews is slightly higher than the NSCAW I survey, which had 5,501 completed key respondent interviews. Based on experience with the NSCAW I analysis, this sample was adequate for many types of analysis that were conducted, both for cross-sectional and longitudinal analysis. As an example, for the descriptive estimates contained in Wave 1 Analysis Report (DHHS, 2005), the coefficients of variation (cv's) of the estimates never exceeded 20%. Further, for comparing estimates between independent subgroups or for comparing estimates of the same subgroup at two points in time, the precision of the comparisons was quite adequate for the eight planned domains of analysis as well as for a number of ad hoc domains.

To illustrate the statistical power of the NSCAW II sample design (Note: The allocation changed after the Consultant Group meeting in May 2007), **Exhibit B.1-2** shows sample sizes for the five sampling domains, their corresponding effective sample sizes, and the power for detecting small effect sizes. These allocations are approximations, and may change based on the achieved sample.   The effective sample sizes in the exhibit were computed as the actual sample size divided by the appropriate design effect. The design effect for the total sample is 5.6[1]. It varied by domain from 2.3 to 2.9. These design effects, which account for both clustering and the unequal weighting roughly on par with those observed for NSCAW I. Power is based on a two-sided test of size α = 0.05 comparing two independent means or proportions; for example, the mean total score on the CBCL for children in two different domains. These calculations assume that a "small effect size" (i.e., an effect size of 0.2 on Cohen's effect size scale; Cohen, 1988) is to be detected.

---

[1] The design effect of 5.6 is higher than the 4.6 design effect projected for the May 2007 Consultant Group Meeting. This is due to changing the allocation scheme used in the Table C.1 in the Sample Design Recommendation Memo (**Volume II**) to the one in Exhibit B.1-2. The allocation in this exhibit is the final allocation scheme that will be used in the study.

Exhibit B.1-2 Sample Size, Effective Sample Size and Power by Domain

NSCAW II Sampling Domain			Actual N	Effective N	Power for Small Effect Size
Age	Open or Closed	In Foster Care (FC)?			
Less than 1 year	Open	Yes	1,100	470	0.86
1-17.5 years	Open	Yes	1,100	432	0.84
All Ages	Closed	No	1,300	522	0.90
Less than 1 Year	Open	No	1,100	429	0.83
1-17.5 years	Open	No	1,100	384	0.79
All Domains			5,700	1,020	0.99

The power shown in *Exhibit B.1-2* represents a worst case scenario since we have assumed independent means and small effect sizes. For example, increasing the effect size from 0.2 to 0.35 (which is still in the small range according to Cohen's rule) increases the smallest power in the table from 0.79 to 0.99. For dependent means (e.g., change estimates) the power may be closed to 1. For example, assuming a moderate correlation between the means will increase smallest power in the table to from 0.79 to 0.99. For medium and larger effect sizes (i.e., effect sizes of 0.50 and 0.80, respectively) the power is close to 1 for testing for differences between independent means. This analysis suggests that even for small subgroups in the population, the design provides adequate power for a wide range of analyses.

B.2 Information Collection Procedures

This section describes the procedures for data collection; sampling procedures were described in detail in the previous section. Also addressed are weighting procedures and analytical techniques to be employed.

B.2.1 Data Collection Procedures

Procedures for data collection mirror closely those developed for and used successfully on NSCAW I.

Data Collection from Caseworkers

Baseline Risk Assessment from the Investigative Caseworker. This Computer-Assisted Personal Interview (CAPI) questionnaire focuses on the case investigation, the caseworker's assessment of risk at the time of the investigation, case history before the index investigation, details about the caregivers from whom the child was removed (if in out-of-home placement), and any referrals made or services received. Upon selection of the monthly sample, the field representative will identify the caseworker associated with each case and schedule the baseline interview. The field representative's call will emphasize the importance of the study, the critical nature of the caseworker's input, the brevity of the interview (about 45 minutes), and the flexibility she can have in accommodating the caseworker's busy schedule. To maximize convenience for the caseworkers and to safeguard confidential case record and interview data, caseworker interviews will be conducted at agency offices. After the call, the field representative will mail a letter explaining the study and a project brochure developed for caseworkers, with a reminder of the day and time the interview has been scheduled. (Drafts of the letter and project FAQ are included in *Volume II*.) The day before the scheduled interview, the field representative will phone the caseworker to provide another reminder, or to reschedule the interview if circumstances have arisen to preclude the conduct of the interview at the scheduled time. Field representatives will be trained to complete all caseworker baseline interviews within ten days after the monthly sample has been fielded in order to collect risk assessment data as close to the close of the investigation or assessment as possible.

18-Month Follow-Up Interviews. Eighteen months after the close of the investigation or assessment, the field representative will interview a service caseworker. This caseworker interview will focus on the services recommended for and received by the sampled child and her/his family, the case history after the report including any subsequent reports or investigations, the living environment in the household, caseworker involvement with the family, progress made by the family, and information about the caseworker. This information cannot be obtained in the earlier baseline interview because of the temporal proximity to the case closing and the need to allow subsequent service decisions and events to unfold. To facilitate the caseworker's preparations for these interviews, the field representative will send a letter with a list of topics to be addressed. Because of the detailed nature of many of the questions, we will request that the caseworker have the child's case record available for the interview session. The caseworker interview will average 60 minutes.

Collection of Data from the Adult Caregiver

Contacting Families and Children Selected into the Sample. Diversity in agency policies and their concerns about releasing names of clients, in NSCAW I we developed several different procedures by which the field representative acquired names and contacting information for selected children and families. In most agencies the liaison will provide us with names and contacting information for each child selected into the sample and the primary adult caregiver.

Different versions of the advance letter and project brochure have been prepared for the different types of caregivers living with the sampled children. (See draft letters and project FAQ included in *Volume II*.) These materials emphasize the importance of the study, ACF's sponsorship and the study's non-affiliation with the local CPS agency, the protections confidential interview data will receive, and the fact that participation in the study provides each family the opportunity to register their experiences.

Selection of the Adult Caregiver. In many instances there will be little question regarding who should be interviewed about the child. In other situations we will have to carefully sift through information to identify the most appropriate adult respondent. Regardless of the family situation, the guiding principle will be to interview the adult in the current household who knows the sampled child best and who can accurately answer as many of the questions as possible.

Specifically, for children who remain in the custody of their parent(s) and for whom there is more than one caregiver in the household, we will ask to interview the adult "most knowledgeable" about the child and who has co-resided with the child for two months or more. In the instance that there are multiple possible respondents who are most knowledgeable and meet the co-residency requirement, we will use the hierarchy of parent-child relationships employed on NSCAW I, the NIMH Methods for Epidemiology of Child and Adolescent Mental Disorders Study (MECA), and other studies of children and adolescents. This selection hierarchy typically results in the selection of the mother or mother figure in the child's living environment. In situations in which the child is in out-of-home placement such as foster care or kin care, we will seek to identify the adult in the household who is "most knowledgeable" about the child and has been co-resident with the child for at least two months. Less frequently we may find sampled children in therapeutic foster care or residential treatment centers; we will examine the specific details of these cases in order to appropriately weigh co-residency and other factors to choose the adult respondent for the sampled child.

Informed Consent Procedures. The adult respondent will be asked to consent to participate for both her/himself and the selected minor child. Written consent procedures will be repeated in each personal interview (baseline, 18 months) Draft consent forms, and authorization forms for teachers to release information and for data linkage are included with other project materials in *Volume II*.

Field representatives will be carefully trained to confirm with the caseworker that the adult respondent chosen has legal guardianship and the resulting legal right to consent to the child's participation. If the chosen adult respondent does not have guardianship rights, the field representative will identify and contact the person who does have the authority to consent for the child. In some sites the agency will have guardianship for out-of-home placement children; in other sites the family court or juvenile court may hold guardianship. The field representative will contact the guardian, explain the

study and the child's selection, and seek permission to interview the child and authorization for others (i.e., a teacher) to release information about the child.

In cases where a non-custodial parent retains guardianship and must be contacted to obtain informed consent for the child's participation, the field representative will talk with the caseworker to get her/his suggestions on the best approach. Depending on the specific case situation, the field representative may send a letter notifying the parent of his/her child's selection into the study sample and the need to discuss our request further, telephone the parent to deliver introductory information, or arrange to meet with the parent in person.

Interview Content. This CAPI questionnaire is focused on the child's health, mental health, services received by the child and the family, the family environment, and experiences with the child welfare system. The field representative will contact the adult caregiver and administer the first in-person interview approximately 40 days after the close of the investigation or assessment. The 18-month follow-up will be fielded two weeks before the anniversary of the investigation closing. The interview is expected to average an administration time of 96 minutes for each in-person caregiver interview.

Collection of Data from Sampled Children

Informed Consent for Selected Children. As noted above, we will obtain signed informed consent from the legal guardian of each sampled child before approaching that child for participation in the study. Additionally, we have assent forms for children 7 through 17 ½ years old and will read the appropriate assent form to the child before beginning the interview, as a means to introduce the child to the study, to assure the child that what they tell us will be kept confidential (with the exceptions surrounding expressed suicidal intent and suspected ongoing serious abuse), and to provide the child with an understanding of the voluntary nature of participation and their right to refuse to answer any question we ask of them. Given the vulnerability of this particular population, we believe it is especially important to provide these children with this information. From our NSCAW I pretest findings, we do not believe that it is feasible to go through informed assent procedures with children younger than 7, as they do not comprehend some of the fundamental concepts necessary to meaningfully process the information.

Collecting Data from Children. The sampled child will be interviewed in the same visit to the household as the adult caregiver. Once a signed consent form has been obtained from the legal guardian and the study has been explained to the adult caregiver (who may also be the same person), the field representative will seek to conduct a CAPI interview with sampled children and collect physical measurement and observation data for infants and toddlers. The timing of the adult caregiver and child interviews will vary by circumstances and the convenience of respondents; field representatives will schedule both interviews in the same visit to the household when possible (about 70% of NSCAW I interviews).

The interview protocol varies considerably depending on the age of the child. Only physical measures (length, weight, and head circumference) and physical development measures will be taken from the very youngest infants; older babies will be assessed through the Battelle Development Inventory. Toddlers and young children will complete several cartoon-based and other simple measures in addition

to the physical measures of height and weight. The interview protocol for older children includes questions on physical health, mental health, assessments of cognitive development and academic achievement, and for 11-14 year olds, questions in Audio Computer-Assisted Self Interview (A-CASI) mode about events that led to their involvement with the child welfare system and their relationship with caregivers. The A-CASI sections include questions on substance abuse, sexual activity, delinquency, injuries, and maltreatment. The interviews with sampled children will average 80 minutes, depending on age.

Special Sensitivities Necessary for Children in this Study. We will provide field representatives with special training on balancing the needs of data collection (e.g., keeping the child focused and on task and remaining emotionally detached and unbiased) with the needs of respondents (e.g., processing emotions evoked by the interview questions and feeling respected and supported) and their own needs (e.g., being confident in their ability to deal with the survey topics and specific sensitive questions and displaying that confidence without threatening or coercing respondents.)   In our NSCAW I experience, a respondent becoming distressed is a rare event, but we will provide field representatives with training on what to do if they encounter such a situation. Lists of local mental health resources will be provided to all adult respondents at the close of the interview.

We will also provide training on interviewing children and especially children who bring various challenges to the interview setting. The training will help field representatives become more aware of the issues of trust and confidence and of building rapport with children of all ages, and will provide them with skills for handling the situation if a child becomes upset during the interview.

The variety and different administration procedures for children of various ages, in addition to the content of the interview, will make this study a challenge for respondents and the field representatives alike. We will recruit, hire, and train persons with demonstrated experience with children, and preferably with experience administering assessments. We will recruit from the local pool of school counselors, social workers, pediatric nurses, teachers, and experienced interviewers.

Collection of Data from Teachers

The purpose of the teacher survey is to obtain an independent measure of the child's academic performance, cognitive abilities, social skills, and relationships with other children. In order to protect the child and parent respondents against their involvement with the child welfare system becoming known, we have created a different name to describe the NSCAW in the teacher survey—the National Teacher's Survey of Children and Adolescents. The survey of teachers will be implemented through a mailed self-administered instrument, with promptings of nonrespondents by mail, telephone, and email. A web option for the Teacher survey may be offered, if IRB approval is obtained. The teacher will be identified in the adult caregiver interview. Note, however, that teachers will only be contacted if the signed authorization form is obtained by the field representative from the legal guardian. This will insure that no teacher will be contacted for participation without the guardian's express approval.

Collection of Administrative Data

State and local agency administrators will be interviewed in late 2008 to obtain information regarding policies and procedures, the current context within which the agency is operating (e.g., presence of a consent decree), budget and resource allocation, characteristics of the caseworker staff, and other systems variables. These data will be used with the child and family data to examine possible relationships between family outcomes and agency structure and resource base. These interviews will average 45 minutes with the administrator or her/his designate.

The field representative will call the agency administrator to schedule an appointment to collect agency-level data. Noting the types of information we will be seeking, the field representative will also identify with the administrator other agency staff who can provide the more detailed data, in order to schedule interview appointments with those staff during the same visit to the agency.

Project staff will also work closely with agency IT staff to specify data files to be used for both sample selection and administrative data acquisition. To minimize burden on agencies, the study will use NCANDS data definitions and file structure. The vast majority of states already contribute data to NCANDS and have their database systems constructed to facilitate file construction for that purpose, and all submit quarterly Child and Family Service Review data files and annual AFCARS files. We are working with each state to define the process to acquire these data in a way that best meets their schedule, resource availability, and security requirements for data release. Details can be found in the Administrative Data File Specifications in *Volume II*.

B.2.2 Weighting Plan

Overview

This section describes the weighting procedures used in nSCAW I, to be mirrored in NSCAW II. The NSCAW sample design presents many significant departures from simple random sampling. First, the NSCAW sampling design is clustered by county thus introducing correlations in the observations that violate the independence assumptions of simple random sampling. Second, the sample is highly stratified according to initial case dispositions, child characteristics, and services received. Third, nonresponse and sample attrition may be related to outcome measures and case characteristics, and therefore cannot be ignored.

For any analysis of the NSCAW data from which inference is to be made to the target population, it would be inappropriate to ignore the complex sample design; therefore, analytic methods that incorporate survey weights or otherwise appropriately reflect departures from simple random sampling should be used. This requirement is routine for large federal demographic surveys.

Calculation of NSCAW I CPS Weights

This section describes the methods used in constructing the final Wave 1 analysis weight for the CPS sample. As noted, the NSCAW I CPS target population includes all children who were subjects of either an investigation or CPS agency assessment of child abuse or neglect, whether or not the investigation was founded or substantiated. In some sites selected for NSCAW, sampling unsubstantiated

cases was problematic for the CPS agency, bound by state law to maintain the privacy and confidentiality of the case files for unsubstantiated investigations. Thus, in some sites unsubstantiated cases were excluded from the sample. However, unlike the exclusion described for the agency contact sites, the weighting procedures include coverage adjustments that account for these missing frame components. Thus, inferences to the entire population of unsubstantiated cases−excluding the aforementioned agency-contact sites−are possible at the national level.

The weight variable for Wave 1 in the Restricted Release is used for making inference at the national level, which includes inference for both substantiated and unsubstantiated cases. A second weight for making inferences at the stratum level, including the eight key states, is also available on the file.

The analysis weight was constructed in stages corresponding to the stages of the sample design, with adjustments due to missing months of frame data or types of children, nonresponse, and undercoverage. This section describes the calculation of:

- the first stage (or PSU) weight

- the initial sampling, or base, weight,

- adjustments that were made to compensate for missing months on the sampling frame (i.e., jagged starts/stop or missing middle months)

- other adjustments for problems or special situations that were encountered in specific sites

- adjustment for nonresponse

- poststratification adjustment

Calculation of the First Stage Weight

Corresponding to the first stage of sample selection, the first stage (or PSU) weight was

constructed as the inverse of the expected inclusion frequency of the PSU. $W_{O,hi} = \dfrac{1}{\pi_{hi}}$ is the initial

base weight for the PSU i in stratum h. This weight was adjusted for the small counties that were combined after sample selection and for nonresponding PSUs to obtain the adjusted first stage weight $W_{1,hi}$.

Calculation of the Base Weights for Stratum and National Estimates

The base weight for a child in the CPS sample is the inverse of the probability of selection of the child. The purpose of this weight is solely to adjust the estimates for the differential probabilities that resulted in the sampling process. The probability of selection for a child is the product of two probabilities: the first stage selection probability and the second stage selection probability. The first

stage probability is the probability of selecting the PSU (county) of residence for the child, and the second stage probability is the probability of selecting the child given that child=s county of residence is sampled. The inverse of the first stage probability is called the first stage base weight and the inverse of the second probability is called the second stage base weight.

First, we define the following notation. Let

- h denote the sampling stratum for the primary sample selection where $h = 1$ for Key State 1, $h=2$ for Key State 2, $h = 3$ for Key State 3, and so on for the eight largest states, and $h = 9$ corresponds to the Remainder Stratum;
- n_h denote the number of PSU=s sampled in stratum h;
- i denote the PSU sampled within stratum h; i.e., $i = 1, ..., n_h$;
- d denote the sampling domain within each PSU where $d = 1, ..., 8$;
- m denote the month of the study where $m = 1, ..., 15$;
- n_{hidm} number of children sampled in month m in stratum h, PSU i, domain d;
- j denote the child sampled within PSU, domain, and month of the study where $j = 1, ..., n_{hidm}$;
- N_{hidm} denote the number of eligible children on the sampling frame in month m in stratum h, PSU i, and domain d; and
- π_{hi} denote the probability of selection for the ith PSU in stratum h.

Let π_{himdj} denote the probability of selection for the jth child selected in month m in domain d in PSU (h, i). Then,

$$\pi_{hidmj} = \pi_{hi} \times \frac{n_{hidm}}{N_{hidm}} \qquad (B.2.1)$$

which is the product of the first stage and second stage selection probabilities, is the overall probability of selection for the child. The inverse of this probability is defined as the base weight for a CPS sample child. However, several adjustments to this base weight were necessary.

In the CPS sampling process, the frame sizes varied from month to month, and in some cases the frame sizes, N_{hidm}, for a given month were very small or even zero. Note that when the domain size is 0, the probability in (B.2.1) is undefined. When the domain size is small, the probability, and consequently the sampling base weight, is unstable and can substantially increase the standard error of the estimates.

To solve this problem, we combined the domains across months so that rather than computing the base weight using the second stage selection probability in (B.2.1), we used n_{hid}/N_{hid} where $n_{hid} = \sum_m n_{didm}$ and $N_{hid} = \sum_m N_{hidm}$. This is equivalent to assuming that the entire 15-month sampling period was compressed into one sampling operation that selected n_{hid} children from the N_{hid} eligible children during the period. This approximation ignores the monthly variations in sampling rates and is therefore a much more stable quantity in most situations; but it still reflects an approximation to

true probabilities of selection and in that sense is not exact. However, any bias incurred by using this approximation is likely to be more than offset by the reduction in variance realized for the estimates. It should be noted that we anticipated using this type of weighting scheme in the design of the optimization algorithms employed in the allocation of the monthly samples.

Thus, taking the inverse of the selection probability in (B.2.1) after pooling over months, the NSCAW CPS sample base weight for all children j in stratum h, PSU i, domain d, regardless of the month of sampling, is

$$W_{B,hidj} = W_{1,hidj} \times W_{2hidj} \tag{B.2.2}$$

where $W_{1,hidj}$ is the first stage weight adjusted for nonresponse, and $W_{2,hidj} = N_{hid}/n_{hid}$ is the second stage weight.

To compensate for the exclusion of siblings of children selected in previous sample months but who are part of the target population, the frame size used, N_{hidm}, was the total number of children on the frame in month m including siblings but excluding the records for children appearing on previously sampled frame files. This adjustment combined with the post-stratification adjustments described later should be effective at reducing the coverage bias due to the exclusion of siblings.

To compute the second stage selection probability for children sampled from the PSU which provided paper reports, the total number of (unique) substantiated and unsubstantiated children appearing on each month's report prior to subsampling was needed. Knowing the exact number of unique children in each of the two domains was problematic, since that would require matching paper reports across the 15 months of sampling for the survey. Instead, we estimated the proportion of duplicate substantiated and unsubstantiated children in a typical county in the Remainder Stratum and used this number to estimate the number of unduplicated children in the PSU across the 15 months in order to obtain the population counts of children to be used for the numerator of the weight component.

Adjustments for Jagged Starts/Stops and Missing Middle Months

The quantity, N_{hid}, in the expression for $W_{2,hidj}$ in (B.2.2) is defined as the total number of children in the target population during the sampling period for the study, October 1999 though December 2000. Therefore, the inferential population for the NSCAW is actually all children investigated for child abuse or neglect during this 15-month sampling period.

In some PSUs no frame data were available for some months of the sample for various reasons. For example, in some PSUs, data collection was delayed until December 1999 or January 2000 and thus N_{hid} for these PSUs excludes the frame counts for months before sampling actually began. As a result, N_{hid} will be smaller than the target population sizes for these PSUs and, thus, the weight, $W_{2,hidj}$ will not appropriately reflect the total target population. Of the 92 PSUs, 4 were missing frame counts for one month and 11 were missing frame counts for more than one month.

The information on frame size by month was pooled across similar PSUs, and these monthly averages were used to estimate the missing frame sizes. The sample PSUs were divided into three groups according to three size categories: large, medium, and small, within approximately equal number of PSUs per group. Let g denote the size group, for $g = 1,...,3$. For a PSU, say i, in size group g with a missing frame count value for some month, say m, let N_{gmdi} denote the missing frame counts. We estimate N_{gmdi} for the domains $d = 1,..., 8$ as follows:

Let A_{gmd} denote the average frame count for all PSUs in size group g for month m, domain d and let A_{gd} denote the group average for domain d across all the months in the survey. Thus, the ratio A_{gmd}/A_{gd} is an adjustment factor that can be applied to any PSU in the group having a missing frame count for month m and domain d. Let A_{id} denote the average frame count for domain d in PSU i for all months available for PSU i. Then an estimate of the frame count for domain d in month m for PSU i is

$$\hat{N}_{gmid} = \begin{cases} \dfrac{A_{gmd}}{A_{gd}} \times A_{id} & \text{if } N_{gmid} \text{ is missing} \\ N_{gmid} & \text{if } N_{gmid} \text{ is not missing} \end{cases} \tag{7.3}$$

Thus, the base weight for all children j in stratum h, PSU i and domain d was adjusted by a factor, $f^{(1)}_{hidj}$ where

$$f^{(1)}_{hidj} = \frac{\displaystyle\sum_{m=1}^{15} \hat{N}_{hidm}}{\displaystyle{\sum_{m}}' N_{hidm}} \tag{7.4}$$

$\hat{N}_{hidm}$ is the estimate in (3) and where Σ' denotes the sum over all m for which N_{hidm} is not missing.

One way to compensate for missing months would have been to simply multiply the count N_{hid} by $15/m$, where m is the number of months actually sampled, to inflate it to a 15-month total. However, if the months of October and November 1999 are very different from the average month in terms of entries into the child welfare system, this type of adjustment would not reflect those differences and the adjusted total could differ considerably from the true total.

Another approach would have been to use the count from October 2000 as an estimate for the count in October 1999 and, likewise, for November 1999 use November 2000 if counts are available for these months. However, in small PSUs, where most of this type of missingness occurred, this type of adjustment is likely to be unstable and cause an increase in the contributions to variance from these PSUs.

Stratum and Site Specific Adjustments

For the majority of PSUs in the NSCAW sample, the frame files that were provided for sampling contained a complete listing of all children who were subjects of child abuse or neglect investigations in the previous month. In a few states, the files sent to us were incomplete in some respect and adjustments to the base weights were required to account for the loss of coverage resulting from these omissions. There were three types of missingness to consider:

• Unsubstantiated cases were excluded from the sampling frame due to legal issues in some states

• The sampling frames appeared to be missing a substantial percentage of unsubstantiated cases, presumably as a result of late data entry into the computer system in some states

• Cases that were not receiving services (Domains 1 and 2) were excluded from the study, as required in one county

This section first discusses the need for different adjustments depending upon the level of inference; i.e., national- or stratum-level. It then describes how we compensated for these frame coverage errors through various adjustments to the base weights.

Need for Different Adjustments Depending on the Level of Inference

When a group of cases, such as unsubstantiated cases, is completely missing from all sites within a stratum, the appropriate weight adjustment to compensate for the coverage error depends upon the level of inference. At the national level, post-stratification adjustments that account for frame noncoverage can be applied to the estimates of the unsubstantiated cases to account for the absence of these in one or more strata. Such adjustments can be very effective at reducing coverage bias because there are many PSUs in the national sample that contribute information on unsubstantiated cases, and that information can be used to estimate the characteristics for missing unsubstantiated cases in the national sample.

Adjustments for Strata Missing Unsubstantiated Cases from the Frame

In two Key States, the sampling frame for each month contained only children who were the subjects of a substantiated investigation. That is, children with unsubstantiated investigations were excluded. For inference at the national-level, we applied a coverage adjustment to account for the loss of unsubstantiated cases in these strata. Weights for sampled unsubstantiated children in the other key states were ratio adjusted to the total number of unsubstantiated cases in the key states including the two Key States. In order to compute this adjustment, we used estimates of the total number of unsubstantiated children in Key State 1 and Key State 2. Further, since child welfare characteristics are believed to differ by urban and rural areas, the totals were estimated separately by age categories for urban and rural PSUs. The data available from NCANDS Detailed Case Data Component (DCDC) were used to estimate the total for unsubstantiated cases in these two states. DCDC data were not available for unsubstantiated cases for the other dimensions of the domain definition. These weight adjustments are similar to applying a poststratification adjustment prior to the nonresponse adjustments. Folsom and Singh (2000) show that

this weighting strategy can improve the precision of the final estimates by reducing the variance in the controls totals used in the nonresponse adjustment.

Adjustments for Missing Unsubstantiated Cases in the Remainder Stratum

Some of the states in the Remainder Stratum were similar to the two Key States above in that only substantiated cases were sampled. Other states appeared to have very low numbers of unsubstantiated cases on the frame. Since the majority of the Remainder Stratum sites provided sampling frames with both substantiated and unsubstantiated cases, the national level estimates for the Remainder Stratum were calculated in the same manner to estimate the total number of children investigated at either level. We used the NCANDS DCDC and Summary Data Component (SDC) data to obtain estimates of the number of children in unsubstantiated cases for states missing unsubstantiated cases from their frame in order to calculate the national level weight. We poststratified the weights of children in the entire Remainder Stratum sample to the total number of children from unsubstantiated cases for those states from available NCANDS DCDC and SDC data.

Adjustments for Missing Children Receiving Services

In one PSU in one of the key states, we were not permitted to sample children who were not receiving some type of service from the local CPS agency. The calculation of the weights for that PSU included children on the sampling frames who were receiving services. Since the national-level estimate needed to include all children investigated, not just those receiving services, the weights for the other PSUs in that stratum were ratio-adjusted to the total number of children receiving services in the stratum. This adjustment was done separately by domain and urban/rural. The sampling frame for that PSU included all children investigated both receiving services and not receiving services. For this reason, we did not need to rely on external data sources to estimate the total number of children who were excluded from data collection.

Adjustment for Other Frame Undercoverage

An adjustment was made to account for a fairly significant frame undercoverage problem that was discovered in one of the PSUs. The undercoverage was discovered when frame counts were compared to the DCDC counts by PSU. We had a number of discussions with the data provider regarding this undercoverage problem, and discovered that the provider used a process to construct the frame that essentially deleted cases at random. We used data available from the DCDC that gave total counts and the distributions by age, gender, and substantiated/unsubstantiated and data from the state that gave the distribution of children by age and race/ethnicity. These data were used to adjust the sample weights within the PSU to agree with these counts and demographic distributions. Even though cases were missing at random, and not likely to cause bias in the estimates, the adjustment was made to help assure that any residual bias in the NSCAW data as a result of this frame problem will be minor.

Adjustment for Nonresponse

For the nonresponse adjustments for the weights for the CPS sample, we used a model-based method. The constrained logistical and exponential model proposed by Folsom and Witt (1994) allows

making the nonresponse and poststratification-type sample weight adjustment at the person level. This method is a generalization of the more traditional weighting class methods used for nonresponse adjustment and offers an attractive alternative to the often cumbersome and computationally intensive iterative proportional fitting (IPF) algorithm generally used for post-stratification, or raking.

The variables considered for inclusion in the model were those for which we had data on most of the children selected for the CPS sample. This allowed us to consider the use of any variables we had on the sampling frames, in the Case Initiation Database (CID, which was updated by the interviewer during the data collection process), or in the caseworker interviews. The variables on the frame include age, gender, race and ethnicity, a sexual abuse indicator, whether the child/family is receiving services, whether the case is substantiated, and whether the child is in out-of-home (OOH) placement. The CID data contains updated versions of these variables, and also the relationship to the caregiver.

From the list of the candidate predictors, variable selection methods were used to determine the significant predictors to be retained for the non-response constrained logistic model. We performed a nonresponse bias analysis of the variables from three sourcesCthe frame, the CID, and the caseworker. Some of the variables which appeared to be significant predictors of response were whether the child was receiving services, whether the child was in OOH placement, whether or not the case was substantiated, whether there was active drug abuse by the primary caregiver, recent history of arrests of the primary caregiver, prior reports of maltreatment, history of domestic violence, and whether the family had trouble paying for basic necessities. However, only about 60 percent of nonrespondents had caseworker data for these variables, and for this reason we chose not to use them in the nonresponse adjustment.

Separate response adjustment models were fitted for each of the strata. Dependent variables were those which form important analysis domains and which were found to be significant predictors of nonresponse. Variables used in the nonresponse adjustment models were age of the child, gender, race/ethnicity, sampling domain (which includes whether the child was receiving services and whether the child is in OOH placement), and urbanicity of the PSU.

Post-Stratification Weight Adjustment

Post-stratification methods are used to reduce the variation in the nonresponse adjusted weights and to reduce noncoverage bias. The weights are adjusted so that they sum to the external population total, generally within classes. The final weights were postratification-adjusted to counts of children available from NCANDS. Different target marginal totals were used for the national-level estimates. Thus, the nonresponse and post-stratification adjustments were carried out independently to produce the national-level weights.

To the extent possible, the post-stratification totals were counts of children available from NCANDS. For some states, data were available at the child level through the DCDC data file. This file contains indicators for substantiated versus unsubstantiated, and the age of the child. Encrypted child identifiers were on the file and it was possible to obtain counts of the numbers of unique children in each domain. For states that are not included on the DCDC file, totals from the SDC file were used. On the

SDC file, only the state by substantiated/unsubstantiated totals are available and, therefore, PSU-level totals cannot be computed. This means that for these states, the adjustments were slightly coarser than what is possible for states on the DCDC file. In some states on the SDC file, the totals are given as the number of children and in other states as the number of cases. Further, the totals sometimes count unduplicated children and sometimes children are counted multiple times. We used the NSCAW frame data to determine factors to use for estimating unduplicated counts of children.

The DCDC and SDC data give counts for 12 months; we ratio adjusted the counts by the factor 15/12 to obtain the poststratification totals used for the NSCAW CPS.

Nonresponse Compensation for Subsequent Follow-Ups

All nonresponse adjustments were fitted using the contractor's proprietary generalized exponential modeling procedure (GEM), which is similar to logistic modeling using bounds for adjustment factors. A key feature and advantage of the GEM software is that the nonresponse adjustment and weight trimming and smoothing are all accomplished in one step. With GEM, lower and upper bounds are set on the weight adjustment factors. The bounds can be varied, depending on whether the weight falls inside or outside a range, such as one defined by the (median - 3* interquartile range, median + 3*interquartile range). This allows different bounds to be set for adjustments for weights that are considered high extreme, low extreme, or nonextreme. In this way, the extreme weights can be controlled, and the design effect due to unequal weighting reduced.

Candidate predictor variables for the nonresponse adjustment models were those thought to be predictive of nonresponse and that were also available for most of the sample (nonrespondents as well as respondents).

To detect important interactions for the logistic models, a Chi-squared automatic interaction detection analysis (CHAID) was performed on the predictor variables. The CHAID analysis divided the data into segments that differed with respect to the response variable. The segmentation process first divided the sample into groups based on categories of the most significant predictor of response. It then split each of these groups into smaller subgroups based on other predictor variables. It also merged categories of a variable that were found to be insignificant. This splitting and merging process continued until no more statistically significant predictors were found (or until some other stopping rule was met). The interactions from the final CHAID segments were then defined.

For the CPS data, the interaction segments from CHAID were based on substantiated/unsubstantiated, type of abuse, sample month, sampling stratum, caseworker length of service, whether the family had trouble paying for basic necessities, the Youth Behavior Checklist, and the child Social Skills Rating System index.

The interaction segments from CHAID and all the main effects were subjected to variable screening in the GEM logistic procedure. The models initially included all of the potentially important variables that contained at least 30 respondents. The interaction segments identified by CHAID were also

retained in the models. The most nonsignificant variables were deleted, and deletion stopped when all variables were significant at level 0.10 or lower. As noted above, the GEM software allowed for different bounds to be set on the weight adjustments, depending on whether the weight was classified as high extreme, non-extreme, or low extreme, and for the nonresponse adjustment and truncation and smoothing to be achieved in one step.

In addition to those variables that were significant, we included other variables important for analysis in the model as main effects. This was because for terms that are included in the model, the GEM procedure has the property that weight sums for Wave 2 respondents using the adjusted weight will equal the original weighted total of the Wave 1 respondents. Keeping a large number of predictor variables in the nonresponse models allowed the estimates to be calibrated based on the respondents to as many totals as possible that were known for both respondents and nonrespondents. For example, as long as the model converged, we included the main effects of the variables age, race/ethnicity, gender, sampling domain, the substantiated/unsubstantiated indicator, and first-stage stratum so that the sums of the Wave 2 weights for the respondents for the marginals of these variables would be equal to the sums of the Wave 1 weights for all Wave 1 respondents.

Variables that were the most important predictors of nonresponse for CPS cases were the CHAID interaction segments, sampling domain, sampling stratum, who was responsible for sexual abuse or maltreatment, who contacted CPS to investigate, active alcohol use by parent or caregiver, active drug use by parent or caregiver, type of insurance coverage of the child, relationship of Wave 1 caregiver to the child, child setting, caseworker race/ethnicity, child's race/ethnicity, child's age, urbanicity of PSU, the standardized social skills score for children ages 3-5, and the child behavior checklist score for children ages 2-3. Variables that were the most important predictors of nonresponse for LTFC cases were the CHAID interaction segments and the caseworker age.

B.2.3 Analytic Techniques to Be Employed

In *Section A.2*, the research questions that will be addressed in the NSCAW analysis were described. The analysis of data from a stratified and clustered national sample is necessarily more complex and problematic than data from a sample selected using a simple random sample. Unfortunately, many statistical software packages that are readily accessible by researchers employ analysis techniques that assume simple random sampling. In spite of this, the benefits of using analysis methods that are appropriate for the sample design employed include improved statistical inference and less reliance on untenable assumptions which increases the robustness of the estimates. To support data licensees' use of the NSCAW I data, the project team has developed a manual that includes guidance on the application of the appropriate weight for the specific analysis, methods for imputing missing data, cross-sectional analysis, longitudinal analysis, and multilevel modeling.

B.3   Methods to Maximize Response Rates

NSCAW's ability to gain the cooperation of potential respondents and maintain their participation through the subsequent waves of the study is key to the success of this endeavor. In preparation for NSCAW I, we carefully reviewed procedures for obtaining respondent cooperation on a wide array of studies, incorporated the best practices of those studies into our data collection procedures, and adapted procedures for continuous improvement through the follow-up waves. The response rates achieved in NSCAW I are presented in *Exhibit B.3-1*.

Exhibit B.3-1 NSCAW I CPS Sample Component Response Rates, by Wave and Respondent Type

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**	**Wave 5***
Key Respondent (weighted) (Child if ≥11 years old, Current Caregiver if < 11 at Wave 1; caregiver or child Waves 2-5)	64.2%	82.4%	86.6%	85.3%	80.2%
Caseworker (unweighted)	92.6%	85.2%	94.0%	97.0%	94.9%

Teacher (unweighted)[2]	69.0%		67.8%	66.4%	83.2%

A caregiver's and child's willingness to participate, both initially and long-term, is affected by the combination of circumstances surrounding the nature of each selected case. These situations define whether our approaches for gaining cooperation are the more conventional and standard practices implemented on household-type surveys, or whether more resourceful and persistent actions are required. In cases in which the child remains in the environment of the reported neglect or abuse and with all non-custodial biological parents who retain legal guardianship we face more difficulty in obtaining cooperation from the caregiver and the child. Participant confidentiality is of course emphasized, as is the lack of association between the study and the child welfare agency and any other law enforcement agency. We also rely on obtaining information from the caseworker about the family situation in order to develop an effective approach to the potential respondents. Advance materials also emphasize that participation in the survey is an opportunity to provide information on how the system works based on their experiences. Additional NSCAW I procedures included:

- Field representatives thoroughly trained on NSCAW procedures and on the resources available and the processes to (1) overcome respondent objections, (2) resolve restricted access problems, (3) safely and successfully work in dangerous neighborhoods, and (4) reach difficult-to-contact respondents such as those seldom at home and teenagers

- Advance mailings of a letter on ACYF letterhead and a project brochure with information customized for that type of respondent on the frequently asked questions about the study

- Use of tailored letters addressing specific reasons for nonparticipation

- Review and approval of all noninterview cases by the field supervisor

- Sufficient numbers of bilingual interviewers so cases are not lost due to a Spanish-language barrier

Analysis of Nonresponse Bias in the NSCAW I

In addition to taking steps to maximize response from each type of respondent, the study team assessed the potential for nonresponse bias. The following section describes the analysis for Wave 1; analyses were conducted for Waves 2-4 and results are presented in Sections 7.9-11

---

[2] When the table was prepared in July 2007, the Teacher Survey had been completed only through the Wave 5 Young Child Follow-Up and therefore the 83.2% weighted response rate applies only to the Wave 5 Infant Follow-Up and the Wave 5 Young Child Follow-Up. We speculate that the higher response rate is partially a function of differences in participation propensities among teachers; it has been our experience that elementary school teachers are more likely to participate in the NSCAW Teacher Survey.

Since preparation of the Supporting Statement we have completed the Wave 5 Adolescent Follow-Up, and its Teacher Survey. Preliminary unweighted response rates are lower than the 83.1% achieved for the younger sample members but higher than the response rates obtained in previous waves. We attribute these increases to improvements in the project materials, and in implementation of additional prompting mechanisms. For example, as email addresses for teachers have become more widely available from both caregivers and on the internet, we now routinely use both telephone and email to prompt nonresponders.

in the *NSCAW Data File User's Manual.* The NSCAW Wave 1 CPS data were analyzed to address the following questions:

- Is the language in the consent forms discouraging respondents from giving complete and accurate information?

- Are item missing rates indicative that current caregivers are concerned about the repercussions of honest and complete answers?

The results of that investigation concluded the following:

- Although respondents may have been concerned about the privacy of their answers, there is no evidence to suggest a tendency for respondents to either falsify or withhold information, either as a result of the consent form or information from the interviewer. In addition, interviewers appear to be neutral collaborators in the interview, whose presence does not seem to have had a detrimental effect on honest reporting.

- Sensitive items are subject to significantly greater item nonresponse than non-sensitive items (98.2 vs. 99.8). However, for sensitive items, the item nonresponse rate is still less than 2 percent, which is negligible for most analyses. Therefore, the tendency for respondents to either actively or passively refuse to answer sensitive questions is quite small in the study.

Another investigation has been conducted in order to provide additional information on the extent of the bias arising from unit nonresponse—the failure to obtain an interview from a NSCAW sample member. An estimate of the nonresponse bias is the difference between the sample estimate (based only on respondents) and a version of the sample estimate based upon respondents and nonrespondents. In the NSCAW, a number of distinct data sources are used to obtain information on the sample child. When the sample child or caregiver did not respond to the survey, other data sources (such as the frame and caseworker data) can be used to provide information about them. Thus, it is possible to compare nonrespondents and respondents for some characteristics in order to investigate the potential nonresponse bias in the NSCAW results. In the remainder of this section, we briefly summarize the results of an investigation of the bias in the NSCAW results due to nonresponse using the data on nonrespondents available from other data sources.

An overall indicator of the severity of the bias due to nonresponse in the NSCAW is simply to count data items in our analysis for which respondents and nonrespondents differ significantly. Although this measure does not take into account either the type of comparisons that are significant or their importance for future analysis, it can be used as an indicator of the extent of the bias for general analysis objectives.

Variables used in this analysis were those that were also collected in the Wave 1 caseworker interview for the nonrespondents. However, only about 60 percent of the nonrespondents had a caseworker interview available. In this regard, the estimates of nonresponse bias are themselves subject to a bias due to incomplete information from caseworkers. However, we did not attempt to account for this potential bias in the analysis. These results assume that nonrespondents for whom caseworker information is unavailable are similar to nonrespondents for whom caseworker data is available.

Using the data collected for CPS and LTFC sample members from caseworkers at Wave 1, we estimated the bias due to using only the data for those with a key respondent interview. Let π denote the true average of the characteristic based upon the entire target population; i.e., π is the average value of C that we would estimate if we conducted a complete census of the target population. Thus, π is the target parameter that we intend to estimate with $\overline{y}_R$. Then bias in $\overline{y}_R$ as an estimate of π is simply the difference between the two, viz.,

$$B(\overline{y}_R) = \overline{y}_R - \pi \tag{1}$$

The bias can be estimated as follows. Let $\overline{y}_{NR}$ denote the estimate of the average value of C for the unit nonrespondents in the sample; i.e., $\overline{y}_{NR}$ is a computed as $\overline{y}_R$ but over the nonrespondents in the sample rather than the respondents. For example, we may have information on the characteristic C that is measured in the child interview from some other source such as the caseworker or caregiver interview or the sampling frame. If that is true, then $\overline{y}_{NR}$ can be computed. From this, we can form an estimate of π using the following formula:

$$\hat{\pi} = (1 - \eta)\overline{y}_R + \eta\,\overline{y}_{NR} \tag{2}$$

where η is the unit nonresponse rate for the interview corresponding to the characteristic C. Thus, an estimator of the bias in $\overline{y}_R$ is obtained by substituting $\hat{\pi}$ in (2) for π in (1). This results in the following estimator

$$\hat{B}(\overline{y}_R) = \overline{y}_R - \hat{\pi} \tag{3}$$

or, equivalently,

$$\hat{B}(\overline{y}_R) = \eta(\overline{y}_R - \overline{y}_{NR}) \tag{4}$$

That is, the estimator of the nonresponse bias for C is equal to the nonresponse rate for the interview that collects C times the difference in the average of C for respondents and nonrespondents.

We estimated these means and their standard errors using the weights and accounting for the survey design, as described in **Section B.2.2.** We estimated $\hat{\pi}$ using the unadjusted base weight. We estimated the mean for respondents, $\overline{y}_R$, in two ways: (1) using the unadjusted base weight, and (2) using the final adjusted analysis weight. This allowed us to see if the bias was reduced by applying the nonresponse and post-stratification adjustments to the weights.

We first tested the null hypothesis that the bias is 0 with α=0.05, i.e., H_O: Bias=0. We used a t-statistic for the test, and Taylor series linearization to estimate the standard errors. Variables with fewer than 20 cases in the denominators of the proportions or means were excluded from the analyses. Because of the dependencies in the tests, we used the largest k-1 categories when a variable had k levels. We counted the number of times that the null hypothesis was rejected.

Exhibit B.3-2 summarizes the results of this analysis. The analysis for children is for those who were key respondents (i.e. age 10 or older); this group of children was eligible to be interviewed and assent from them was necessary in order for the interview to proceed. In the CPS data, for the child interview, the number of tests that were deemed significant is slightly more than the number expected purely by chance (6.9 percent using the final analysis weight). This analysis indicates for the caregiver that there are more variables with significant bias than would be expected by chance (13.8 percent).

We examined the variables with significant bias. The biases, while statistically significant due to the large NSCAW sample size, were generally small and not practically significant. For this reason, we also tested a hypothesis of practical significance. We tested that the relative bias is small, and counted the number of times that the hypothesis was rejected. Specifically, we tested the null hypothesis H_O: |Relative Bias|<5 percent**,** where the relative bias is calculated as 100*Bias/ $\hat{\pi}$. **Exhibit B.3-2** shows the number of times that the null hypothesis was rejected at =0.05, using both sets of weights. This exhibit shows that for the CPS sample, with the final analysis weight, the number of variables with practically significant relative bias is four percent, or within the range of what would be expected by chance. Thus, we conclude that nonresponse bias in the CPS sample is unlikely to be consequential for most types of analyses.

Variables showing practically significant bias in the CPS sample were variables related to the type and severity of abuse/neglect, relationship of the primary caregiver to the child, likelihood of abuse/neglect in the next 12 months without services, child placement in a group

home, and the outcome of the investigation being substantiated. The actual bias in these variables was small (less than 10%).

Exhibit B.3-2 also shows the results for the LTFC sample. When using the final response adjusted analysis weight, approximately four percent of the tests that the bias is zero were significant at a five percent alpha, and less than one percent of the tests that relative bias is small were significant at a five percent alpha. This analysis also suggests that the bias was reduced by applying the nonresponse adjustment to the weights. Thus, there is no evidence of nonresponse biases in the LTFC data.

Exhibit B.3-2. Number of Significant Biases Observed by Type of Respondent for the CPS and LTFC Samples

Caregiver	CPS Sample		LTFC Sample	
	Base Weight	Final Analysis Weight	Base Weight	Final Analysis Weight
Items with more than 20 cases in the denominator	500	500	1,107	1,107
Items where H_O: Bias=0 was rejected	83 (16.6%)	69 (13.8%)	187 (16.9%)	50 (4.5%)
Items where H_O: |Relative Bias|<5% was rejected	33 (6.6%)	19 (3.8%)	32 (2.9%)	4 (0.4%)
Child	Base Weight	Final Analysis Weight	Base Weight	Final Analysis Weight
Items with more than 20 cases in the denominator	478	478	802	802
Items where H_O: Bias=0 was rejected	48 (10.0%)	33 (6.9%)	108 (13.5%)	33 (4.1%)
Items where H_O: |Relative Bias|<5% was rejected	45 (9.4%)	19 (4.0%)	26 (3.2%)	8 (1.0%)

Exhibit B.3-3 indicates that the response rate tends to be slightly lower for children in the LTFC sample component aged 11 to 14 than for children 10 or younger. This suggests that the potential for nonresponse bias is greater for older children and their caregivers. This effect of age on nonresponse was not apparent in the previous analysis because those data were analyzed

separately by key respondent type: child and caregiver. (For NSCAW, the caregiver was the key respondent when the child was less than 11 years old.) Therefore, the nonresponse bias results for children included only children who were at least 11 years old. Still, the lack of evidence for nonresponse bias in the previous analysis suggests that the greater relative bias for older children was quite small.

Exhibit B.3-3. Response Rates by Age of Child for the LTFC Sample at Wave 1

Age	# of respondents	% unweighted response rate	% weighted response rate
0 - 2 years old	246	76.64	78.94
3 - 5 years old	122	71.35	64.37
6 B 10 years old	196	73.41	76.07
11- 14 years old	163	69.07	69.41
TOTAL	727	73.07	73.41

B.4 Tests of Procedures

The NSCAW I study is the best test of procedures for preparation for NSCAW II. For both sample components combined through the 36-month follow-up (Wave 5 is still ongoing), we had completed 21,914 interviews with caregivers, 16,015 interviews with sampled children, 14,495 interviews with investigative and services caseworkers, and received 5,046 completed mail surveys from teachers. This experience, coupled with commitment to continuous improvement, is the best test. In preparation for NSCAW II, we will complete a full-systems test with 9 respondents of each type. These efforts will help us find the final instrument application problems, and the final tweaks to project materials such as advance letters and fact sheets.

B.5 Statistical Consultants

Statistical consulting was provided by
J.N.K. Rao, Ph.D.
Mathematics Department
Carleton University
1125 Coloney By Drive
Ottawa, Ontario K1S 5B6
(613) 788-214