

Weighting and Adjustment for Non-Response

Calculation of NSCAW II Wave 1 Weights

This section describes the methods that will be used in constructing the NSCAW II Wave 1 analysis weight. As noted, the NSCAW II 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 both NSCAW I and NSCAW II, 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 to 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 so-called Restricted Release¹ file will be used for making inference at the national level, which includes inference for both substantiated and unsubstantiated cases. In NSCAW I, a second weight for making inferences at the stratum level, including the eight key states, was also available on the file. This weight will not be created for the NSCAW II since the sample design did not provide for state-level estimates.

The analysis weight will be 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 Base Weights

The base weight for a sample child 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.

¹ The Restricted Release file contained confidential information such as PSU identification number and some child demographic characteristics that increase the risk of re-identification of a sample member. However, this file was only available to institutions having IRBs and after considerable assurances to maintain confidentiality of the data. These variables were masked on the Public Release file.

First, we define the following notation. Let

n denote the number of PSUs sampled;

i denote the i th PSU sampled;

d denote the sampling domain within each PSU where $d = 1, \dots, 5$;

m denote the month of the study where $m = 1, \dots, 15$;

n_{idm} number of children sampled in month m in PSU i , domain d ;

j denote the child sampled within PSU, domain, and month of the study where $j = 1, \dots, n_{idm}$;

N_{idm} denote the number of eligible children on the sampling frame in month m in PSU i , and domain d ; and

π_i denote the probability of selection for the i th PSU.

Let π_{idmj} denote the probability of selection for the j th child selected in month m in domain d in PSU i . Then,

$$\pi_{idmj} = \pi_i \times \frac{n_{idm}}{N_{idm}} \quad (\text{B.2.1})$$

which is the product of the first stage and second stage selection probabilities, is the overall probability of selection for the child. Note that this probability is the same for all children, j , in the same domain, d . The inverse of this probability is defined as the base weight for a sample child. However, several adjustments to this base weight will be necessary.

In the sampling process, the frame sizes have varied from month to month, and in some cases the frame sizes for some domains in a given month, N_{idm} , 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, can be very large and unstable thus substantially increasing 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_{id}/N_{id} where $n_{id} = \sum_m n_{idm}$ and $N_{id} = \sum_m N_{idm}$. This is equivalent to assuming that the entire 15-month sampling period was compressed into one sampling operation that selected n_{id} children from the N_{id} 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. However, any bias incurred by using this approximation is likely to be more than offset by the reduction in variance realized for the estimates by stabilizing the weight calculations. It should be noted that we anticipated using this type of weighting scheme in the design of the optimization algorithms being 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 sample base weight for the j th child in domain d and PSU i regardless of the month sampled is

$$W_{B,ijd} = W_{li} \times W_{2id} \quad (\text{B.2.2})$$

where W_{li} is the first stage weight – i.e., $W_{li} = 1/\pi_i$ – and $W_{2id} = N_{id}/n_{id}$ is the second stage weight which is the same for all children j in domain d of the i th PSU.

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_{idm} , 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 and $N_{id} = \sum_m N_{idm}$. This adjustment combined with the post-stratification adjustments described later should be effective at reducing any potential coverage bias due to the exclusion of siblings.

Adjustments for Jagged Starts/Stops and Missing Middle Months

The quantity, N_{id} , in the expression for $W_{2,ijd}$ in (B.2.2) is defined as the total number of children in the target population during the sampling period for the study, February 2008 though April 2009. Thus, the inferential population for the NSCAW II is all children investigated for child abuse or neglect during this 15-month sampling period.

In some PSUs no frame data have been available for some months of the sample for various reasons. For example, in some PSUs, data collection was delayed until May or June 2008 and thus N_{id} for these PSUs excludes the frame counts for months before sampling actually began. As a result, N_{id} will be smaller than the target population sizes for these PSUs and, thus, the weight, $W_{2,ijd}$ will not appropriately reflect the total target population. To address this issue, information on frame size by month will be pooled across similar PSUs and these monthly averages will be used to estimate the missing frame sizes. The sample PSUs will be 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 $g = 1, \dots, 3$. 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 $d = 1, \dots, 5$ 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

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

Thus, the base weight for all children j in PSU i and domain d will be adjusted by a factor, $f_{idj}^{(1)}$ where

$$f_{idj}^{(1)} = \frac{\sum_{m=1}^{12} \hat{N}_{idm}}{\sum_m N_{idm}} \quad (7.4)$$

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

One way to compensate for missing months is to simply multiply the count N_{id} by $15/m$, where m is the number of months actually sampled, to inflate it to a 15-month total. However, if the missing months 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 is to use the count from the same month in 2007 as an estimate for the missing month's count in 2008. 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.

Post-stratification Adjustments

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

1. Unsubstantiated cases excluded from the sampling frame due to legal issues
2. The sampling frames appeared to be missing a substantial percentage of cases for whatever reason. In some states, this is the result of late data entry into the computer system.

For (1), post-stratification adjustments that account for frame noncoverage will be applied to the estimates of the unsubstantiated cases to account for the absence of these in one or more states. Such adjustments can be very effective at reducing coverage bias because there are many PSUs in the 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.

For (2), an adjustment will be made to account for any significant frame undercoverage problem that we discover in one or more PSUs. Frame undercoverage will be identified by comparing the sampling frame counts to the Detailed Case Data for Children (DCDC) counts by PSU. We will use data available from the DCDC on total counts by age, gender, and substantiated/unsubstantiated and race/ethnicity. These data will be used to adjust the sample weights within the PSU to force agreement with these counts and demographic distributions.

Adjustment for Nonresponse

Nonresponse weight adjustments will use 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 will be those for which we have data on most of the children selected for the sample. This will allow us to consider the use of any variables we had on the sampling frames, or in the Case Initiation Database (CID, which was updated by the interviewer during the data collection process). The variables on the frame include age, gender, race and ethnicity, 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 child's relationship to the caregiver.

From the list of the candidate predictors, variable selection methods will be used to determine the significant predictors to be retained for the non-response constrained logistic model. We will perform a nonresponse bias analysis of the variables from two sources: the frame and the CID. In NSCAW I, some of the variables that were significant predictors of nonresponse included whether the child was receiving services, whether the child was in OOH placement, and whether or not the case was substantiated,. All of the variables will be explored for the NSCAW II adjustment.

Final Post-Stratification Weight Adjustments

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 will be post-stratified to counts of children available from NCANDS. For most states, data are available at the child level through the DCDC data file. This file contains indicators for substantiated versus unsubstantiated, and the age of the child. Using the encrypted child identifiers on the file, it will be 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 will be used

All nonresponse and post-stratification adjustments will be computed 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.

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

In NSCAW I, 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. These interactions 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 allows 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.

Variables that were the most important predictors of nonresponse in NSCAW I were the CHAID interaction segments, sampling domain, type of perpetrator of abuse or maltreatment, person contacting 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's placement type, 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.

Possible Special Adjustment for NSCAW II State Drop-Outs

As previously noted, four states that participated in NSCAW I declined to participate in NSCAW II citing new agency-first-contact requirements that were not in place in NSCAW I. The weighting procedures above do not attempt to adjust for any states that were declared out-of-scope for NSCAW by virtue of their state policies. However, in the case of the four aforementioned states, there is a wealth of information available from the NSCAW I interviews, unlike other agency first contact states that were excluded from both NSCAW I and NSCAW II. Add these data to the data from other sources such as NCANDS and DCDC, and it becomes quite feasible to construct an adequate weighting adjustment that would bring these states into the NSCAW II inferential population. Such an adjustment has not yet been explored in detail, but will be in the NSCAW II weighting process.

If it is possible to extend the NSCAW II inferential population to include these four agency first contact states, the benefits are improved coverage (94.8% versus 87%) as well as more valid comparisons between NSCAW II and NSCAW I which includes these states. The adjustment process would model the differences between these four states and the remainder of the CW population for a number of key analysis variables using NSCAW I data for the modeling process. Then a set of weights adjustments would be constructed that minimize these differences using weight calibration models (see, for example, Deville and Sarndal, 1992). These models would also incorporate external data (NCANDS, etc) on the differences that are available for the NSCAW II sample.

To illustrate, let *A* denote the states that participated only in NSCAW I and let *B* denote the states participating in both 1996 and 2008. Then the data available for calibration or nonresponse adjustment is displayed in the following table:

Survey	State Group A	State Group B
NSCAW I	y_{A1}, x_{A1}, z_{A1}	y_{B1}, x_{B1}, z_{B1}
NSCAW II	z_{A2}	y_{B2}, x_{B2}, z_{B2}

where *y*'s denote the characteristics of interest, the *x*'s are characteristics highly correlated with *y* from the questionnaire and the *z*'s are frame variables that may not be as highly correlated with the *y* as the *x*'s. Let Y_{ij} denote the total for y_{ij} and let N_j denote the population total for $i = A, B$ and $j = 1, 2$. We want to estimate $\bar{d}_y = \bar{y}_1 - \bar{y}_2$ where

$$\bar{y}_1 = w_{A1}\bar{y}_{A1} + w_{B1}\bar{y}_{B1} \text{ and the weights } w_{A1}, w_{B1} \text{ are such that } E(\bar{y}_1) = \frac{Y_{A1} + Y_{B1}}{N_1} \text{ and where } E(\bar{y}_2) = \frac{Y_{A2} + Y_{B2}}{N_2}.$$

Imputation methods can be used to estimate the missing Y_{A2} totals using the *y*, *x* and *z* variables. However, this method would require a separate model for each *y* variable. Alternatively, weight calibration can be applied to obtain a set of

weights w_{B2idj} such that $\sum_{ij} w_{B2idj} Z_{B2idj} = Z_{A2d} + Z_{B2d}$ for NSCAW II domain totals Z_{A2d} and Z_{B2d} . Then these weights can be applied to the y -variables to obtain estimates of $E(\bar{y}_2)$ for any y .

Wave 2 Weighting

For weighting purposes, respondents at Wave 2 are defined as those children who had either a Wave 2 caseworker or caregiver interview. A set of Wave 2 weights will be constructed by applying an adjustment for nonresponse to the final analysis weights for Wave 1. This Wave 1 weight is constructed by first calculating the selection probabilities for each of the sampled children, and then applying a series of adjustments for undercoverage and nonresponse. An adjustment of the final Wave 1 weights for the additional Wave 2 nonresponse will result in a Wave 2 response-adjusted analysis weight. This weight variable is appropriate for producing national estimates from the Wave 2 datasets.

Adjustment for Wave 2 Nonresponse

Wave 2 nonresponse adjustments will be computed using RTI's GEM modeling procedure. This procedure is similar to logistic modeling using bounds for adjustment factors. Since all the Wave 2 sample members are also Wave 1 respondents, the predicted response probability for Wave 2 could be written as the product of the probabilities of response for Wave 1 and Wave 2 conditional on Wave 1:

$\Pr(\text{unit } j \text{ responds at Wave 2}) =$

$\Pr(\text{unit } j \text{ responds at Wave 2 and Wave 1}) =$

$\Pr(\text{unit } j \text{ responds at Wave 2} | \text{unit } j \text{ responds at Wave 1}) \times \Pr(\text{unit } j \text{ responds at Wave 1}),$

where "Pr()" denotes the probability of the event. The Wave 1 analysis weights are constructed by multiplying the inverse of the second term, $1/\Pr(\text{unit } j \text{ responds at Wave 1})$, by the NSCAW base weights. Thus, the Wave 1 analysis weights are used and adjusted for the additional nonresponse at Wave 2 in order to obtain Wave 2 analysis weights.

Further, by applying the adjustment to the Wave 1 analysis weight, data collected from the Wave 1 respondents (as well as frame data) could be used in the adjustments.

The universe for the modeling of the Wave 2 nonresponse adjustment is the set of Wave 1 respondents; i.e., nonrespondents at Wave 1 will not be included in the Wave 2

nonresponse adjustment model. The Wave 2 adjustment is then applied to the final Wave 1 analysis weights. Let w_{Fidj} denote the final Wave 1 weight for the j th child in domain d and PSU i computed as described above. Denote $\Pr(\text{unit } idj \text{ responds at Wave 2} | \text{unit } idj \text{ responds at Wave 1})$ estimated from the GEM model by p_{idj} and let $w_{idj}^* = p_{idj}^{-1}$. Then the Wave 2 weight is given by

$$w_{2idj} = w_{idj}^* w_{Fidj}$$

Finally, to ensure that the weighted Wave 2 totals agree with the weighted Wave 1 totals for all strata and sampling domains used in the sampling design, marginal effects will be included in all the nonresponse adjustment models.

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.

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 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 \mathcal{T} denote the true average of the characteristic based upon the entire target population; i.e., \mathcal{T} is the average value of C that we would estimate if we conducted a complete census of the target population. Thus, \mathcal{T} is the target parameter that we intend to estimate with \bar{y}_R . Then bias in \bar{y}_R as an estimate of \mathcal{T} is simply the difference between the two, viz.,

$$B(\bar{y}_R) = \bar{y}_R - \mathcal{T} \quad (1)$$

The bias can be estimated as follows. Let \bar{y}_{NR} denote the estimate of the average value of C for the unit nonrespondents in the sample; i.e., \bar{y}_{NR} is computed as \bar{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 \bar{y}_{NR} can be computed. From this, we can form an estimate of \mathcal{T} using the following formula:

$$\hat{\mathcal{T}} = (1 - \eta) \bar{y}_R + \eta \bar{y}_{NR} \quad (2)$$

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

$$\hat{B}(\bar{y}_R) = \bar{y}_R - \hat{\mathcal{T}} \quad (3)$$

or, equivalently,

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

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 $\hat{\mathcal{T}}$ using the unadjusted base weight. We estimated the mean for respondents, \bar{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 $\alpha=0.05$, i.e., $H_0: \text{Bias}=0$. 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 $H_0: |\text{Relative Bias}| < 5 \text{ percent}$, where the relative bias is calculated as $100 * \text{Bias} / \hat{\pi}$. **Exhibit B.3-2** shows the number of times that the null hypothesis was rejected at $\alpha = 0.05$, 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 $H_0: \text{Bias}=0$ was rejected	83 (16.6%)	69 (13.8%)	187 (16.9%)	50 (4.5%)
Items where $H_0: \text{Relative Bias} < 5\%$ 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 $H_0: \text{Bias}=0$ was rejected	48 (10.0%)	33 (6.9%)	108 (13.5%)	33 (4.1%)
Items where $H_0: \text{Relative Bias} < 5\%$ 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