Chapter 3
WEIGHTING AND POPULATION ESTIMATES

Leyla Mohadjer, John Burke, James Green, and Joseph Waksberg; Westat, Inc.

3.1 GOALS OF WEIGHTING

Sample weights were produced for National Adult Literacy Survey respondents who completed the exercise booklet; those who could not start the exercises because of a language barrier, a physical or mental disability, or a reading or writing barrier; and those who refused to complete the exercises but had completed background questionnaires. Separate sets of weights were computed for the incentive and non-incentive samples (refer to section 2.3 for a description of the non-incentive sample).

The purpose of calculating sample weights for the National Adult Literacy Survey was to permit inferences from persons included in the sample to the populations from which they were drawn, and to have the tabulations reflect estimates of the population totals. Sample weighting was carried out to accomplish the following five objectives:

1) To permit unbiased estimates, taking account of the fact that all persons in the population did not have the same probability of selection;
2) To combine the state and national samples in an efficient manner;
3) To bring data up to the dimensions of the population totals;
4) To use auxiliary data on known population characteristics in such a way as to reduce sampling errors; and
5) To minimize biases arising from differences between cooperating and non-cooperating persons in the sample.

Objective 1 was accomplished by computing base weights for the persons selected into the sample. To produce unbiased estimates, different weights must be used for various subsets of the population, whenever these subsets have been sampled at different rates. Weighting was required to account for the oversampling of Black and Hispanic persons in high-minority segments of the national sample. Furthermore, the survey specifications called for the selection of one person in households with fewer than four eligible members and two persons in households with four or more eligible members. Using this approach, members of households with only one eligible member had twice the chance of selection of those in households with two eligible members, three times the chance of selection of those in households with three eligible members, etc. Weighting was needed in these situations to prevent potentially serious biases.

The base weight was calculated as the reciprocal of a respondent’s final probability of selection. For the household sample, it was computed as the product of the inverse of probabilities of selection at the primary sampling unit (PSU), segment, household, and person levels. For the prison sample, the base weight was equal to the reciprocal of the product of the selection probabilities for the facility and the inmate within the facility. Section 3.2.2 provides a summary of the base weight computation.

The second objective of weighting was to provide composite weights for the respondents in the 11 state samples and the respondents in the national sample PSUs in the 11 states. The national and state components applied the same sampling procedures in terms of stratification method, PSU construction, sample design, and selection at the various stages of sampling. Furthermore, the same forms were used to screen households and to collect background information and literacy assessment data in the state and national surveys. To take full advantage of this comparability, the samples were combined to produce both state- and national-level statistics. The advantage of compositing the samples was the increased sample size, which improved the precision of both state and national estimates. It should be noted that composite estimates apply only to persons ages 16–64, because data for persons age 65 and older came only from the national sample. Section 3.2.4 describes the composite estimation procedures used for the National Adult Literacy Survey.

For the household components, the post-stratified base weight was multiplied by a compositing factor that combined the national and state component data in an optimal manner, considering the differences in sample size and sampling error between the two components. Up to four different compositing factors were used in each of the 11 participating states, and a pseudo factor (equal to 1) was used for all persons age 65 and older and for national component records from outside of the 11 states. The product of the post-stratified base weight and the compositing factor for a record was the composite weight. A particular state analysis can include data from all respondents, age 16 and older, in that state. However, the sampling error for state estimates will increase with the inclusion of records for respondents over age 64, because these records came from the national component only.

Objectives 3, 4, and 5 were accomplished in one step by adjusting for nonresponse through post-stratification and raking[1] to adjusted 1990 census totals. If every selected household had agreed to complete the screener, and every selected person had agreed to complete the background questionnaire and the exercise booklet, weighted estimates based on the data would be approximately unbiased (from a sampling point of view). However, nonresponse occurs in any survey operation, even when participation is not voluntary. The best approach to minimizing nonresponse bias is to plan and implement field procedures that maintain high cooperation rates. For example, the payment of a $20 incentive in the household survey and repeated callbacks for refusal conversion were very effective in reducing

[1]Raking is a special kind of poststratification in which the weights of the adjustment cells are adjusted in such a way that the weighted sample marginal totals correspond to known population totals.

nonresponse, and thus nonresponse bias. However, because some nonresponse occurs even with the best strategies, adjustments are always necessary to avoid potential nonresponse bias.

Although the data collection was carried out in 1992, adjusted 1990 census data were used for poststratification. Undercount rates estimated by the U.S. Bureau of the Census were applied to the 1990 census count to correct for the undercoverage of some population subgroups. It was concluded that the estimates would not have been improved by extrapolating 1990 census data to the 1992 estimates of the population.

The composite weights were raked so that numerous totals calculated with the resulting full sample weights would agree with the 1990 census totals, adjusted for undercount. The cells used for the raking were defined to the finest combination of age, education level, and race/ethnicity that the data would allow. Raking adjustment factors were calculated separately for each of the 11 states and then for the remainder of the United States. Section 3.2.5 describes the details of the poststratification and raking approaches. Demographic variables that were critical to the weighting were re-coded and imputed, if necessary, before the calculation of base weights.

Full-sample and replicate weights were calculated for each record to facilitate the computation of unbiased estimates and their standard errors. The full-sample and replicate weights for the household components were calculated as the product of a record's post-stratified base weight and a compositing and raking factor.

The weighting procedures were repeated for 60 strategically constructed subsets from the records in the sample to create a set of replicate weights for variance estimation using the jackknife method. The replication scheme was designed to produce stable estimates of standard errors for the national and 11 individual state estimates.

The full-sample and replicate weights for the prison component were calculated as the product of a record's base weight and a nonresponse and raking factor. The base weight was calculated as the reciprocal of the final probability of selection for a respondent, which reflected the two stages of sampling (sampling facilities and sampling inmates within facilities). The base weights were then adjusted for nonresponse to reflect both facility and inmate nonresponse. The resulting nonresponse-adjusted weights were then raked to agree with independent estimates for certain subgroups of the population.

3.2 CALCULATING SAMPLE WEIGHTS FOR THE HOUSEHOLD POPULATION

3.2.1 Preliminary Steps in Weighting

The data used in weighting underwent edit, frequency, and consistency checks to prevent any errors in the sample weights. The checks were performed on fields required for data weighting and were limited to records that required weights (i.e., records for respondents who completed the exercise booklet and those who failed to complete a screener.

The consistency checks also helped to identify any unusual values. Listings were prepared of records with missing values in any of the fields used in weighting. The listings showed the entire record: the respondent's identification number, age, date of birth (from the background questionnaire), sex, race/ethnicity, level of education, the race of the head of household, and the number of age-eligible members and respondents in the household. The printed listings were used to review the extent of missing data, identify the pattern of missing data, and prepare for imputation. The sex and race/ethnicity data from the screener and background questionnaire were also compared for consistency. Overall, these checks found little missing data and very few records with values that differed between the screener and the background questionnaire.

Most of the fields required for data weighting (race/ethnicity of the head of household; sex, age, race/ethnicity and education of the respondent) were at finer levels of detail than were necessary for the later steps of weighting. The data in these fields were, therefore, collapsed to the required levels. Most of these fields were present in both the screener and the background questionnaire, thereby providing two measures of the same item. The background questionnaire measure was preferred for all items except the race of the head of household, which was collected only on the screener. For the few cases in which the background questionnaire measure was missing, the screener measure was generally available and was used as a direct substitute. Frequencies were prepared for each item after collapsing and making direct substitutions to gauge the magnitude of the imputation task.

The amount of missing data remaining after substitution was small, making the imputation task fairly straightforward. The Westat imputation macro WESDECK was used to perform hot-deck imputation for particular combinations of fields that were missing. Imputation flags were created for each of the five critical fields to indicate whether the data were originally reported or were based on substitution or imputation via WESDECK. The imputed values were used only for the sample weighting process.

Several special cases required attention before the calculation of base weights. In some dwelling units, the number of eligible household members exceeded nine, the maximum allowable number on preprinted labels used by the interviewers for respondent selection. In these instances, field staff provided the total number of eligible household members to the main office, where statisticians randomly selected respondents for interview and relayed this information back to the field staff. Detailed records indicated the PSU, segment number, total number of eligible household members, and number of respondents selected in each dwelling unit. This information was retrieved and attached to each of these records before the calculation of base weights.

Some additional dwelling units came into the sample as part of the missed structure and missed dwelling unit procedures (refer to section 2.2.3.3 for more information), which allow units that were missed in the segment listing activities to be included in the sample with a known probability of selection. All missed dwelling units within a segment were included unless the total number of missed units in the segment was unusually large, in which case a sample of missed dwelling units was taken. Detailed records indicated the PSU, segment, number of missed dwelling units selected, and total number of missed dwelling units whenever a sample of missed units was selected. This information was retrieved and attached to each of these records prior to the calculation of base weights.

A few final checks were run before base weight calculation to ensure the availability and validity of all fields required by the base weights program (fields created for the special cases mentioned above and fields for the total number of age-eligible household members and the number of sample persons for each dwelling unit). A detailed description of base weight computation is provided in the next section.

3.2.2 Computing Base Weights

A base weight was calculated for each record. The base weight was initially computed as the reciprocal of the product of the probabilities of selection at each stage of sampling (as given in section 2.2.3.2). The base weight reflected the probabilities of selection at the PSU, segment, dwelling unit, and respondent levels. The final base weight included adjustments to reflect the selection of the reserve sample (see section 2.2.2), the selection of missed dwelling units (see section 2.2.3.3), and the chunking process conducted during the listing of the segments (section 2.2.2.5), and to account for the subsample of segments assigned to the non-incentive experiment (section 2.3) and the sub-sampling of respondents within households (section 2.2.4). The base weight was given by

$$W_{bij} = \frac{1}{P_{ij}} R \, k \, h_i \, C_i \, S_j \tag{1}$$

where

P_{ij}	=	the initial probability of selection of household j in segment i;
R	=	the adjustment factor for the selection of the reserve sample;
k	=	the adjustment factor to reflect the sub-sampling of the non-incentive sample;
h_i	=	the adjustment factor for the addition of missed structures and dwelling units in segment i;
C_i	=	the adjustment factor to reflect the chunking of the segments during the listing operation; and

S_j = the factor to reflect the sub-sampling of persons in household j with multiple eligible members.

Twelve respondents in the national sample had extremely high base weights resulting from various features of the design. The base weights of these respondents were trimmed down to about three times the mean value of the base weights to avoid unnecessary increases in variances of estimates from the National Adult Literacy Survey.

3.2.3 Nonresponse Adjustments and Poststratification

Before compositing the national and state samples, the base weights for each sample were post-stratified separately to known population totals. This first-level poststratification provided sampling weights with lower variation and adjusted for nonresponse. Poststratification implicitly adjusts for unit nonresponse through adjustments to the weights of the responding units. Typically, the adjustments are made for subgroups of the sample that are likely to be quite different or for subgroups with high nonresponse rates. Poststratification is appropriate when population totals are known for the subgroups, or weighting classes, of the sample.

For purposes of poststratification, the entire sample was partitioned into classes, with the classification based on available survey data from respondents. Each class contained sample persons with the survey characteristics provided below. The adjustment was then implemented within each weighting class. The national and state records were split into 45 mutually exclusive and exhaustive groups, according to the state the record came from, whether the record came from the national or a state sample, and whether the record came from a PSU that was included in the national sample with certainty. The 45 groups were defined as follows:

Groups 1–11	State records from PSUs that were not selected with certainty for the national component, separated by state;
Groups 12–22	State records from PSUs that were selected with certainty for the national component, separated by state;
Groups 23–33	National records from one of the states participating in the state survey, from PSUs that were not selected with certainty for the national component, separated by state;
Groups 34–44	National records from one of the states participating in the state survey, from PSUs that were selected with certainty for the national component, separated by state; and
Group 45	National records from states not participating in the state survey.

State records were post-stratified separately from national records to provide a common base for applying the composite weighting factors. Population totals were calculated separately for each distinct group, based on 1990 census figures adjusted for undercount, thereby providing the control totals for poststratification. (More detail on poststratification totals is presented in section 3.2.5.)

A post-stratified base weight was calculated for each person in the sample as follows:

$$WPS_{hi} = Wb_{hi}\left[\frac{NT_h}{\sum_{i=1}^{n_h} Wb_{hi}}\right] \qquad (2)$$

where

WPS_{hi}	=	the post-stratified base weight for the i^{th} person record in the h^{th} group;
Wb_{hi}	=	the base weight for the i^{th} person record in the h^{th} group;
NT_h	=	the population total for the h^{th} group; and
n_h	=	the number of respondents in the h^{th} group.

3.2.4 Compositing Data from the National and State Components

3.2.4.1 Composite estimation procedure

Composite estimates were developed so that National Adult Literacy Survey data could be used to produce both state and national statistics. The original plan was to consider the national and state samples as two separate surveys, so that national statistics would be prepared from the national sample only and state data would be prepared from the state samples only. Upon reconsideration, it was clear that sampling error would be reduced by combining the state and national samples for each state that participated in the state survey. The combined sample had the advantages of producing a single database for state and national statistics and improving precision.

The method of combining data from the state and national samples is referred to as composite estimation. The composite estimation procedure and issues associated with the choice of composite weights for the national and state samples are discussed in the following sections.

The composite estimator for the national/state sample is given by

$$\hat{Y} = \beta_i \hat{Y}_{st} + (1 - \beta_i)\hat{Y}_{nt} \qquad (3)$$

where

$\hat{Y}$ = the composite estimate for variable Y in state i;

β_i = the composite factor for state i ($0 < \beta_i < 1$);

$\hat{Y}_{st}$ = the estimate of Y coming from the state sample; and

$\hat{Y}_{nt}$ = the estimate of Y coming from the national sample.

The variance of a composite estimator will be smaller than the variance of both the national and state estimates if appropriate composite factors are used. Optimal factors can be found when unbiased estimators exist for the two components and approximate estimates of their variances are available. It should be noted that a composite estimator will produce unbiased estimates for any value of β_i. The optimum value of β_i is the one that results in the lowest variance. However, there is generally only a slight loss in efficiency if a reasonable approximation of the optimum value of β_i is used. In most practical situations (including the national and state components of the National Adult Literacy Survey), approximations are necessary because there is insufficient information available to provide the optimal value of β_i when sample weights are produced.

As stated earlier, the national and state samples were selected independently, and each could thus produce unbiased estimates of sub-domain statistics for persons 16–64 years of age. Therefore, factors could be derived to produce composite estimators with variances that were smaller than those of either of the two estimates. For statistic Y, the optimal composite factor for state i is

$$\beta_i = \frac{V(\hat{Y}_{nt})}{V(\hat{Y}_{nt}) + V(\hat{Y}_{st})} \tag{4}$$

where

$V(\hat{Y}_{nt})$ = the variance of the estimate of Y coming from the national sample; and

$V(\hat{Y}_{st})$ = the variance of the estimate of Y coming from the state sample.

A different optimal value of β_i might be found for each statistic of interest. However, data analyses would be complicated if item-specific values of β_i were used, because items would not add up to totals, or totals derived by summing different items would not agree. Consequently, the goal for the National Adult Literacy Survey was to associate with each person in the sample a single compositing factor that, while not precisely optimal for any particular statistic, would be robust enough to enhance the precision of virtually all composited statistics. This objective was accomplished by focusing on aspects of the sample design that were likely to affect the variance, regardless of the choice of statistic. Under simple random sampling, the variance of the estimator is inversely proportional to the sample size, and the expression for β_i simplifies to the following:

$$\beta_i = \frac{n_{st}}{n_{st} + n_{nt}} \tag{5}$$

where

n_{st} = the number of respondents age 16-64 in the state sample; and

n_{nt} = the number of respondents age 16-64 in the national sample.

Because of the complexity of the National Adult Literacy Survey sample design, it was useful to think of deriving β_i in terms of the effective sample size, i.e., the actual sample size divided by the design effect. Three aspects of the survey design tended to inflate the design effect and thereby reduce the effective sample size: clustering, stratification, and the differential sampling rates used for Black and Hispanic adults.

In both the national and state components, clustering occurred at the PSU and segment levels and, to a trivial extent, at the household level, where two respondents were sampled in a small proportion of households. Geographic clustering kept the cost of survey administration down but reduced the effective sample size because of within-PSU and within-segment intraclass correlations. For example, in the Current Population Survey, which has a PSU and segment sample design similar to that of the National Adult Literacy Survey, the within-PSU and within-segment intraclass correlations have been estimated to average about 0.00075 and 0.042, respectively (Train et al., 1978). It seemed reasonable to use these values as approximations of intraclass correlations for the national and state components of the National Adult Literacy Survey.

Ordinarily, stratification enhances sample efficiency, but the national PSU sample was designed to optimize the precision of national estimates. As a result, stratum boundaries did not always conform with state boundaries; in fact, because PSUs sometimes contained counties from more than a single state, the measure of size used for PSU sample selection was not always optimal for producing state estimates. This aspect of the national design affected the variances of the state-level estimates coming from the non-certainty PSUs included in the national sample. (Note that stratum boundaries do not cause any problem for PSUs selected with certainty, because they are self-representing.)

In the national sample, minority households were oversampled in segments containing a high proportion of Black and Hispanic households. This practice introduced variability in the weights and increased the design effect. Minority households were not oversampled in the state survey. A separate source of variability in weights for both the national and state samples was the within-household sampling of persons, although this variability was dampened somewhat by increasing the sample size to two persons in households containing four or more eligible adults.

To best reflect the influence of these design aspects on the effective sample size, distinct compositing factors were derived for up to four subsets of data in each participating state. Those subsets were defined according to (1) whether or not the data came from a PSU chosen with certainty for the national sample and (2) whether or not the respondent was Black or Hispanic.

3.2.4.2 Deriving the PSU design effect

As mentioned in the previous section, the national PSU sample was not designed to maximize the efficiency of state-level estimates. To estimate the relative loss of efficiency for state data resulting from the inclusion of the national non-certainty PSUs, special tabulations were produced for each of the 11 participating states. The analysis was based on a variable that was likely to be correlated with literacy at the PSU level: the percentage of persons age 25 or older who had 0–8 years of schooling. Although the use of 1990 census data would have been preferable, only 1980 figures were available at the time.

First, all possible PSU samples under the national sample design were enumerated, and the between-PSU variances were computed for the estimated percentage using a Taylor series approximation. This process was repeated for the state design. These variances, which are presented in the third column of Table 3-1, were used to calculate provisional compositing factors that would have been appropriate had no within-PSU sampling been performed. These compositing factors reflect the limitations of the national stratification procedures for producing efficient state estimates. The table shows that the national design was quite adequate for producing state estimates in California but was greatly deficient in Louisiana.

Under the hypothesis that the national and state designs were equally efficient, another set of compositing factors, based strictly on the counts of PSUs (excluding the certainty PSUs in the national sample), was computed. These figures are presented in the fifth and sixth columns of Table 3-1. A factor similar to a design effect was computed by taking the quotient of the ratio of the state and national compositing factors derived using the two approaches:

$$F_{ij} = \left[\frac{\beta_{\text{Between- PSU variance}}}{(1-\beta_{\text{Between- PSU variance}})}\right] / \left[\frac{\beta_{\text{PSU count}}}{(1-\beta_{\text{PSU count}})}\right] \tag{6}$$

$$= \left[\frac{\text{National between-PSU variance}}{\text{number of national PSUs}}\right] / \left[\frac{\text{State between-PSU variance}}{\text{number of state PSUs}}\right]$$

This factor plays a role in calculating the effective sample size, as described in the next section.

3.2.4.3 Estimating composite factors

For data collected in PSUs selected with certainty for both the national and state samples, the effective sample size was estimated as:

$$n_{eff}^{ijk} = \frac{n_{ijk}}{1 + (\bar{n}_{ijk} - 1)\rho_1 + V^2_{w_{ijk}}} \qquad (7)$$

where

i	=	a participating state;
j	=	national or state sample;
k	=	minority (Black or Hispanic) or non-minority;
n_{ijk}	=	total number of respondents ages 16-64
$\bar{n}_{ijk}$	=	mean number of respondents per segment;
ρ_1	=	0.042, the intraclass correlation within segment, assumed to be equal to the Current Population Survey average and to be constant across states; and
$V^2_{w_{ijk}}$	=	the relvariance[2] of the weights.

[2]Relvariance, short for relative variance, is calculated by dividing the variance on an estimate by the squared value of the estimate.

Table 3-1. Between-PSU variance and provisional compositing factors for the National Adult Literacy Survey national and state PSU sample designs

State	Data source	Between variance*	Provisional compositing factors	PSU count[t]	Provisional compositing factors	F_{ij}**
California	National	0.000498	0.4644	4	0.5000	1.15
	State	0.000432	0.5356	4	0.5000	1.00
Illinois	National	0.001375	0.1735	3	0.3750	2.86
	State	0.000289	0.8265	5	0.6250	1.00
Indiana	National	0.000401	0.0865	4	0.2500	3.52
	State	0.000038	0.9135	12	0.7500	1.00
Iowa	National	0.001812	0.0324	2	0.1429	4.97
	State	0.000061	0.9676	12	0.8571	1.00
Louisiana	National	0.002499	0.0210	1	0.1000	5.19
	State	0.000053	0.9790	9	0.9000	1.00
New Jersey	National	0.000430	0.0000	4	0.2857	1.00
	State	0.000000	1.0000	10	0.7143	1.00
New York	National	0.000127	0.3964	2	0.3333	0.76
	State	0.000083	0.6037	4	0.6667	1.00
Ohio	National	0.000140	0.1703	5	0.2941	2.03
	State	0.000029	0.8297	12	0.7059	1.00
Pennsylvania	National	0.000214	0.2571	4	0.3333	1.44
	State	0.000074	0.7429	8	0.6667	1.00
Texas	National	0.001482	0.1715	4	0.3333	1.44
	State	0.000307	0.8285	8	0.6667	1.00
Washington	National	0.000390	0.0681	1	0.1111	1.71
	State	0.000029	0.9319	8	0.8889	1.00

* Of the estimated percentage of persons 25 or older (1980) with 0-8 years of schooling.
t Excluding National Adult Literacy Survey certainty PSUs.
** A design-effect-like factor descriptive of the relative inefficiency of the national PSU sample design for making state estimates

For data collected in other than the certainty PSUs included in the national sample, the effective sample size was estimated as

$$n_{eff}^{ijk} = \frac{n_{ijk}}{1 + (\bar{n}_{ijk} - 1)\rho_1 + (\bar{m}_{ijk} - 1)\rho_2 P_{ijk} F_{ij} + V_{w_{ijk}}^2} \tag{8}$$

where

i	=	a participating state;
j	=	national or state sample;
k	=	minority (Black or Hispanic) or nonminority;
n_{ijk}	=	total number of respondents ages 16-64;
$\bar{n}_{ijk}$	=	mean number of respondents per segment;
ρ_1	=	0.042, the intraclass correlation within segment, assumed to be equal to the Current Population Survey average and to be constant across states
$\bar{m}_{ijk}$	=	mean number of respondents per segment;
ρ_2	=	0.00075, the intraclass correlation within PSU, assumed to be equal to the CPS average and to be constant across states;
P_{ijk}	=	the proportion of respondents in non-certainty PSUs;
F_{ij}	=	a design-effect-like factor descriptive of the relative inefficiency of the national PSU sample design for making state estimates; and
$V^2_{w_{ijk}}$	=	the relvariance[2] of the weights.

Then an estimate of the optimal composite factor for state i is given by

$$\beta_{i(State)k} = \frac{n_{eff_{i(State)k}}}{n_{eff_{i(State)k}} + n_{eff_{i(National)k}}} \qquad (9)$$

$$\beta_{i(National)k} = 1 - \beta_{i(State)k} = \frac{n_{eff_{i(National)k}}}{n_{eff_{i(State)k}} + n_{eff_{i(National)k}}} \qquad (10)$$

Table 3-2 presents each of the quantities contained in the above formulas and the final compositing factors.

Table 3-2. Derivation of factors used to composite National Adult Literacy Survey national and state data

State	National Certainty PSU	Race/ethnicity	Data source	Sample size	Persons/ segment	Persons PSU	P_{ijk*}	F_{ij*}	Relvariance of weights	Effective design effect	Sample size	Compositing factor
California	No	Black or Hispanic	National	196	3.5	49.0	1.0	1.2	0.3305	1.48	132.7	0.7098
			State	62	2.1	20.7	1.0	1.0	0.0804	1.14	54.2	0.2902
		Other	National	200	3.7	50.0	1.0	1.2	0.1393	1.30	154.4	0.4401
			State	260	4.7	65.0	1.0	1.0	0.1191	1.32	196.4	0.5599
	Yes	Black or Hispanic	National	675	13.0	-	-	-	0.3666	1.87	361.0	0.6883
			State	226	7.5	-	-	-	0.1083	1.38	163.5	0.3117
		Other	National	414	8.3	-	-	-	0.1177	1.42	290.9	0.5232
			State	457	15.2	-	-	-	0.1261	1.72	265.1	0.4768
Illinois	No	Black or Hispanic	National	56	4.3	18.7	1.0	2.9	0.3629	1.54	36.4	0.5968
			State	29	1.8	7.25	1.0	1.0	0.1414	1.18	24.6	0.4032
		Other	National	202	5.3	67.3	1.0	2.9	0.0844	1.41	143.5	0.3210
			State	417	6.1	83.4	1.0	1.0	0.0965	1.37	303.5	0.6790
	Yes	Black or Hispanic	National	161	5.2	-	-	-	0.1764	1.35	119.0	0.4378
			State	198	5.0	-	-	-	0.1292	1.30	152.9	0.5622
		Other	National	121	4.8	-	-	-	0.1243	1.29	94.1	0.2502
			State	378	7.0	-	-	-	0.0882	1.34	282.1	0.7498
Indiana	No	Black or Hispanic	National	107	5.4	35.7	1.0	3.5	0.3943	1.67	64.1	0.3834
			State	126	3.1	11.5	0.3	1.0	0.1324	1.22	103.1	0.6166
		Other	National	215	5.8	71.7	1.0	3.5	0.0628	1.45	148.1	0.1746
			State	947	5.9	78.9	0.7	1.0	0.1072	1.35	700.1	0.8254
Iowa	No	Black or Hispanic	National	2	1.0	2.00	1.0	5.0	0.1837	1.19	1.7	0.0441
			State	45	1.7	5.63	0.8	1.0	0.2007	1.23	36.5	0.9559
		Other	National	146	7.3	73.0	1.0	5.0	0.0997	1.63	89.4	0.1073
			State	1027	6.2	85.6	0.8	1.0	0.1083	1.38	743.7	0.8927
Louisiana	No	Black or Hispanic	National	80	4.7	80.0	1.0	5.2	0.2808	1.74	45.9	0.1559
			State	315	3.4	35.0	0.5	1.0	0.1562	1.27	248.4	0.8441
		Other	National	55	4.2	55.0	1.0	5.2	0.1222	1.47	37.5	0.0649
			State	718	5.5	79.8	0.6	1.0	0.1043	1.33	539.9	0.9351
N. Jersey	No	Black or Hispanic	National	132	4.3	33.0	1.0	1.0	0.4060	1.57	84.2	0.3293
			State	209	3.4	26.1	0.0	1.0	0.1182	1.22	171.6	0.6708
		Other	National	163	3.5	40.8	1.0	1.0	0.0917	1.23	132.7	0.2375
			State	535	4.6	53.5	0.0	1.0	0.1057	1.26	426.0	0.7625
	Yes	Black or Hispanic	National	15	3.0	-	-	-	0.2381	1.32	11.3	0.3438
			State	28	4.7	-	-	-	0.1391	1.29	21.7	0.6562
		Other	National	38	5.4	-	-	-	0.1346	1.32	28.8	0.2554
			State	103	4.9	-	-	-	0.0636	1.23	83.9	0.7446

Table 3-2. Derivation of factors used to composite National Adult Literacy Survey national and state data – continued

State	National Certainty PSU	Race/ethnicity	Data source	Sample size	Persons/ segment	Persons PSU	P_{ijk*}	F_{ij*}	Relvariance of weights	Effective design effect	Sample size	Compositing factor
New York	No	Black or Hispanic	National	69	5.3	34.5	1.0	0.8	0.2721	1.47	46.9	0.7075
			State	24	1.6	6.00	1.0	1.0	0.2096	1.24	19.4	0.2925
		Other	National	154	5.9	77.0	1.0	0.8	0.1035	1.35	113.8	0.2994
			State	370	6.1	92.5	1.0	1.0	0.1083	1.39	266.3	0.7006
	Yes	Black or Hispanic	National	275	7.6	-	-	-	0.3344	1.61	170.5	0.5812
			State	170	7.1	-	-	-	0.1283	1.38	122.9	0.4188
		Other	National	186	5.0	-	-	-	0.2343	1.40	132.5	0.3766
			State	317	9.1	-	-	-	0.1063	1.44	219.4	0.6235
Ohio	No	Black or Hispanic	National	158	4.8	31.6	1.0	2.0	0.3722	1.58	100.0	0.4724
			State	138	2.8	11.5	0.2	1.0	0.1579	1.23	111.9	0.5277
		Other	National	309	4.8	61.8	1.0	2.0	0.0962	1.35	229.0	0.2583
			State	871	5.7	72.6	0.4	1.0	0.1040	1.32	657.4	0.7417
Pennsyl-vania	No	Black or Hispanic	National	25	2.3	6.25	1.0	1.4	0.6318	1.69	14.8	0.2555
			State	52	2.5	7.43	0.5	1.0	0.1427	1.21	43.1	0.7445
		Other	National	309	5.9	77.3	1.0	1.4	0.0818	1.37	225.2	0.3048
			State	704	6.2	88.0	0.7	1.0	0.1055	1.37	513.6	0.6952
	Yes	Black or Hispanic	National	60	3.5	-	-	-	0.1565	1.26	47.5	0.4881
			State	64	3.4	-	-	-	0.1848	1.28	49.8	0.5119
		Other	National	79	4.2	-	-	-	0.1581	1.29	61.2	0.2693
			State	210	5.3	-	-	-	0.8570	1.26	166.1	0.7308
Texas	No	Black or Hispanic	National	235	3.5	58.8	1.0	2.4	0.3547	1.56	150.4	0.4069
			State	272	3.9	34.0	0.9	1.0	0.0942	1.24	219.3	0.5932
		Other	National	250	3.6	62.5	1.0	2.4	0.1670	1.39	180.0	0.3210
			State	497	5.1	62.1	0.8	1.0	0.0971	1.30	380.9	0.6790
	Yes	Black or Hispanic	National	194	6.3	-	-	-	0.3709	1.59	121.9	0.5185
			State	145	5.0	-	-	-	0.1132	1.28	113.2	0.4815
		Other	National	155	5.7	-	-	-	0.1429	1.34	115.5	0.3532
			State	320	10.7	-	-	-	0.1068	1.51	211.5	0.6468
Washington	No	Black or Hispanic	National	13	1.6	13.0	1.0	1.7	0.4044	1.45	9.0	0.1578
			State	55	1.3	6.88	0.3	1.0	0.1305	1.15	48.8	0.8422
		Other	National	99	6.2	99.0	1.0	1.7	0.0945	1.44	68.8	0.0821
			State	1064	6.5	133.0	0.4	1.0	0.1096	1.38	769.7	0.9179

* As defined in Section 3.2.4.3.

3.2.5 Computing Final Weight—Poststratification Through Raking Ratio Adjustments

Poststratification is commonly used in sample surveys to accomplish three purposes: (1) It generally reduces the sampling errors; (2) it is frequently an effective way of making nonresponse adjustments; and (3) it creates consistency with statistics from other studies. The National Adult Literacy Survey used a particular form of poststratification referred to as raking ratio adjustments. The final sampling weights were computed by raking the composited weights to known population totals. In poststratification, classes are formed from cross-tabulations of certain variables. In some instances, such cross-tabulations may lead to sparse cells, or population distributions may be known for the marginal but not the joint distributions for variables used to define the weighting classes. Weighting class adjustments based on small cell sizes can result in a large amount of variation in the adjusted weights. Raking ratio adjustments are useful for maintaining the weighted marginal distributions of variables used to define weighting classes. For this type of adjustment, population distributions are required for the marginal distributions of the weighting class variables and not for their joint distribution.

An objective of raking ratio adjustments is to adjust the weights of cells in such a way that the marginal distributions for the weighted sample correspond to known population distributions. To illustrate the algorithm, consider a simple case of two variables that are cross-tabulated. Using an example from Kalton (1981), the marginal and joint distributions for the population and sample are as follows.

	Population 1 2 … K	Total		Sample 1 2 … K	Total
1	W_{11} W_{12} … W_{1K}	$W_{1.}$	1	q_{11} q_{12} … q_{1K}	$q_{1.}$
2	W_{21} W_{22} … W_{2K}	$W_{2.}$	2	q_{21} q_{22} … q_{2K}	$q_{2.}$
⋮	⋮ ⋮ ⋮	⋮	⋮	⋮ ⋮ ⋮	⋮
H	W_{H1} W_{H2} … W_{HK}	$W_{H.}$	H	q_{H1} q_{H2} … q_{HK}	$q_{H.}$
Total	$W_{.1}$ $W_{.2}$ … $W_{.K}$	$W_{..}$	Total	$q_{.1}$ $q_{.2}$ … $q_{.K}$	$q_{..}$

The iterative procedure makes successive modifications to the weights until the process stabilizes. The algorithm used for raking in the National Adult Literacy Survey, and described by Kalton, first weights each cell in row h (h=1, …, H) by the factor $W_{h.}/q_{h.}$. The result is that the sum of the weighted cells for a given row h, $\frac{W_{h.}}{q_{h.}} \sum_{h} q_{hi}$, will be equal to $W_{h.}$. Because of the adjustments to the weights, the column totals for the sample now become $\sum_{h} q_{hk} \frac{W_{h.}}{q_{h.}} = \sum_{h} q'_{hk} = q'_{.k}$. At the second step in the iterative procedure, the sampled units in each cell in column k (k=1, …, K) are weighted by the factor $W_{.k} \_ q'_{.k}$. Then, the sum of the weights in a given column k is equal to $W_{.k}$. At this point, the $q'_{.k}$ values have been changed to $\sum_{h} q'_{hk} \frac{W_{h.}}{q'_{h.}} = \sum_{h} q''_{hk} = q''_{.k}$. The process now repeats with step one. The procedure is completed when the process converges or, alternatively, is terminated after a pre-specified number of iterations. The result is a set of adjusted weights that are then used for estimation. It has been shown that the raking ratio estimation procedure produces best asymptotically normal estimates under simple random sampling. At the same time, the procedure minimizes the adjustments to the sample weights based on one measure of closeness (Ireland & Kullback, 1968).

Construction of weighting classes is an important consideration in poststratification, particularly when it is used as an adjustment for unit nonresponse. A purpose of using weighting classes is to bring together respondents and nonrespondents with similar characteristics not only for the variables defining the classes but also for variables that are unknown for nonrespondents only. The variables used to construct raking classes for the National Adult Literacy Survey were age, race/ethnicity, sex, education, and geographic indicators, i.e., metropolitan statistical area (MSA) vs. non-MSA for the 11 states and census region for the remainder of the United States.

The 1990 census totals used for raking were adjusted separately by age, race/ethnicity, sex, and region of the country to account for undercoverage. The undercoverage rates used in this process were supplied by the U.S. Bureau of the Census.

3.3 REPLICATED WEIGHTS FOR VARIANCE ESTIMATION IN THE HOUSEHOLD POPULATION

Variance estimation must take into account the sample design. In particular, the estimate of sampling variance for any statistic should account for the effects of clustering, the use of nonresponse and poststratification adjustments, and the component of sampling variability arising from the variation in the weights used to compute the statistic. Treating the data as a simple random sample will produce underestimates of the true sampling variability.

The jackknife method can be used to estimate the variance for most statistics. Jackknifing estimates the sampling variability of any statistic Y, as the sum of components of variability that may be attributed to individual pairs of first-stage sampling units. The variance attributed to a particular pair is measured by estimating how much the value of the statistic would change if only one unit in the pair had been sampled. When using replication techniques such as jackknifing to calculate standard errors, it is necessary to establish a number of subsamples (or replicates) from the full sample, calculate the estimate from each subsample, and sum the squared difference of each replicated estimate from the full-sample estimate. The 60 replicates formed for the National Adult Literacy Survey provided the degrees of freedom necessary for the production of stable estimates of variance.

Variance estimation requires three steps: (1) forming the replicates, (2) constructing the replicate weights, and (3) computing estimates of variance for survey statistics. The formation of replicates is discussed in detail in sections 3.3.1 through 3.3.3. After the replicates had been formed, a replicate factor was constructed for each variance stratum. Let $f_{ijk}(r)$ denote the r^{th} replicate factor for the k^{th} respondent in the j^{th} variance unit in the i^{th} variance stratum. Then, in general:

$$f_{ijk}(r) = \begin{cases} 2 \text{ if } i = r \text{ and } j = 1 \\ 0 \text{ if } i = r \text{ and } j = 2 \\ 1 \text{ if } i \neq r \end{cases} \tag{11}$$

and the replicated base weight, $Wb_{ijk}(r)$, was obtained as $Wb_{ijk}(r) = Wb_{ijk}\, f_{ijk}(r)$ for r = 1, 2, …, 60. (A variation on this scheme, used for only non-certainty PSUs in the state component, is described in section 3.3.2.)

After obtaining a person base weight for each replicate, all remaining full-sample weighting steps leading to the final person weight were performed on each replicate. By repeating the various weight adjustment procedures on each set of replicate base weights, the impact of these procedures on the sampling variance of the estimator Y is appropriately reflected in the variance estimator, v(Y).

After the replicate weights had been constructed, the estimate of variance could easily be computed for any statistic. The statistic was computed 61 times, once using the full-sample weight and an additional 60 times using each of the 60 replicate weights. The variance estimate is the sum of the 60 squared differences between the estimate derived using the full-sample weight and the estimate derived using each of the 60 replicate weights. That is, the estimate of the variance of a statistic Y is,

$$v(Y) = \sum_{r=1}^{60} (Y_r - Y)^2 \tag{12}$$

where Y_r = the weighted estimate obtained using the r^{th} replicate weight; and Y = the weighted estimate obtained using the full-sample weight.

The National Adult Literacy Survey pooled data from a nationally representative sample of 101 PSUs and from 11 independently selected state PSU samples. The threefold objective of the replication scheme was (1) to reflect the actual sample design of each sample; (2) to ensure the production of stable estimates of standard errors by having sufficient degrees of freedom for national estimates, individual state estimates, and regional estimates; and (3) to limit the total number of replicates so that variance estimation would not be prohibitively expensive. The general approach in setting up the replication was to devise an appropriate scheme for each component of the sample, the national sample and the 11 states, and then to collapse replicates to a reasonable number.

3.3.1 Household Sample Replication for the National Component

The national sample contained 101 PSUs, 25 of which were selected with certainty. The remaining 76 PSUs were selected 2 per stratum using the Durbin method (1967), with probabilities proportional to size and with known joint probabilities. Ordinarily, replicates are formed by pairing first-stage sampled units, that is, segments are paired in PSUs selected with certainty and whole PSUs are paired in non-certainty strata. However, under the Durbin scheme, an unbiased estimate of variance can be obtained by treating PSUs in some non-certainty strata as if they had been chosen with certainty, that is, by pairing segments instead of whole PSUs. For the 101-PSU sample, the natural pairing led to 74 replicates. These replicates were examined carefully to see which contained data from any of the 11 participating states. In certainty PSUs where segments from a participating state had been paired to form a replicate, the segments were grouped into subsets and were paired within each subset to increase the number of replicates and hence the degrees of freedom of the state variance estimator. This procedure expanded the number of national sample replicates to 111.

3.3.2 Household Sample Replication for the State Component

An independent sample of 8 to 12 PSUs was selected in each of the 11 participating states. The largest PSUs were taken with certainty. Within each state, the remaining PSUs were grouped into strata, and from each stratum a single PSU was sampled with probability proportional to size. In PSUs selected with certainty, segments were paired to form replicates. However, the segments were grouped into subsets and paired within each subset to increase the degrees of freedom. This procedure created from 2 to 8 replicates for each PSU chosen with certainty, with a total of 113 replicates across the 11 states.

Ordinarily, non-certainty PSUs would be paired to form replicates so that, for instance, a state with n such PSUs would yield $n/2$ replicate pairs. With the goal of increasing the degrees of freedom, an alternative procedure was adopted. The same n PSUs were used to create n-1 replicates, as follows: The active part of each replicate contained data from exactly n-1 of the n PSUs, and the base weight was multiplied by $n/(n-1)$ rather than the usual factor of 2. One randomly selected PSU was active in all n-1 replicates, and a successively different one of the remaining n-1 PSUs was inactive in each of the n-1 replicates. It was possible to create n replicates from the n PSUs, but only at the expense of a bothersome complication in the variance estimation formula. The applied method kept estimation consistent with the rest of the sample and created 54 replicates across the 11 states.

3.3.3 Final Household Sample Replication for the National and State Components

A total of 278 replicates had been formed at this point: 111 from the national sample, 113 from PSUs chosen with certainty for the state samples, and 54 from non-certainty PSUs chosen for the state samples. These replicates reflected the actual design of each sample and provided sufficient degrees of freedom to produce stable estimates of variance for the nation, each state, and the four census regions. However, using 278 replicates to estimate variances would be computer intensive and expensive, while providing only a slight gain in the precision of the overall estimates. Therefore, the replicates were collapsed to 60, a much more realistic number. To preserve the total number of replicates for each state, replicates from the same state were never collapsed. As often as possible, the same constraint was used by region as well.

Table 3-3 presents the results of the replication scheme, showing which replicates are active for the major sub-domains of analysis.

3.4 CALCULATING SAMPLE WEIGHTS FOR THE PRISON POPULATION

The final inmate weight was constructed in four major steps. The first step was to construct the inmate base weight, which was the reciprocal of the overall probability of selection for each inmate. The second step was to adjust the inmate base weight for the one facility that did not cooperate, so that weighted estimates for inmates from cooperating facilities would also represent inmates from the non-cooperating facility. The third step was to adjust the inmate weight to compensate for not obtaining a completed background questionnaire for every inmate in the sample. The fourth step was to post-stratify the weight so that the weighted counts from the sample agreed with independent estimates for certain subgroups of the population.

3.4.1 Computing Inmate Base Weights

The initial correctional facility sample consisted of 96 facilities, of which eight facilities were randomly selected and set aside as the reserve sample. The reserve sample was never used because the actual response rates were higher than those originally estimated for the sample of 96 facilities. The reduced sample of facilities was drawn by taking a systematic sample, with equal probabilities of selection, from a listing of all sample facilities in their initial selection order.

The facility weight for the remaining 88 facilities in the sample was computed as a product of the reciprocal of the probability of the i^{th} facility (PSU) being selected to the initial sample and the reciprocal of the probability of its not being selected to the reserved sample; that is:

$$W_{bi} = \frac{1}{P_i} \frac{1}{88 \;\; / 96} \tag{13}$$

where

$W_i =$ the weight for the i^{th} facility; and

$P_i =$ the probability of selection of the i^{th} facility.

The inmate base weight is the reciprocal of the overall probability of selecting the j^{th} inmate in the i^{th} facility.

$$WI_{bij} = W_{bi} \frac{N_i}{n_i} \tag{14}$$

where

N_i = the inmate population size for the i^{th} facility; and

n_i = the inmate sample size for the i^{th} facility.

Table 3-3. Active replicates for sub-domains of the National Adult Literacy Survey analysis file

Replicate	Household sample																Prison sample
	U.S.	Northeast	Midwest	South	West	California	Illinois	Indiana	Iowa	Louisiana	New Jersey	New York	Ohio	Pennsylvania	Texas	Washington	
1	x	x	x	x	x	x	x		x	x	x						x
2	x	x	x	x	x	x			x	x	x						x
3	x	x	x	x	x	x	x		x	x	x		x				x
4	x	x	x	x	x	x	x	x	x	x	x						x
5	x	x	x	x	x	x	x		x	x	x						x
6	x	x	x	x	x	x	x		x	x	x						x
7	x	x	x	x	x	x	x		x	x	x						x
8	x	x	x	x	x	x	x		x	x	x						x
9	x	x	x	x	x	x	x		x	x	x						x
10	x	x	x	x	x	x	x		x	x	x						x
11	x	x	x	x	x	x	x		x	x	x						x
12	x	x	x	x	x	x	x		x	x	x						x
13	x	x	x	x	x	x	x		x	x	x						x
14	x	x	x	x	x	x	x		x	x	x						x
15	x	x	x	x	x	x	x		x	x	x						x
16	x	x	x	x	x	x	x		x	x	x						x
17	x	x	x	x	x	x	x		x	x	x						x
18	x	x	x	x	x	x	x		x	x	x						x
19	x	x	x	x	x	x	x		x	x	x						x
20	x	x	x	x	x	x	x		x	x	x						x
21	x	x	x	x	x	x	x				x						x
22	x	x	x	x	x	x	x				x						x
23	x	x	x	x	x	x	x				x						x
24	x	x	x	x	x	x	x				x						x
25	x	x	x	x	x	x		x			x						x
26	x	x	x	x	x	x		x			x						x
27	x	x	x	x	x	x		x			x						x
28	x	x	x	x	x	x		x			x						x
29	x	x	x	x	x	x		x			x						x
30	x	x	x	x	x	x		x				x					x
31	x	x	x	x	x	x		x				x					x

Table 3-3. Active replicates for sub-domains of the National Adult Literacy Survey analysis file – continued

Replicate	Household sample																Prison sample
	U.S.	Northeast	Midwest	South	West	California	Illinois	Indiana	Iowa	Louisiana	New Jersey	New York	Ohio	Pennsylvania	Texas	Washington	
32	x	x	x	x	x	x		x				x	x				x
33	x	x	x	x			x				x	x					x
34	x	x	x	x			x				x	x					x
35	x	x	x	x			x				x	x					x
36	x	x	x	x	x			x				x	x		x		x
37	x	x	x	x	x			x				x	x		x		x
38	x	x	x	x	x			x				x	x	x	x		x
39	x	x	x	x	x			x				x	x	x	x		x
40	x	x	x	x	x			x				x	x	x	x		x
41	x	x	x	x	x			x				x	x	X	x	x	x
42	x	x	x	x	x							x	x	X	x	x	x
43	x	x	x	x	x							x	x	x	x	x	x
44	x	x	x	x	x							x	x	x	x	x	x
45	x	x	x	x	x							x	x	x	x	x	x
46	x	x	x	x	x							x	x	x	x	x	
47	x	x	x	x	x							x	x	x	x	x	
48	x	x	x	x	x							x	x	x	x	x	
49	x	x	x	x	x							x	x	x	x	x	
50	x	x	x	x	x							x	x	x	x	x	
51	x	x	x	x	x							x	x	x	x	x	
52	x	x	x	x	x								x	x	x	x	
53	x	x	x	x	x								x	x	x	x	
54	x	x	x	x	x								x	x	x	x	
55	x	x	x	x	x								x	x	x	x	
56	x	x	x	x	x									x	x	x	
57	x	x	x	x	x									x	x	x	
58	x	x	x		x											x	
59	x	x	x		x											x	
60	x	x	x														
# active	60	60	60	57	59	32	23	18	20	20	29	22	25	20	22	19	45

3.4.2 Nonresponse Adjustments

3.4.2.1 Facility nonresponse adjustment

Only one correctional facility did not cooperate. As described in section 2.6.1.1, the sample facilities were stratified on the basis of certain characteristics. Using this stratification scheme, the non-cooperating facility was classified as a state maximum security facility, in the southern region of the United States, with a male-only inmate population. To adjust for the non-cooperating facility, two nonresponse adjustment classes were constructed: (1) all facilities in the same sampling stratum (implicit stratum) as the non-cooperating facility and (2) all remaining facilities. The facility nonresponse adjustment factor was computed for each nonresponse class as the ratio of the weighted (facility weight times the facility inmate population size) sum of all eligible sample facilities to the respondent facilities. That is, the nonresponse adjustment factor for the α^{th} class, $A_{F\alpha}$, was computed as

$$A_{F\alpha} = \frac{\sum_{i \varepsilon S(\alpha)} W_{b\alpha i} N_{\alpha i}}{\sum_{i \varepsilon SR(\alpha)} W_{b\alpha i} N_{\alpha i}} \tag{15}$$

where

$W_{b\alpha i}$	=	the facility weight for the i^{th} facility in the α^{th} facility nonresponse adjustment class;
$N_{\alpha i}$	=	the inmate population count for the i^{th} facility in the α^{th} facility nonresponse adjustment class;
$S(\alpha)$	=	the collection of all eligible (cooperating and non-cooperating) sample facilities in the α^{th} facility nonresponse adjustment class; and
$SR(\alpha)$	=	the collection of all cooperating facilities in the α^{th} facility nonresponse adjustment class.

Table 3-4 presents the facility nonresponse adjustment factors for both nonresponse adjustment classes.

Table 3-4. National Adult Literacy Survey correctional facility sample counts and facility nonresponse adjustment factor, by facility nonresponse adjustment classes

Nonresponse adjustment class	Sample count		Nonresponse adjustment factor
	Eligible	Respondent	
1	8	7	1.122
2	80	80	1.000

3.4.2.2 Inmate nonresponse adjustment

The inmate sample consisted of 1,340 inmates, of whom 1,147 completed background questionnaires. The main reason for adjusting the sampling weights was to remove potential bias on statistics of interest as a result of the inability to collect completed background questionnaires for all sample inmates. If the probability of nonresponse were independent of the statistics of interest, then no bias would arise. Therefore, the objective was to obtain adjustment classes such that the probability of nonresponse within each class was as independent of statistics of interest as possible. There are several alternative methods of forming the classes to achieve this result. For the prison sample, the classes were formed so that the variation in the response propensity within the classes was minimized.

A set of potential predictive variables was selected for the response propensity. These variables had to be available for respondents and nonrespondents alike. They were

- State vs. Federal facility;
- Region: Northeast, Midwest, South, West;
- Sex of inmates: male only, both sexes, female only; and
- Facility type: maximum security, medium security, minimum security, medical, all other.

To form the nonresponse adjustment classes, a technique similar to the automatic interaction detection type of algorithm was used. Pearson chi-square statistics were computed between the response and each one of the predictive variables. The predictor with the smallest p-value was selected as the "best" predictor. Then, the same process was applied within the subgroups of the population, defined by the levels of the "best" predictor chosen in the preceding step. This process was continued until no significant predictor was found or until a specified minimum class size had been reached. The procedure is stepwise and creates a hierarchical, tree-like structure. The inmate nonresponse classes are shown in Table 3-5.

Table 3-5. National Adult Literacy Survey inmate sample counts and nonresponse adjustment factors, by inmate nonresponse adjustment classes

Region	Facility type	State/Federal facility	Sample counts		Nonresponse adjustment factor
			All	Respondent	
Northeast and West	Maximum security and medical	All	171	121	1.386
Northeast and West	All other	State	330	275	1.196
Northeast and West	All other	Federal	54	51	1.063
South and Midwest	Maximum security and medical	All	212	174	1.214
South and Midwest	Medium security	All	337*	302*	1.117
South and Midwest	Minimum security and other	All	235	224	1.051

*This class actually contained 338 and 303 responding inmates, with the additional unit representing one inmate who was selected into the sample twice from two different facilities. The number of records is adjusted here to be consistent with the number of records (1,147) receiving weights.

The inmate nonresponse adjustment factor for the h$^{\underline{th}}$ nonresponse adjustment class, $INRAF_h$, was computed as

$$AIh = \frac{\sum_{i \, \varepsilon \, A(h)} WI_{bhi} A_{Fhi}}{\sum_{i \, \varepsilon \, AR(h)} WI_{bhi} A_{Fhi}} \quad (16)$$

where

WI_{bhi}	=	the base weight for the ith inmate in the hth inmate nonresponse adjustment class;
A_{Fhi}	=	the facility nonresponse adjustment factor for the ith inmate in the hth nonresponse adjustment class;
A(h)	=	the collection of all sample inmates in the hth facility nonresponse adjustment class; and
AR(h)	=	the collection of all sample inmates with completed background questionnaires in the hth facility nonresponse adjustment class.

3.4.3 Poststratification Procedures

To reduce the mean square error of estimates, the weights were further adjusted so that the weighted totals obtained from the sample as estimates for certain subgroups of the population would be consistent with presumably more precise estimates available from external sources. Control totals were obtained from the U.S. Department of Justice's Bureau of Justice Statistics and were partly based on data from the 1991 Survey of Inmates in State Correctional Facilities. Both sets of estimates were obtained from larger samples than the one utilized in this survey and thus were expected to have greater precision.

Poststratification was intended to reduce nonresponse-related residual bias on the estimates and simultaneously to increase the precision of the post-stratified estimates. This beneficial effect on the variance was not restricted to the post-stratified variables. The precision of any substantive variable correlated with the post-stratified variables was also expected to improve.

For the male inmates, the poststratification estimation utilized raking ratio estimation. The inmate nonresponse adjusted weights were alternately adjusted by an iterative process to provide consistency with the independent estimates of population by age and then by education within each race/ethnicity category. Table 3-6 shows the sample estimates for male inmates (before raking) and the independent control totals by age and by education within race/ethnicity categories.

Table 3-6. Comparison of National Adult Literacy Survey sample estimates (before raking) and independent control totals, by age and by education within ethnicity, for male inmates

Race/ethnicity	Age or education	Sample		Control total
		Size	Estimate	
White and other	Age			
	less than 30	173	117,604	107,332
	30 or more	255	175,019	167,488
	Education			
	0-8 years	49	34,375	31,496
	9-12 years	272	186,001	189,149
	Some college	107	72,246	54,175
Black	Age			
	less than 30	240	165,229	155,912
	30 or more	210	145,130	164,931
	Education			
	0-8 years	40	27,475	35,968
	9-12 years	333	230,590	239,645
	Some college	77	52,118	45,230
Hispanic	Age			
	less than 30	107	76,144	6,400
	30 or more	91	61,543	65,569
	Education			
	0-8 years	59	41,256	34,035
	9-12 years	109	76,144	77,758
	Some college	30	20,289	15,176

Raking ratio estimation was used rather than a straightforward poststratification procedure because the cell sizes were too small to obtain stable estimates when age and education were cross-classified within race/ethnicity. Refer to section 3.2.5 for a detailed description of raking ratio estimation.

Table 3-7 shows the raking ratio estimate and the adjustment factor for each adjustment class for the male inmates. The small adjustment factors for inmates with some college education could be related to the tendency of better educated inmates to be more cooperative. A similar pattern can be observed for inmates who were less than 30 years old.

Table 3-7. Raking ratio estimates and weight adjustment factors for male inmates in the National Adult Literacy Survey sample

Adjustment cell	Race/ethnicity	Education	Age	Adjustment factor
1	White and other	0-8 years	less than 30	0.859
2			30 or more	0.943
3		9-12 years	less than 30	0.067
4			30 or more	1.061
5		Some college	less than 30	0.700
6			30 or more	0.768
7	Black	0-8 years	less than 30	1.130
8			30 or more	1.397
9		9-12 years	less than 30	0.952
10			30 or more	1.177
11		Some college	less than 30	0.741
12			30 or more	0.910
13	Hispanic	0-8 years	less than 30	0.684
14			30 or more	0.950
15		9-12 years	less than 30	0.892
16			30 or more	1.239
17		Some college	less than 30	0.614
18			30 or more	0.853

One-dimensional poststratification was used for female inmates mainly because of the small sample size for this group. The poststratification adjustment factor for the g^{th} poststratification adjustment class, PA_{Ig}, was

$$PA_{Ig} = \frac{C_g}{\sum_{i \varepsilon E(g)} WI_{bgi} A_{Fgi} A_{Igi}} \qquad (17)$$

where

C_g = the female inmate control total for the g^{th} poststratification class;

$E(g)$ = the collection of female respondent inmates in the g^{th} poststratification class;

WI_{bgi} = the inmate base weight for the i^{th} inmate in the g^{th} poststratification class;

A_{Fgi} = the facility nonresponse adjustment factor for the i^{th} inmate in the g^{th} poststratification class; and

A_{Igi} = the inmate nonresponse adjustment factor for the i^{th} inmate in the g^{th} poststratification class.

The poststratification factors for the female inmates are shown in Table 3-8.

Table 3-8. Control totals and poststratification adjustment factors for female inmates in the National Adult Literacy Survey sample, by poststratification classes

Poststratification adjustment cell	Race/ethnicity	Sample size	Control total	Poststratification factor
19	Black	30	19,465	0.906
20	All other	41	23,554	0.875

3.4.4 Final Inmate Weights

Final inmate weights were obtained as a product of the inmate base weight, the facility nonresponse adjustment factor, the inmate nonresponse adjustment factor, and the raking/poststratification adjustment factor:

$$Fw_{gh\alpha i} = WI_{bgh\alpha i} \; A_{F\alpha i} \; A_{Ihi} \; PA_{Igi} \qquad (18)$$

where

$WI_{bgh\alpha i}$ = the base weight for the i^{th} inmate in the α^{th} facility nonresponse adjustment class, the h^{th} inmate nonresponse adjustment class, and the g^{th} poststratification class;

$A_{F\alpha i}$ = the facility nonresponse adjustment factor for the i^{th} inmate in the α^{th} facility nonresponse adjustment class;

A_{Ihi} = the inmate nonresponse adjustment factor for the i^{th} inmate in the h^{th} inmate nonresponse adjustment cell; and

PA_{Igi} = the poststratification/raking adjustment factor for the i^{th} inmate in the g^{th} poststratification class.

Table 3-9 presents statistics for the sampling weights at each stage of weight adjustment. The table shows that the variation in the base weight was rather small and that nonresponse adjustments had only a trivial effect on the weight variation. The poststratification/raking increased the weight variation moderately. Despite the increase in weight variation, poststratification/raking usually decreases the variance of estimates for any characteristics that are correlated with the raked variables (Brackstone & Rao, 1979; Oh & Scheuren, 1978). The post-stratified/raked variables in this survey are known to be strongly correlated with many substantive characteristics. The poststratification procedure was effective in simultaneously reducing the residual nonresponse bias and the sampling variance.

Table 3-9. Statistics for the distribution of the weight-by-weight adjustment stage for the National Adult Literacy Survey incarcerated sample

Statistic	Base weight	Facility nonresponse adjusted weight	Inmate nonresponse adjusted weight	Post-stratified raked weight
Sample size	1,340	1,340	1,147	1,147
Mean	582.52	588.16	687.13	667.52
cv (%)	16.51	16.56	18.43	24.94
Minimum	110.22	110.22	115.89	110.29
5th Percentile	491.47	491.47	530.58	458.49
Median	593.20	596.27	684.30	644.83
95th Percentile	680.51	700.67	877.48	937.89
Maximum	1,012.37	1,012.37	1,682.92	1,785.87

3.5 REPLICATED WEIGHTS FOR VARIANCE ESTIMATION IN THE PRISON POPULATION

The use of a complex sample design, adjustments for nonresponse, and poststratification procedures resulted in dependence among the observations. The application of the usual formulae of variance estimation, which were based on simple random sampling assumptions, would result in the underestimation of sampling variance in this survey. To estimate sampling variability, therefore, 45 jackknife replicates were formed to provide adequate degrees of freedom for the production of reliable estimates. The variance estimation was carried out in three steps: (1) the replicates were formed, (2) the replicate weights were computed, and (3) the estimates of the variances of the survey statistics were computed.

The replicates were designed in accordance with the sample design. The 86 non-certainty facilities were placed in their sample selection order. Then, the facilities were paired consecutively, and each pair was assigned to a variance stratum. This process resulted in 43 variance strata. Within each variance stratum, one facility was assigned randomly to variance unit 1 and the other to variance unit 2. The two largest facilities in the sample were assigned to separate variance strata. These facilities were certainty selections and therefore their only contribution to the total variance was from within-facility sampling. Therefore, the inmate records within each facility were placed in their sample selection order and numbered sequentially. The odd-numbered inmates were assigned to one variance unit and the even-numbered inmates to the other. Thus, a total of 45 variance strata and 90 variance units were obtained. After the replicates had been formed, the replicate weights were constructed. A replicate factor was constructed for each variance stratum. If $f_{ijk}(r)$ denotes the r^{th} replicate factor for the k^{th} inmate in the j^{th} variance unit and the i^{th} variance stratum, then

$$f_{ijk}(r) = \begin{cases} 2 \text{ if } i = r \text{ and } j = 1 \\ 0 \text{ if } i = r \text{ and } j = 2 \\ 1 \text{ if } i \neq r \end{cases} \qquad (19)$$

The r^{th} replicate inmate base weight for the k^{th} inmate in the i^{th} variance stratum and the j^{th} variance unit, $WI_{bijk}(r)$, was then obtained as

$$WI_{bijk}(r) \quad = \quad WI_{bijk} \ f_{ijk}(r) \tag{20}$$

for r = 1, 2,........,45.

After obtaining an inmate base weight for each replicate, all remaining full-sample weighting steps leading to the final inmate weight were performed on each replicate. For each replicate, a facility nonresponse adjustment factor, an inmate nonresponse adjustment factor, and a poststratification adjustment factor were computed, and these factors were then applied to the replicate inmate base weight to obtain 45 replicate final inmate weights. Replicate weights 46 through 60 were “inactive” for the prison sample and were set equal to the full-sample weight in the data file. The variance estimation procedures were similar to those used for the household sample, as described in section 3.3.