Appendix B

Computation and Usage of Analysis Weights

Part I: Description

1. **Adjustments for Sample Size.** The *sampling weight* is defined as the reciprocal (or inverse) of the sampling fraction or selection probability for a particular stratum. For example, if the sampling fraction is 1/100 (1 out of every 100 mothers are sampled), the reciprocal (sample weight) would be 100. Each mother in the sample represents 100 mothers in the population. Because the sampling fractions vary by stratum, mothers in different strata start off with different sampling weights. Inconsistencies in sampling procedure and post-stratification may result in different sampling weights among mothers in the same stratum. The sum of the sampling weights for all mothers sampled in a particular stratum should equal the population of that stratum. The sum of the sampling weights of mothers sampled from all of the strata should equal the total number of mothers in a state's sampling frame.

2. **Adjustments for Nonresponse.** Failure of a mother in the sample to complete a questionnaire is called *unit nonresponse*. The component of the analysis weight that adjusts for unit nonresponse is called the *unit nonresponse weight.* Development of the unit nonresponse weight involves identifying, within each stratum, factors that are associated with response. The factors that we consider for PRAMS are maternal age, education, marital status, trimester of first visit for prenatal care, parity, race, Hispanic ethnicity, and unknown birthweight.

   Depending on the set of factors related to response, response categories and computed unit nonresponse weights for respondents in each category were defined. For example, if in a particular stratum marital status (defined as married, other) and maternal education (defined as >12 years, 12 years, <12 years/other) were related to likelihood of response, six response categories for unit nonresponse could be defined:

MARITAL STATUS	EDUCATION
Married	> 12 years
Other	> 12 years
Married	12 years
Other	12 years
Married	< 12 years
Other	< 12 years

An analysis weight that includes an adjustment for the selection probability (sampling weight) and the unit nonresponse (unit nonresponse weight) is computed for respondents in a particular response category. This analysis weight is the product of the sampling weight times the following ratio:

$$\frac{\text{number of sampled mothers in the response category}}{\text{number of respondent mothers in the response category}}$$

This ratio may be adjusted by a constant factor for all response categories in a single stratum to account for different sampling weights within stratum.

If all of the mothers in the category respond, the ratio is 1 and the initial weight is retained (i.e., there is no adjustment for nonresponse). Within a stratum, the size of the unit nonresponse weight is inversely related to the category-specific response rate: the lower the response rate for a particular response category, the larger the adjustment for nonresponse. For a particular stratum, the sum of unit nonresponse weights among respondents should equal the number of sampled mothers in that stratum. Likewise, the sum of unit nonresponse weights among respondents in all strata should equal the total number of sampled mothers.

Computation of the unit nonresponse weight rests on the assumption that, within a stratum and response category, the average of the answers of the respondents is the same as the average of the answers of the nonrespondents. As a rule, if there are fewer than 25 respondents in any response category, the response category should be combined with one or more other response categories until all response categories have at least 25 respondents. The goal is to have enough respondent mothers in each category so that the average of their responses is not unduly influenced by a few mothers.

3. **Adjustments for Omissions in the Sampling Frame (Incomplete Frame).** Adjustment for omission in the sampling frame is defined as a *sampling frame noncoverage weight.* In other words, adjustments are made for mothers who were not included in, or "covered" by, the sampling frame.

   The general approach for computing the noncoverage weight is similar to the approach for computing the unit nonresponse weight. The first step in developing noncoverage weights is identifying the factors related to noncoverage. Noncoverage may be related to stratum, maternal county of residence, hospital of delivery, etc. Noncoverage categories are defined depending on the set of factors related to noncoverage. For each noncoverage category, a "corrected" frame size will be computed by summing the "original" sampling frame and the missed mothers. Then, using the

corrected frame size, the noncoverage weight for the respondents in each noncoverage category will be computed.

An analysis weight that adjusts for the sampling weight, the nonresponse weight, and the noncoverage weight is computed for each respondent in the survey. This analysis weight is the product of the analysis weight computed previously times the following ratio:

$$\frac{\text{number of mothers on the "corrected frame" in the noncoverage category}}{\text{number of mothers in the original frame in the noncoverage category}}$$

If the corrected frame is very similar to the original frame, this ratio will be very close to one, and the adjustment for noncoverage will therefore be very small. In some cases, it may be so small as to hardly be noticeable. In others, a large difference between the corrected frame and the original frame will result in a large adjustment for noncoverage.

Computation of the noncoverage weight rests on the assumption that, within a noncoverage category, the average of the answers of the respondents who were included in the original sampling frame is the same as the average of the answers of the mothers who were not included in the original sampling frame.

When adjustments for sampling design, nonresponse, and noncoverage are included in the analysis weight, the sum of the analysis weights for respondents should equal the number of mothers eligible for PRAMS in the state.

Part II: Formulas for PRAMS Analysis Weights

Definitions:

N = Total number of records actually in the sampling frame

N_i = Total number of sampling frame records in stratum I

n = Total number of records in sample

n_i = Total number of records sampled for stratum I

r_i = Total number of respondents in stratum I

Then the sampling weight for a record in stratum i, Ws_i, is:

$$Ws_i = N_i / n_i.$$

Suppose several maternal characteristics are found to be associated with response in a given stratum i. Let

n_{ij} = total number of records sampled for the j-th category of nonresponse associated characteristics in stratum I

and

r_{ij} = total number of respondents in j-th category of nonresponse associated characteristics within stratum i.

Then the nonresponse weight for a record in stratum i and nonresponse associated characteristic category j, Wn_{ij}, is

$$Wn_{ij} = n_{ij} / r_{ij} .$$

Suppose it is discovered that a nontrivial number of records, N^*, were omitted from the sampling frame (most likely because they were being processed too late to make the sampling frame). Let

N' = total number of records that should have been in the sampling frame.

Then,

$$N' = N + N^* .$$ (N' is the true population of interest.)

Suppose one or more variables are found to be associated with the noncoverage of records on the sampling frame. Let

N^*_k = total number of eligible records omitted from sampling frame from k-th category of associated variables

and

N_k = total number of records actually in sampling frame from k-th category of associated variables.

Then the noncoverage weight for the k-th category of associated variables, Wc_k, is:

$$Wc_k = (N_k + N^*_k) / S,$$

where S is the sum of $Ws^*_i * Wn_{ij}$ over all respondents in category k. The analysis weight is the product of the sampling weight, nonresponse weight, and noncoverage weight.

The analysis weight for a record in the i-th stratum, in the j-th stratum category of nonresponse associated variables within stratum i, and the k-th category of noncoverage associated variables, W_{ijk}, is

$$W_{ijk} = Ws_i * Wn_{ij} * Wc_k$$

If no noncoverage adjustment is made, the analysis weight is just the product of the sampling weight and the nonresponse weight. In this case

$$W_{ij} = Ws_i * Wn_{ij}$$

$$= (N_i / n_i) * (n_{ij} / r_{ij}).$$

To verify, let

I_{ijm} = 1 for m-th respondent in the j-th category of nonresponse associated variables within stratum I

and

= 0 otherwise.

Then,

$$\text{э}_i\text{э}_j\text{э}_m W_{ij} = \text{э}_i\text{э}_j\text{э}_m I_{ijm} * (N_i / n_i) * (n_{ij} / r_{ij})$$

$$= \text{э}_i\text{э}_j (N_i / n_i) * (n_{ij} / r_{ij}) \ \text{э}_m I_{ijm}$$

$$= \text{э}_i\text{э}_j (N_i / n_i) * (n_{ij} / r_{ij}) * r_{ij}$$

$$= \text{э}_i\text{э}_j (N_i / n_i) * n_{ij} = \text{э}_i (N_i / n_i) \ \text{э}_j n_{ij}$$

$$= \text{э}_i (N_i / n_i) * n_i = \text{э}_i N_i = N.$$