4 Sampling

4.1 Definition

The population of interest for PRAMS is all mothers who are residents of <STATE> who delivered within <STATE> a live-born infant during the surveillance period. To draw a sample from this population, a *sampling frame* must be identified. The sampling frame is a list of mothers that represents the population eligible for inclusion in the sample. Often a complete list of all resident women who delivered a live-born infant within the state does not exist. Therefore, the operational sampling unit for PRAMS is *infants* who were born alive within <STATE> to resident mothers during a specified interval. <STATE>'s vital records birth certificate file serves as the best available source of the sampling frame representing live births.

By definition, the target population for PRAMS is limited to pregnancies resulting in a live-born infant. By using the birth certificate file as the sampling frame, PRAMS implicitly excludes stillbirths, fetal deaths and induced abortions. The reason for this is twofold: 1) reporting systems for these outcomes are not routinely in place in many states, and 2) the standard definitions for these outcomes vary widely.

Because of the importance of learning about the maternal behaviors of mothers whose infants have died, they will be *included* in the sampling frame. The original PRAMS states were concerned that the inclusion of sampled mothers whose infants had died or had severe health problems would affect response rates because of a possible lack of sensitivity in approaching them. PRAMS data analysis has found that the response rates for mothers of infants who have died are comparable or higher than the response rates for other mothers. The literature states, and PRAMS experience indicates, that these mothers are usually eager to participate in epidemiologic studies because they want to help researchers understand the problems that their babies experienced. PRAMS uses a separate letter for these women that expresses sympathy for their loss and encourages their participation by emphasizing that their input may be helpful to the health of other mothers and babies. Additionally, the questionnaire is sensitive to this issue and has had little difficulty in eliciting responses from this group of women.

4.2 Adjustments to the Sampling Frame

4.2a ***Exclusions.*** Exclusions are made because of particular concerns or operational difficulties in including them in the sample. These exclusions are described here.

i. ***Out-of-State Births to Residents.*** For practical reasons, the sampling frame is restricted to infants who were delivered in <STATE>. There are often substantial delays in obtaining birth certificate information from other states. In most cases these records are obtained too late to be sampled and followed up within the two to six month time frame prescribed in this protocol.

ii. ***In-State Births to Nonresidents.*** As in-state births to nonresident mothers may be difficult to trace, they will be excluded from the sampling frame. Furthermore, a state's target population for public health action does not extend beyond its residents, and information on residents would be more relevant for better serving the public health needs of <STATE>.

iii. ***Missing Information.*** Infants whose birth certificates lack the mother's last name are excluded from the sampling frame because this information is crucial for follow-up. Birth certificates that are only missing a mailing address will *not* be excluded because often it is possible to obtain that information from another source. Birth certificates missing any other information will not be excluded.

iv. ***Delayed or Early Processing of Birth Certificates.*** Birth certificates that are processed too late after the birth occurred (more than six months afterwards) are excluded from the sampling frame. Use of these records would raise concerns about recall bias, ability to locate the mother, and comparability with other respondents. Birth certificates that appear on the sampling frame less than two months after the birth occurred will be temporarily excluded. These records will be included in the sampling frame two months after the birth.

Protocol Development Task

Estimate to what extent delayed receipt of information and/or early processing of birth certificates will affect your state's sampling procedures in the format below. In the following table, this information is translated into estimates of the percentage of births in a given month that will be available for inclusion in the sampling frame by the specified number of days after birth.

Births Available on Computer File for Sampling

Number of Days After Birth	30	60	90	120	150	180	210	240
Percent of Births on File	>46	>84	>94	>97	>98	>99	>99	>99

v. ***Multiple Gestation Infants.*** The issue of multiple gestation in the sampling frame is complex. These infants are of interest because of the elevated rates of low birth weight and infant mortality. However, infants of a multiple gestation have had the same intrauterine environment, thus are not independent of one another. It is necessary to establish measures to ensure that only one infant of a multiple gestation is included in the sampling frame.

- For each twin or triplet set, *one member* will be randomly selected for inclusion in the sampling frame. If one of the deliveries results in a fetal death, the surviving twin or triplet(s) will be included. **See Section 4.4a** for the procedure used to make this selection.

- Multiple gestations involving four or more siblings, however, should be excluded from the sampling frame. Because of the rarity of multiple births of this number, little is lost in representing the population. Quadruplets and quintuplets are almost always below normal birthweight, but not because of risk factors that are of interest in epidemiologic studies. Alerted to possible complications, the mother most likely received more than the usual prenatal care. Because of the publicity that often surrounds such births, the

mother's pregnancy is unlikely to have been representative of the general experience.

vi. ***Adopted Infants.*** The sampling of adopted infants can also be a sensitive issue. The adoptive mother is often named on the birth certificate. As the majority of the survey questions involve the time period prior to and during the pregnancy, the adoptive mother is not qualified to respond to the PRAMS survey. For this reason, any adoptions that are identified as such on the birth certificate should be excluded from the sample.

It is not always possible to identify adoptions before the sample is drawn. If an adoption is not identified for exclusion prior to drawing the sample but is discovered during the data collection period, one of two situations may occur. In the first case, if the name of the adoptive mother is provided on the birth certificate, this mother should be dropped from further follow-up for the reasons given above (using the "Drop Mom" feature of PRAMTrac). If, however, the name of the birth mother is provided on the birth certificate, PRAMS staff should attempt to contact the mother and encourage her to participate. This mother is able to answer the questions regarding her experiences prior to and during the pregnancy. The survey has been designed to allow these mothers, whose infants do not live with them, to skip the questions pertaining to the child's early infancy.

Protocol Development Task

Explain how adoptions are processed in your state and explain if and how they are identifiable for exclusion.

vii. ***Surrogate Births.*** Current medical technology has made it possible for couples to have a baby through surrogacy. Unlike the process of adoption, no legal consensus has been reached about surrogate motherhood in the United States. As a result, the sampling of an infant delivered by a surrogate carrier (or gestational carrier) can be a sensitive issue. Under certain circumstances the intended mother (the woman who will raise the child) is named on the birth certificate. Since the majority of survey questions involve the time period prior to and during the pregnancy, the intended mother is not qualified to respond to the PRAMS survey. For this reason, any births that can be identified as such should be excluded from the sample.

If a surrogate birth is not identified for exclusion prior to drawing the sample but is discovered during the data collection period, one of two situations may occur. In the first case, if the name of the intended mother is provided on the birth certificate, this mother should be dropped from further follow-up for the reasons given above (using the "Drop Mom" feature of PRAMTrac). If, however, the name of the surrogate carrier is provided on the birth certificate, PRAMS staff should attempt to contact this woman and encourage her to participate. This woman will be able to answer the questions regarding her experiences prior to and during the pregnancy. The survey design will also allow these women to skip those questions pertaining to the child's early infancy.

Protocol Development Task

Explain how surrogate births are processed in your state and explain if and how they are identifiable for exclusion.

viii. Additional State-Required Exclusions.

Protocol Development Task

Your state may have additional exclusions that are required by law or because of the nature of your vital records processing. If you have any state-required exclusions, list them here beginning with $4.2a_{viii}$. Define what the exclusions will be and estimate their expected numbers.

4.2b Assessment of the Exclusions on the Generalizability of the Sampling Frame. It is important to understand the effects of exclusions from the frame on potential generalizability of the results to the total birth population. We anticipate that the magnitude of the exclusions listed above will be minimal with the possible exception of out-of-state births to residents. The tables below show sample calculations of one state's exclusions by stratification variable:

Example Table 4.2b$_i$
1992 Resident Births

Stratification Variable	Total Births	Singleton Births/Mothers	Number of Births With Plurality = 2	Mothers Giving Birth to Twins	Number of Births With Plurality =3	Mothers Giving Birth to Triplets	Number of Births With Plurality>=4	Total PRAMS Mothers
URBAN AREA								
Low Birthweight	657	600	50	25	3	1	4	626
Normal Birthweight	10,909	10,747	150	75	12	4	0	10,826
TOTAL Urban Area	11,566	11,347	200	100	15	5	4	11,452
RURAL AREA								
Low Birthweight	775	678	88	44	9	3	0	725
Normal Birthweight	8,000	7,814	168	84	18	6	0	7,904
TOTAL Rural Area	8,775	8,492	256	128	27	9	0	8,629
TOTAL								
Low Birthweight	1,432	1,278	138	69	12	4	4	1,351
Normal Birthweight	18,909	18,561	318	159	30	10	0	18,730
GRAND TOTAL	20,341	19,839	456	228	42	14	4	20,081

Source: *State Vital Statistics Report, 1992*

Example Table 4.2b$_{iii}$
1992 Resident Births Born Inside State

Stratification Variable	Number With Complete Information	Row / Col Pct Complete Information	Number With Missing Information	Row / Col Pct Missing Information	Total	Row / Col Pct Total
URBAN AREA						
Low Birthweight	657	99.7 / 3.2	2	0.3 / 11.1	659	100.0 / 3.2
Normal Birthweight	10,909	99.9 / 53.6	7	0.1 / 38.9	10,916	100.0 / 53.6
TOTAL Urban Area	11,566	99.9 / 56.9	9	0.1 / 50.0	11,575	100.0 / 56.9
RURAL AREA						
Low Birthweight	775	99.6 / 3.8	3	0.4 / 16.7	778	100.0 / 3.8
Normal Birthweight	8,000	99.9 / 39.3	6	0.1 / 33.3	8,006	100.0 / 39.3
TOTAL Rural Area	8,775	99.9 / 43.1	9	0.1 / 50.0	8,784	100.0 / 43.1
TOTAL						
Low Birthweight	1,432	99.7 / 7.0	5	0.3 / 27.8	1,437	100.0 / 7.1
Normal Birthweight	18,909	99.9 / 93.0	13	0.1 / 72.2	18,922	100.0 / 92.9
GRAND TOTAL	20,341	99.9 /100.0	18	0.1 /100.0	20,359	100.0 /100.0

Source: *State Vital Statistics Report, 1992*

Example Table 4.2b$_v$
1992 Resident Births Born Inside State With Complete Information

Stratification Variable	Total Births	Singleton Births/Mothers	Number of Births With Plurality = 2	Mothers Giving Birth to Twins	Number of Births With Plurality =3	Mothers Giving Birth to Triplets	Number of Births With Plurality>=4	Total PRAMS Mothers
URBAN AREA								
Low Birthweight	657	600	50	25	3	1	4	626
Normal Birthweight	10,909	10,747	150	75	12	4	0	10,826
TOTAL Urban Area	11,566	11,347	200	100	15	5	4	11,452
RURAL AREA								
Low Birthweight	775	678	88	44	9	3	0	725
Normal Birthweight	8,000	7,814	168	84	18	6	0	7,904
TOTAL Rural Area	8,775	8,492	256	128	27	9	0	8,629
TOTAL								
Low Birthweight	1,432	1,278	138	69	12	4	4	1,351
Normal Birthweight	18,909	18,561	318	159	30	10	0	18,730
GRAND TOTAL	20,341	19,839	456	228	42	14	4	20,081

Source: *State Vital Statistics Report, 1992*

Protocol Development Task

1. After selecting your stratification variables **(Section 4.3b)**, return to this task box. Create tables identical with the examples above for *each* of your state's exclusions numbered $4.2a_i$, $4.2a_{iii}$, $4.2a_v$ - $4.2a_{vii}$ and for any additional exclusions identified beginning with $4.2a_{viii}$.

2. For residents giving birth out of state ($4.2b_i$), records with missing information ($4.2b_{iii}$), adoptions ($4.2b_{vi}$), and surrogate births ($4.2b_{vii}$), describe the differences between the strata for the excluded and the included (the sample frame) records for the exclusions that are of substantial magnitude that they may affect the population. What proportion do the excluded records represent of the total birth population of interest?

   Example: Of 22,071 births, 7.8% (n = 1,712) were born outside the state. This percentage varied by level of Geographic Area: within Urban, 9.7% (1,250/12,825) were born outside the state; within Rural, 5.0% (462/9,246) were born outside the state. Additionally, within the levels of Geographic Area, the proportion born outside the state varied by Birthweight. On the other hand, the percentage of resident births born outside the state was more consistent by Birthweight, 8.2% (129/1,566) within Low Birthweight and 7.7% (1,583/20,505) within Normal Birthweight.

   With respect to Geographic Area, the excluded records have a significantly higher proportion of Urban than the included records, 73.0% vs. 56.9%. With respect to Birthweight, there do not appear to be differences in the distribution when comparing within-state and out-of-state occurrences.

3. If there are differences in the sampling frame regarding stratification characteristics caused by an exclusion as compared with the true birth population of the state, then the sample will not be an accurate reflection of the population it is meant to represent. For residents giving birth out of state ($4.2b_i$), records with missing information ($4.2b_{iii}$), adoptions ($4.2b_{vi}$), and surrogate births ($4.2b_{vii}$), explain how the generalizability of your estimates will be influenced by these differences.

   Example: Although the out-of-state occurrences differ from within-state occurrences with respect to Geographic Area, there are not sufficient numbers of out-of-state occurrences to significantly affect the remaining distribution when they are excluded. Comparing the column percents of column 2 with column 6 of Table 4.2b, the distribution of Birthweight is an accurate representation of the total birth population. There will be a small underrepresentation of the Urban records. However, the total state distributions by Geographic Area and Birthweight are not altered appreciably by limiting the sample to within-state occurrences. Therefore, despite this exclusion, findings from analyses are still generalizable to the total state birth population.

4. Place your newly developed tables describing each exclusion ($4.2a_i$, $4.2a_{iii}$, $4.2a_v$-$4.2a_{vii}$) following the description of that exclusion.

4.3 Sampling Plan

4.3a Why Use Stratified Sampling? For PRAMS surveillance, there is often a particular interest from a public health perspective in certain subpopulations. These subpopulations may not represent a large portion of a state's overall population. To make inferences about specific subpopulations and make comparisons among several subpopulations, infants in those subpopulations (commonly called *strata*) will need to be oversampled (i.e., sampled at a higher rate than other subpopulations).

The main advantage of *stratified sampling* is that it permits separate estimates of subgroups of interest and permits comparisons across these subgroups.

An alternative to a stratified sample is a *proportional sample*. With proportional sampling, specific subpopulations are represented in the sample in the same proportion as they are represented in the actual population. Thus, if 3% of a state's births are low birthweight, about 3% of the sample would be expected to be low birthweight. In the context of PRAMS, the main advantages of proportional sampling over stratified sampling are that weights are not always required, which makes the analysis of such data simpler, and statewide prevalence estimates will be more accurate and have smaller variance.

For PRAMS, however, the ability to make inferences about subgroups of high public health interest and to make comparisons across the groups is important. This need outweighs the difficulties of computing and using sampling weights. Therefore, PRAMS incorporates a stratified sampling scheme.

4.3b Stratification Scheme. The sampling plan is designed so that inferences about prevalence rates for maternal behaviors can be estimated with sufficient precision both at the state level and within selected strata. The choice of stratification variables is limited to the information available on birth certificate records. For reasons of data quality and the efficiency of estimates, this choice of variables is further limited to information reported with high accuracy.

Protocol Development Task

Your state may choose up to two stratification variables from the following list:

Birthweight	Maternal Age
Maternal Race and Ethnicity	Geographic Area
Maternal Education	Medicaid Status

To limit annual samples to a manageable size, each stratification variable can have from two to four levels, but the total number of strata cannot exceed six. For example, you may choose to have one stratification variable with three levels; two stratification variables with two levels each; or two stratification variables, one with three levels and one with two levels. In addition, separate strata may be created for unknowns.

States are strongly encouraged to choose a single stratification variable in their first year of data collection to help keep the sample manageable. The sampling plan can be modified in the second year of operations to stratify by an additional variable, depending on the state's success in contacting hard-to-reach mothers.

Any stratification variable you define must be included in birth certificate sample and frame files if not part of the core variables. These variables must further be added to the state-specific portion of the file.

Provide the following information where relevant in the preceding section:

1. Description and detailed rationale for the stratification scheme selected.

2. Definitions of all strata.

4.3c ***Determining Sample Size.*** Required sample sizes for PRAMS are determined in relation to the given proportion that is being estimated, at a given level of precision, and with a given level of statistical confidence. For specified levels of precision and confidence, the sample size required is at its maximum when the advance estimate (the number used in sample size calculations) of the proportion being estimated equals 0.50. PRAMS data are used in estimating proportions for risk factors that range from common (such as delivery paid for by Medicaid) to rare (such as drinking alcohol during pregnancy). Using 0.50 in sample size calculations leads to the largest sample sizes for a specified level of precision and confidence, whatever the true population proportions are for the various risk factors.

Based on the stratification measures found above, a sample size of about 400 (n = 400) is necessary in each stratum to estimate a dichotomous variable with reasonable precision and 95% confidence, assuming an infinitely large population size (N). The assumption of an infinitely large population will be violated in the oversampled strata. In any stratum where the desired sample size of 400 comprises more than 5% to 10% of the population, it is appropriate to apply the Finite Population Correction (FPC). The FPC will reduce the desired sample sizes in such cases without compromising the precision of the estimates.

The formula for FPC is:

adjusted size = n / (1 + (n/N)),

where n = desired sample size,

N = population size.

Mothers in some strata may be more difficult to contact than mothers in other strata. Thus, actual stratum sample sizes must be larger than theoretically needed to achieve a given level of statistical power. Based on the estimated stratum-specific response rates, the stratum-specific sample sizes will be inflated to ensure an adequate number of responses for analysis.

This practice, as noted by Don Dillman, author of *Mail and telephone surveys: The Total Design Method*[1] and a consultant to PRAMS during its initial development, amounts to substituting "an available respondent for one that cannot be contacted." Dillman quotes the statistician W. Edward Deming, who said that "Substitution does not help; it is only equivalent to building up the size of the initial sample, leaving the bias of nonresponse undiminished." To assess validity, one cannot compensate for a low response rate by inflating the numbers in the strata. Increasing the sample size does reduce the random component of error in estimates obtained from PRAMS. Nonrandom or systematic error from response bias can best be reduced by improving response rates.

When calculations lead to a high sampling fraction (greater than half) and the cost is not prohibitive, common survey practice is to select all the records in a stratum. Small strata are sometimes volatile, and creating a take-all stratum guards against the shortfall in responses that an unexpected decline in births could cause. If such a decline is unlikely,

[1] Dillman DA (1978). *Mail and telephone surveys: the total design method*. New York: Wiley-Interscience

however, 100% sampling is usually unwarranted. Increasing the sample beyond the calculated initial size in small strata brings little gain in precision for the expense incurred. Given typical PRAMS states' budgets and standards for precision, 100% sampling is out of the question if the stratum population size exceeds 571 (minimum sample size of 400 inflated by minimally acceptable response rate of 70%). Even if the stratum population size is below 571, 100% sampling is inadvisable because it is possible to get precise estimates without using 100% of the sample.

In the context of a state-wide, population-based surveillance system like PRAMS, it may not be reasonable to select as strata demographic groups comprising less than 2% of the population or fewer than 500 women. States must balance competing priorities and give careful consideration to precluded opportunities when choosing groups to oversample.

To reduce nonresponse bias, one of the components of the analysis weights (See **Chapter 6, Section 6.8 and Appendix B** for a discussion of analysis weights) adjusts for nonresponse patterns because response rates vary among strata. However, weights may not adequately compensate for low response rates. The nonresponse weight assumes that the average of the answers of the respondents within a particular stratum and response category under consideration is the same as the average of the answers of the nonrespondents in that stratum and response category. Although the assumption seems reasonable for strata with high response rates, it becomes increasingly implausible for strata with lower response rates. For strata with response rates below 50%, this assumption is unjustified.

Protocol Development Task

1. For each stratification variable, provide the following information where relevant:

   a. Determine the estimated annual population and proposed annual sample size for each stratum. Provide the rationale for all sample sizes.

   b. Determine the anticipated response rates and provide rationale.

   c. Determine the level of precision for the anticipated analyses.

   NOTE: Response rates are typically influenced by two main factors: the study population and the methodology employed. If your state has conducted other research studies on similar populations, you may be able to use response rates from those studies to estimate response rates for PRAMS. However, if the study population is different from the PRAMS target population, results from these studies

may not be useful predictors for PRAMS. The mail/telephone methodology used by PRAMS typically elicits higher response rates than those encountered with most mail-out surveys. Response rate results from studies that used other modes of data collection may not be comparable. Based on the experiences of states conducting PRAMS, the expected response rates average between 70% and 75% for high-risk strata and between 75% and 85% for low-risk strata, and the annual samples will range from 1,600 to 2,500 infants.

2. Provide any details on response rates from other studies if you feel they are comparable.

3. Provide a table showing derivation of desired sample sizes for each stratum. Include the annual birth distribution, the expected sample for each stratum, and adjustments to the sample size. (See the example below.)

The following table illustrates this state's figures and provides an example of how your state can choose to present your figures.

Example Table 4.3.c
Calculation of Annual Sample Size

Stratum	1992 Resident Births (from Table 4.2b$_i$)	Resident Births Minus Exclusions (from Table 4.2b$_v$)	Estimated Annual Unadjusted Sample Size	FPC Corrected Sample Size#	Estimated Response Rates	Estimated Annual Sample Size Adjusted for Non-response*
Urban, Low Birthweight	745	626	400†	244	70%	349
Urban, Normal Birthweight	12,080	10,826	400†	N/A	70%	571
Rural, Low Birthweight	821	725	400†	258	70%	369
Rural, Normal Birthweight	8,425	7,904	400†	N/A	75%	533
Total	22,071	20,081	1,600	1,302		1,822

Finite Population Correction (FPC)

* Adjusted sample sizes calculated by dividing unadjusted sample sizes (or FPC when applicable) by response rates (as proportions).

† Sample size estimates calculated assuming a risk factor proportion of 0.50, desired precision of +/-0.05, and 95% confidence interval. Calculations use formula from Levy, P.S. and S.L. Lemeshow, *Sampling for Health Professionals*, Belmont, California: Lifetime Learning Publication, 1980.

Source of birth data (column one): *State Vital Statistics Report,* 1992.

4.3d Evaluation of Sample Size. Because birth distributions and response rates can change over time, CDC has developed procedures for evaluating the sampling scheme on a regular basis and making modifications if necessary. For further information regarding the evaluation of sample size procedures see **Section 11.1a.**

4.4 Selection of Sample

Producing a sample of mothers has two steps, constructing the frame and drawing the sample.

4.4a Preparation of the Frame. <STATE>'s vital records birth certificate file serves as the source of the sampling frame representing live-born infants. Records not yet processed are not available for inclusion on the frame.

Records identified as exclusions as listed in **Section 4.2** are excluded as part of the construction of the frame. No infant should be included in the sampling frame twice. The sampling frame will be routinely checked for duplicate records to eliminate the possibility of an infant being included twice.

Protocol Development Task

Describe in detail the birth registration process including time frames for this process. If there are separate procedures for electronic birth certificates, describe them as well. Specify the proportion of certificates that are obtained by each method that will be available for sampling during a range of time periods after birth. It may be useful to display this information in a flowchart (see example below). Be sure to describe when and how the identifying number (sometimes referred to as the state file number or birth certificate number) is added to the records. Describe exactly where the PRAMS sampling procedures take place in the registration process.

The specifics of the way your state processes records may vary, but your protocol should include comparable detail as shown in the example below.

Example of a State Birth Registration Process

“The Health Department's Center for Health Statistics is responsible for collecting, processing, coding, filing, and certifying Vital Records, including birth and death certificates. Although records are handled as monthly batches, incoming certificates are processed daily by the Division of Vital Records and are keyed and added weekly to the mainframe file by Informational Systems.

“Birth certificates are completed at the hospital by hospital personnel, who are usually in the Medical Records Department. Less than 0.4 percent of births occur outside hospitals, with most of these being filed within 30 days of birth. Certificates are mailed by the hospital within five days of birth to the Center for Health Statistics, except for counties A, B, and C, where they are mailed to the local Vital Statistics Registrar in the county health department where the birth occurred. The certificates are reviewed daily for errors, completeness, and appropriate signatures. If the certificate is not correct, it is sent back to the hospital. These 'send-back' certificates are monitored to be sure that they are corrected and returned.

"Daily, as soon as the certificates are reviewed for completeness, they are given a state file number. The second day after receipt in Vital Records, certificates are sent to the Coding/Quality Control Section in Vital Records for coding of

demographic and medical items. Any obvious data inconsistencies are queried at this time.

"The third day after receipt in Vital Records, certificates are sent to Information Systems for keying of all information. In the future the Vital Records division will be keying certificates. (Legal and medical/statistical data are keyed at the same time, so linkage of medical and legal portions is not necessary.) After being keyed, certificates are returned to Vital Records, where the complete certificate is microfilmed. The confidential medical/statistical part of the birth certificate is then removed from the legal portion of the certificate, and the legal portion is re-microfilmed for certification purposes. The legal portions of the certificates are then placed in groups of 500 in special vital records boxes for storage in the vault.

"Weekly, Information Systems loads the keyed birth information to a disk file on the state's centralized mainframe computer and creates a cartridge of the year-to-date file for the Statistical Analysis Division in Health Statistics to use for analysis and quality control. Also at this time, edits are run for the Coding/Quality Control Section to check certificates for errors and inconsistent information. Coding/Quality Control staff query hospitals about inconsistent data.

"For 1991 records, edits are being rewritten by Information Systems to edit records one time for all errors. This edit procedure will eliminate the need to look up records more than once and will enable Coding/Quality Control staff to process edits more quickly. The current system for correcting errors requires that Coding/Quality Control staff complete a form that is sent to Information Systems to have the correction keyed, then batch added to the mainframe file.

"In the past, Health Statistics has had problems getting support from Information Systems to write programs to produce the monthly data file. Now, however, additional staff have been hired by Information Systems to provide needed support. Information Systems is currently developing an on-line system for Coding/Quality Control to key the corrected birth information directly onto the disk file. The new on-line update program should be completed in the next few months, and will allow any keying or coding errors to be corrected within two weeks of receipt of the certificate.

"Procedures for processing birth certificates vary slightly for counties A and B, which have about 27 percent of the births that occur in the state. These counties number their certificates with the state file number (obtained by calling the Vital Records Division for the file numbers for the batch to be numbered) and code their certificates following state requirements before sending them weekly to Vital Records. After these certificates arrive in Vital Records, they are checked, and then added to the batch to be keyed the next day.

"To improve the process of filing birth certificates, Health Statistics staff are currently working on obtaining an Electronic Birth Certificate (EBC) software package. This software will be placed on a personal computer at the hospital where the birth certificate is completed. Since the software will have multiple edits and will not allow blanks, the quality of data should improve greatly. Also, since "white-out" will not be needed because certificates can easily be reprinted if a mistake is made, most "send-backs" will be eliminated. In addition, data will be downloaded to a diskette or transmitted over a modem directly to Vital Records. Paper copies of birth certificates will also be sent directly to Vital Records. In Vital Records, paper copies will be given the state file number.

"Staff in Coding/Quality Control will add the state file number to the digitalized data received from the hospital through personal computers located in Vital Records. Coding/Quality Control staff will also resolve any coding problems that hospital staff could not handle. After problems are resolved, data will be uploaded from the personal computers to the mainframe file. Data from hospitals not yet on the EBC system will continue to be processed as described above. Editing and quality control procedures will continue as described above. Editing and quality control procedures will continue to be done for all data on the file. As more hospitals use the system, less coding, keying and editing will need to be done by Health Department staff.

"To insure that certificates for all births are sent to Vital Records, hospitals are required to send Vital Records a list of all births that occur in their facility each month. At the end of each processing month, a computer listing sorted by hospital is run showing all births on the file. This listing is compared to the lists sent by individual hospitals to identify missing records. Records found to be missing are turned over to Field Program staff in Vital Records for follow-up. Staff contact the hospital to learn why the record was not filed. If necessary, Field Program staff personally visit the hospital to obtain the certificate. This procedure virtually eliminates missing certificates, since over 99 percent of all births occur in hospitals.

"Many other edit checks are done monthly. Inconsistencies among items are checked. Plural births are checked to be sure all certificates have been filed, and infant deaths are also checked to be sure births have been received. High-risk births are checked to obtain the status of the infant at time of release from the hospital. This procedure catches a number of infant deaths that might not otherwise be filed. Questionable data such as mother's age less than 15 and very low birth weights or gestational ages are examples of items verified.

"Questionable items that have been queried by staff are updated on the computer file whenever the information is received. To insure that no errors are made, the editing process is repeated any time a record is updated during the year. Our state is also part of the Vital Statistics Cooperative Program (VSCP) of

the National Center for Health Statistics (NCHS) and must maintain quality to meet contract requirements. Since we have been part of the VSCP, we have consistently met quality requirements specified in the contract and no information has had to be recoded or rekeyed.

"At the end of the processing month, tape files are created for the VSCP and also for the Social Security Enumeration Project. Monthly data are completed by the end of the month following the month of the event. A sample of records filed shows that 46 percent of the monthly certificates are on the computer file one week after the end of the birth month; 84 percent one month after the end of the birth month; 97 percent three months after the end of the birth month; and 99 percent five months after the end of the birth month.

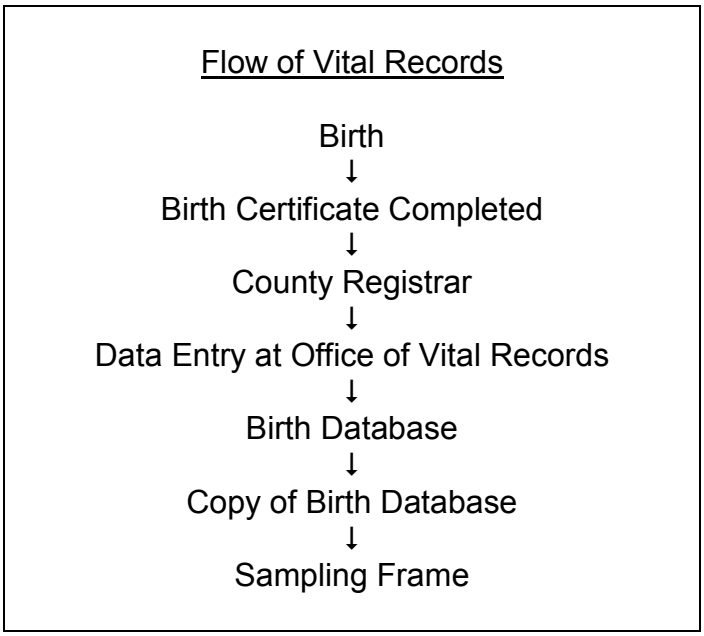
"The PRAMS sample will be drawn once the records have been received, coded, keyed, and edited."

As part of the process of developing the sampling frame, the file of birth certificate records is divided into strata, from each of which a sample is drawn. Preparing the frame also includes deleting the records for all but one infant in each multiple birth. One easy manner to accomplish this selection involves basing the selection on birth order and date of birth (DOB). For example, for twins, choose the first-born if the DOB lies between the first and fifteenth of the month; otherwise choose the second-born. Similarly, the month can be broken down into three segments for triplets. This selection process is used only if all the deliveries resulted in a live birth. If only one twin was born alive, it is treated as a singleton birth for sampling purposes.

Protocol Development Task

1. Describe the preparation of the records before sampling. The procedures for preparing the records before sampling will include development of the stratum-specific sampling frames, identifying and eliminating exclusions, and dealing with records for multiple births.

2. The following two flow charts show signal events in the processing of birth certificate records and preparation of the frame. The first example shown is of the process utilized for a paper certification system. Please adapt the flow chart that follows to your state-specific situation. If you process your birth certificates via EBC or use a combination of both, include a flowchart for each.



This flowchart describes the typical flow of the vital records process. The process begins with a birth. The birth certificate is completed and forwarded to the county registrar. From the county registrar data entry is performed by the Office of Vital Records which then establishes a birth database. From the birth database a sampling frame is drawn.

FLOWCHART OF COMPUTER PROCESS FOR SELECTING PRAMS SAMPLING FRAME AND SAMPLE (for states using mail/telephone surveillance)				
				READ BIRTH RECORD
				↓
EXCLUDE	←	YES	←	IS BIRTH TO OUT-OF-STATE RESIDENT?
				↓ NO
EXCLUDE	←	YES	←	DID BIRTH OCCUR IN ANOTHER STATE?
				↓ NO
EXCLUDE	←	YES	←	DID BIRTH OCCUR MORE THAN 6 MONTHS BEFORE SAMPLING DATE?
				↓ NO
EXCLUDE FROM THIS SAMPLING FRAME AND INCLUDE IN NEXT ELIGIBLE SAMPLING FRAME	←	YES	←	DID BIRTH OCCUR LESS THAN 2 MONTHS BEFORE SAMPLING DATE?
				↓ NO
RANDOMLY SELECT ONE FOR INCLUSION IN SAMPLING FRAME	←	YES	←	IS THIS BIRTH IN A MULTIPLE GESTATION?
↳				↓ NO
				1. ASSIGN TO ONE OF THE SAMPLING STRATA 2. EXTRACT BIRTH CERTIFICATE INFORMATION
				↓
NOT IN SAMPLE	←	NO	←	SELECTED FOR SAMPLE?
				↓ YES
				EXTRACT INFORMATION FROM THE SAMPLE FILE TO PRODUCE: 1. INITIAL CONTACT LETTERS 2. FOLLOW-UP LETTER 3. MAILING LABELS

This flowchart describes the computer process flow for selecting the PRAMS sampling frame and sample for states using mail and telephone surveillance. The birth record is read to determine if the birth is to an out-of-state resident. If the answer is Yes, the record is excluded. If the answer is No, the next question is whether the birth occurred in another state? If the answer is Yes, the birth record is excluded. If the answer is No, the next question is did the birth occur more than 6 months before sampling date? If the answer Yes, the birth record is excluded. If the answer is No, the follow-up question is did the birth occur less than 2 months before the sampling date? If the answer is Yes, the birth record is excluded from this sampling frame and included in the next eligible sampling frame. If the answer is No, the follow-up question is the birth in a multiple gestation? If the answer is Yes, randomly select one for inclusion in the sampling frame. If the answer is No, assign to one of the sampling strata and extract birth certificate information. Next question, is the birth record selected for the sample? If the answer is No, the birth record is not included in the sample. If the answer is Yes, extract information from the sample file to produce the initial contact letters, the follow-up letter and the mailing labels.

4.4b Selecting the Sample. The selection procedures must satisfy the probability requirements of the sample. The sample is chosen so that, within each stratum, each record has an equal probability of being selected. Based on these probabilities, weights can be determined for statewide estimates.

Systematic sampling within each stratum is used. The probability of being selected, commonly referred to as the sampling fraction (*1/f*) for a given stratum, is based on the estimated size of the sampling frame for the stratum and the desired sample size for the stratum. *f* is computed by dividing the population size by the sample size (i.e., $f = N/n$). A random number between 1 and *f* is chosen, and that record, as well as every *f-th* record thereafter, is selected for the sample.

Protocol Development Task

Compute sampling fractions for each stratum, and display them in table form, as illustrated in the following example.

This table displays the sampling fractions and the resulting estimated operational sample sizes for the two stratification variable examples displayed earlier.

Example Table 4.4b
PRAMS Sampling Fractions and Estimated Sample Sizes by Stratum
(Based on 1992 Vital Records Data)

Stratum	PRAMS Population Size (from Table 4.3c)	Estimated Adjusted Sample Size (from Table 4.3c)	f = N/n	f = N/n expressed as common fraction	Operational Sample Size, Annual	Operational Sample Size, Monthly
Urban, Low Birthweight	626	349	1.79	7/4	358	30
Urban, Normal Birthweight	10,826	571	18.96	19	570	48
Rural, Low Birthweight	725	369	1.96	2	363	30
Rural, Normal Birthweight	7,904	533	14.83	59/4	536	45
Total	20,081	1,822			1,827	153

4.5 Frequency and Timing of Sampling

The timetable for sampling is expressed in terms of actual calendar time and in terms of the approximate length of time after delivery. The particular day chosen to sample may depend on vital records processing. The schedule for drawing the sample permits births to be sampled within two to six months after delivery. However, the ideal sampling period is two to four months after delivery. To collect information about factors that occur in early infancy, mothers are contacted no earlier than two months after delivery. Sampling on a monthly basis ensures that mothers are contacted in a timely manner and that there is a balanced workload for PRAMS staff.

Protocol Development Task

Describe when the sampling frame will be constructed and when the sample will be drawn. Estimate the number of birth certificates that will be missed because of delayed registration with the sampling frame you have established.

Complete this statement:

A sample will be drawn on the _____th day of each month. This will make the records _____ to _____ months old.