

7 Analysis, Use, and Limitations of PRAMS Data

Analysis and use of the PRAMS data is the culmination of establishing and operating PRAMS surveillance. However, many of the tasks associated with analysis and data use must be considered before the inception of the surveillance system operation, however. An analysis plan supports the effective use of PRAMS data to plan or evaluate programs, affect state policies, or identify program priorities. Beginning the development of the analysis plan early in the process of establishing PRAMS can be helpful in deciding on stratification designs (see **Chapter 4 (Sampling)**), state-specific questions (see **Chapter 5 (Data Collection)**), and state-specific birth certificate variables (see **Chapter 6 (Data Management)**). This chapter discusses the tasks associated with analysis.

7.1 Analyses by the State and CDC

7.1a Role of the State in Data Analysis. PRAMS states are responsible for analyzing their PRAMS data. CDC may collaborate with members of the <STATE> PRAMS team in analyzing a topic of interest to them, but CDC will not independently undertake analysis of a single state's data without first consulting with the state. CDC is available to provide technical assistance and review of data analysis from each state. State-level analyses may include:

- Prevalences of selected health behaviors or outcomes that address the year 2010 or 2020 health objectives or Title V Block Grant needs assessments.
- Descriptive analyses of particular topics for state reports, newsletters, and other publications or presentations.
- Descriptive or analytic articles for *MMWR* or state journals.
- Periodic evaluation of data collection methods, stratum-specific response rates, and characteristics of nonrespondents.

7.1b Role of CDC in Data Analysis. CDC is responsible for comparisons of data between states, and descriptive and analytic studies of selected topics or methodologic issues using data aggregated across multiple states. CDC analyses may include:

- Periodic reports of selected prevalences of health behaviors or outcomes from all states.
- Descriptive and analytic studies of particular topics for publication in journals or presentation at national meetings.

- Descriptive and analytic articles for *MMWR*.
- Description and analysis of PRAMS methodology.

7.2 Preparing for Analysis

Before any analyses of PRAMS data can be undertaken, several preparatory activities must take place. Some of these activities, including the preparation of the analysis data files, are described in detail in **Chapter 6 (Data Management)** and are reviewed here. The activities discussed below are the fundamental components of planning and conducting analyses.

7.2a Preparing the Data Analysts. All individuals involved in the analyses of PRAMS data should be familiar with the surveillance system; its goals, purpose, data collection methods, questionnaire topics, and limitations. If possible, primary collaborators in the analyses should participate in developing and selecting the state-specific questions and in developing the analysis plan. The analysts must understand the process of creating the PRAMS master analysis data file (see **Section 6.8**), each component of the file, the variables contained in the file, and the source of the variables. To ensure analysts have the necessary knowledge, they should have:

- A copy of the state's protocol, especially Chapters 4 - 8.
- Codebooks for all of the variables in the analytic file, including questionnaire, birth certificate, and operations variables.
- A summary of the PRAMS methodology, such as the one in the CDC Surveillance Report.
- Reports sent with the weighted data sets, including item and unit non-response reports.
- If available, evaluation summaries for the core, standard, and state-developed questions to be analyzed.

Analysts should also review the **Guidelines for State Analysis of PRAMS Data**, which can be found on the **PRAMS SharePoint site**.

In addition, analysts should be trained in the software used for analysis of PRAMS data, in most states, SAS and SUDAAN. For more information concerning the use of SAS and SUDAAN, see Chapters 5-6 in the **Guidelines for State Analysis of PRAMS Data** or the SAS and SUDAAN examples page of the **PRAMS SharePoint site**.

The CDC PRAMStat online data query system is available to the general public, and accessible via the CDC PRAMS website (www.cdc.gov/prams) for quick analyses of most core PRAMS indicators.

7.2b Preparing the PRAMS Data. The master analysis data set is the file from which individual project or topic-specific analysis files will be created. This SAS data set is prepared by CDC for each state and contains information from the birth certificate, operations, and questionnaire files. It includes several computed variables and the analysis weights required by SUDAAN. Several criteria need to be met for the analysis data file to be prepared. These criteria, together with the computation and use of analysis weights, are discussed in **Section 6.8 and Appendix B**.

7.3 Features of Analysis

PRAMS data can be analyzed in many ways. The specific objective of the analysis will determine the type of analysis to be conducted and the population included in analysis.

7.3a Types of Analyses. At least three analytic study designs may be used in the analysis of PRAMS data, including descriptive, inferential, and methodologic designs. Analyses may be performed on the entire data set or on subsets of the entire data set that are created for a particular analysis.

i. Descriptive Studies. These analyses describe a health problem or behavior in a state. They include estimating the prevalence of a problem, examining trends over time, or describing the characteristics of women who experience the problem or behavior. Most information requests received by PRAMS projects will be descriptive studies.

ii. Inferential or Analytic Studies. These studies investigate relationships between behaviors or health outcomes in order to make inferences about possible causes or contributing factors to health problems.

iii. Methodologic Studies. These analyses assess and evaluate the PRAMS data collection methodology to determine how to improve it, such as increasing response rates or improving the validity of the data collected.

7.3b Levels of Analyses. PRAMS data can be analyzed for several population levels. These include statewide, stratum-specific, and domain.

i. Statewide Analyses. Statewide analyses describe the experience of <STATE> as a whole. They can provide overall estimates of prevalences of behaviors or health problems that are useful for monitoring trends and assessing <STATE>'s need for services. Prior to the 2007 data year, a

weighted response rate of 70% was recommended for statewide analysis. From 2007 forward, the recommended weighted response rate is 65%.

- ii. **Stratum-Specific Analyses.*** Statewide results may not be the same as those for particular subpopulations within <STATE>. In developing <STATE>'s sampling scheme, subpopulations of particular public health interest were selected as strata. Health behaviors may vary between the strata and <STATE> as a whole, as well as among the strata themselves. Stratum-specific analyses provide important information about each separate subpopulation or stratum and for comparing behaviors between strata. Prior to the 2007 data year, response rates of 70% for the individual stratum were recommended for stratum-specific analysis. Starting with 2007 data, the response rate recommended for stratum-specific analysis is 65%.
- iii. **Domain Analyses.*** A third type of analysis examines health behaviors and experiences within domains, which (for the purposes of PRAMS) are defined as subpopulations that are not strata. As in stratum-specific analyses, domains can be used to gain insights into subgroups within <STATE> that may be different from each other and from <STATE> as a whole. Domains of frequent interest to PRAMS states are women of specific racial groups, age groups, (e.g. teenagers or older mothers), and women living in small geographic regions such as counties or cities. See **Sections 7.4a and 7.4b** for issues to be considered when planning domain analyses.

7.4 Analytic Considerations

- 7.4a Appropriate Software.*** PRAMS uses a stratified sampling scheme to ensure that subpopulations of women within each state are present in the sample in sufficient levels for analysis. The probability of being selected for the sample varies among women in <STATE>. Analysis weights are used to adjust for these differences, which result from the sampling design, nonresponse pattern, and omissions from the sampling frame. Standard software packages can incorporate weights and provide accurate point estimates, but most procedures provide incorrect variance estimates and statistical tests because the calculations used by those packages assume a simple random sample. SUDAAN (Software for the Statistical Analysis of Correlated Data, Research Triangle Institute, North Carolina) or other software for analysis of complex survey data (e.g., STATA) should be used to analyze PRAMS data. SAS Version 8 and higher has some procedures for analyzing survey data. Information on these procedures can be found by searching the online manuals for 'survey'. CDC provides technical assistance only for SUDAAN.
- 7.4b Adequate Sample Size.*** A planned analysis should include at least 30 women, and depending on a state's sampling design, some analyses may require as many as 50 women to produce stable estimates. State and strata sample sizes are selected to ensure that analysis at these levels can be conducted on most

topics with one year of data. Some domain analyses or analyses of rare behaviors or health outcomes may require combining two or more years of data. For information on combining multiple years of data, see the **Guidelines for State Analysis of PRAMS Data**.

7.4c Adequate Response Rates. For any level of analysis, a 70% response rate in the overall population or subpopulation being analyzed was recommended prior to the 2007 data year. From 2007 forward, a 65% response rate in the overall population or subpopulation being analyzed is recommended. For domain analyses, the response rate for the domain should be determined, if possible, i.e. if the defining characteristic is from the birth certificate. When presenting descriptive or stratified analysis within these populations, some of the subgroups may have response rates below 70% or 65% starting with 2007 data. It is important to remember that some estimates within an analysis may be less accurate or precise than others, and to take that into account in your interpretation of results. Estimates for groups with very low response rates (below 50%) should be noted if possible or not presented.

7.4d Suitability of PRAMS Data for the Analysis. PRAMS data may not be suitable for all uses even when they contain relevant information. Following are some types of analyses that may require special considerations, and some points to consider.

- i. Etiologic analyses.** These analyses investigate a possible causal relationship between a health event or behavior (exposure) and a health outcome (disease).
 - Such investigations require a clear temporal relationship between the exposure and the disease, with the exposure preceding the disease. For some combinations of variables, it is not possible to determine this temporal relationship.
 - Such relationships may be affected or confounded by a variety of other characteristics or behaviors. To ensure a valid analysis, information must be available on important confounders and modifiers.
 - PRAMS covers many topics, but has in-depth information on few topics. The information available from PRAMS may not be sufficiently detailed for etiologic analyses.
- ii. Program or Policy Evaluation.** Thorough program evaluation requires several pieces of information. PRAMS can usually supply some, but not all, of this information. The needed information includes:
 - Who received the intervention or service? Did the women most at risk receive the intervention? Were most of the women who received the

intervention at risk for the outcome it was meant to prevent? PRAMS data can often answer these questions if the intervention was targeted to a large proportion of pregnant women. It is not generally helpful for highly targeted interventions in small populations.

- Was the intervention administered effectively and correctly? PRAMS cannot usually provide sufficiently detailed information to answer this question.
- Did the outcome of interest change in the targeted population? In women who received the intervention? Depending on the outcome of interest, PRAMS data may be able to measure changes in the population. Changes in women who received the intervention can be measured only if PRAMS can effectively identify these women.

iii. Small Area Analysis. PRAMS is designed to provide statewide estimates and may not be suitable for analyses of very small geographic regions, such as counties or particular public health clinics. See **Section 7.4b** for information on minimum sample sizes for analysis.

7.4e Limitations of PRAMS Data. PRAMS provides valuable population-based data on an ongoing basis for numerous state uses. As with any study design, PRAMS data has some limitations and these limitations must be considered when planning analyses and presenting results.

i. Limitations in Generalizing to the Study Population (Selection Bias).

These are limitations in the ability to apply findings from PRAMS to the population of women who have had a recent live birth. They arise if women who participate in PRAMS are different from all women who had a live birth, and those differences are not completely corrected for by the weighting process. Many of these differences are identified and corrected during the weighting process using the birth certificate information on all women in the population. Uncorrected differences result in estimates that are inaccurate for the population or underrepresented subgroups.

- **Noncoverage Bias.** This bias occurs when certain groups are underrepresented in a study sample. Noncoverage bias could arise in PRAMS if birth certificate records from one area of the state or one hospital are systematically excluded from the sample because there is a regular and substantial delay in submitting records to the state health department. If noncoverage is small, bias from this source should be minimal. The percent coverage and any comments on substantial noncoverage for a given year of data can be found in the <STATE> Weights and Comments report for that year.

- **Nonresponse Bias.** Nonresponse bias occurs when some subgroups of the sample do not respond to the survey or are less likely to respond than other groups. In PRAMS, nonresponse bias among some subgroups can be assessed using data from the birth certificate, and weights are used to adjust for identified differences in response. These weights assume that the women in a particular subgroup who responded have the same response as those who did not respond. If response rates are 65% or higher, this is a reasonable assumption. However, as response rates drop below 65%, the potential for bias increases.

ii. Inability to Generalize to Related Populations. It is important to remember that PRAMS only includes in-state residents who had a live, in-state (for most states) birth. Findings from PRAMS cannot be generalized to other populations. Some specific examples are:

- If a substantial proportion of residents give birth outside of <STATE>, and these residents differ from those who give birth within <STATE>, findings from PRAMS may not be applicable to questions concerning all state residents, or to those areas of the states where many women deliver out of state.
- PRAMS findings do not represent all pregnant women. They are not applicable to women who had abortions, stillbirths, or fetal deaths.

Protocol Development Task

List any exclusions to the sampling frame specific to <STATE> and discuss any limitations on generalizability of PRAMS data due to these exclusions.

iii. Limitations of PRAMS Information (Information Bias). The PRAMS analytic data file includes questionnaire data, birth certificate data, and operations data. Each of these can have problems with missing or inaccurate data, but the causes and result of the problems may be different for each data source. When data are equally likely to be missing or inaccurate among all groups of interest, the bias is referred to as nondifferential. If some groups of interest are more likely to have missing or inaccurate data than others, the bias is referred to as differential. When interpreting your PRAMS results, you should consider whether your data may be subject to any of these biases, and if so, how the bias may have affected your results. Any problems with questionnaire data should be reported to the <STATE> PRAMS coordinator so the issue can be addressed in the next questionnaire revision.

- 1. Questionnaire Data.** PRAMS questionnaire data are self-reported and may be subject to inaccurate reporting. These inaccuracies may occur for a variety of reasons and have differing effects on estimates.

- a. Recall Bias.** Recall bias occurs because respondents asked about events in the past may not remember them accurately. The bias can be differential if a subgroup of respondents is more or less likely to recall events accurately.
 - b. Reporting Bias.** Women may be unwilling to report some behaviors or events, leading to an underestimate of the prevalence of these events, or they may over report socially desirable behaviors such as car seat use.
 - c. Mode Bias.** Women who complete a telephone interview may answer differently than they would have if they had completed a self-administered questionnaire by mail. Similarly, answers could differ if a mother chooses to answer the survey using the web. These differences may be caused by differences in trust or concerns about confidentiality, or differences in the way the question is understood when it is read by the interviewer than when it is read by the woman herself, as would be the case for both mail and web versions of the survey.
 - d. Misunderstanding the Question.** Generally, the interpretation of PRAMS questions is left up to the respondents. The questionnaire does not instruct women in the meaning of the questions. If women interpret the question differently than expected, or some subgroups interpret the question differently than others, inaccurate information may result.
- 2. Birth Certificate Data.** The medical and public health literature includes several reports on the reliability of birth certificate data. In general, infant information for current birth is usually reliable, but information on previous births may not be. Some maternal information is also questionable, e.g., cigarette and alcohol use is known to be underreported on birth certificates. Information on birth certificates can be from maternal self-report or from medical records and the source of the data may affect the results. The information may be collected differently at different hospitals within <STATE>. Some variables are missing for a large percentage of women (>10%), and some groups of women may be more likely to have missing data than others.
- 3. Operations Data.** Operations data are used most often in methodological analyses. These data are derived from the PIDS system. The operations data includes information about the timing of letters and phone calls and related contacts with mothers. If the web survey module is used, PIDS can provide information about usability and duration of the web sessions. Inaccurate information may result from data entry errors or miscoding. Data entry errors are most likely to be random and so would not bias

results. Biased results could occur if a variable is systematically entered or coded incorrectly.

- iv. Confounding.** Many factors can affect an observed association between two variables. PRAMS includes information on many of these factors, but may not include all of them. For example, core PRAMS data does not include information on drug abuse. The interpretation of an association may be incorrect if it is affected by an unmeasured confounding variable.
- v. Power.** Standard errors for PRAMS estimates are frequently larger than they would be from a simple random survey. This increase in error may lead to a lack of power for some analyses, even if the total sample is large. Lack of power is a particular concern when an event or behavior is very rare.

7.5 The State Analysis Plan

States begin identifying their data needs and how PRAMS will be used to meet those needs when they apply for PRAMS, and they continue this process during the implementation of the project. These needs form the basis of the <STATE> Analysis Plan. The <STATE> Analysis Plan is used as the guide for planning and conducting analyses of PRAMS data. The process of developing the <STATE> Analysis Plan will be helpful in developing and selecting state-specific questions, selecting appropriate stratification variables, determining sample sizes, and modifying data collection procedures. Individuals responsible for analyzing PRAMS data, in collaboration with other state PRAMS team members and programs that will use PRAMS data, should be involved in developing this plan. CDC PRAMS epidemiologists and statisticians will provide any needed technical assistance in developing the analysis plan and with subsequent analyses. <STATE> should update its analysis plan as needs and priorities change.

7.5a Developing the State Analysis Plan.

- i. Identify Data Needs.** Develop a list of data <STATE> will need in the next three to five years. Data may be needed for policy development, changes or evaluation; legislative action; needs assessment; program monitoring, modification, and development; mobilizing community support for MCH services; obtaining funding; setting priorities for programs; publications; or informing the health and lay community about particular MCH issues. The data needs identified in the background and needs section of the <STATE> application for PRAMS should provide a good starting place. Other data needs may be identified while discussing sample design and choosing the state-specific questions.

- ii. **PRAMS Data.*** Develop a list of PRAMS core and standard questions that will be useful in addressing the identified data. Needs that cannot be addressed by the current PRAMS questionnaire should be retained but shown separately and used to develop or select state-specific questions in the future. Identify information needed from the birth certificate for the analyses. List any birth certificate information needed that was not incorporated into the PRAMS data set and outline the procedures for linking that information to PRAMS. Explore other sources of information that address these data. For example, are there programmatic or vital records data sources available that could be linked with PRAMS data (e.g., Medicaid data files or infant death certificate records)?
- iii. **Time Frame.*** Note when PRAMS data are expected to be available and whether the data will be available in time to address the identified data needs. This time line should incorporate the expected time needed for the analysis and should be updated as needed.
- iv. **Priorities.*** Rank the identified data needs in order of their priority and determine what information is needed on an ongoing or periodic basis.

7.5b Components of the State Analysis Plan.

- i. **Profile of State Data Needs and Priorities.*** The profile is the summary of the developmental steps outlined above in **Section 7.5a**. The profile is used to create the listing of analyses for the state.
- ii. **Listing of Analyses.*** This listing presents a brief description of each analysis; the rationale for the analysis, including its importance and the data needs it addresses; the analytic design and methods; and a plan for disseminating the results. For further discussion of dissemination of findings, see **Section 7.7**.
- iii. **Policy for Review of Reports, Manuscripts, and Presentations.*** Describe the review process for PRAMS reports, manuscripts, and presentations in <STATE>, including the line of review for publications with internal and external authors, and for publications or presentations that include PRAMS data provided through a data request.
- iv. **Policy for Access to Data by External Researchers.*** Establish guidelines for access to PRAMS by individuals who are not members of the <STATE> PRAMS team (e.g., individuals who work in the state health department, other state health employees, contracted data analysts, university researchers, graduate students). Describe how such studies will be developed, reviewed, analyzed, reported, authored, used and disseminated. Examples of guidelines for PRAMS data use by external researchers and data sharing agreements may be found in **Appendix R (Guidelines for Collaborative Studies Using PRAMS Data)**.

- v. **Policy on Authorship.** Develop a policy for authorship in authored publications. An individual may be identified as an author or may be recognized in the acknowledgment section. **See Chapter 8 (Authorship)** for further details.

Protocol Development Task

Identify the individual who will have overall responsibility for analytic activities in your state. Identify others who will be collaborators or conduct their own analyses.

Develop your State Analysis Plan and place the plan in Appendix S. Your State Analysis Plan should include: (1) the profile of state data needs and priorities; (2) the listing of analyses; (3) the policy for review of reports, manuscripts, and presentations; (4) the policy for access to data by external researchers; and (5) the policy on authorship.

Updates to the State Analysis Plan should be shared with CDC and also placed in **Appendix S**.

7.6 Conducting Analyses

7.6a Before Beginning. Review all documents listed in **Section 7.2a**. Note any potential concerns for analysis, including groups with low response rates, questions with high rates of missing data, or other issues that may impact the analyses listed in the analysis plan. Run frequencies of birth certificate variables, assess missing data and note any problems that will impact the analyses in the analysis plan. Update the analysis plan with any problems noted.

7.6b Conducting Specific Analyses. When the weighted analysis data set has been provided to <STATE> and state analysts have become familiar with it, specific analyses can be undertaken.

- i. Determine the complexity of the analysis. Some data needs will be addressed by very simple analysis, such as prevalence estimates. More complex analyses may need an analytic proposal that states the purpose and goals of the analysis, the expected use of the findings, and the projected time frame for the analysis. A written proposal can assist in identifying collaborators in the analysis, other data sources that might add to the PRAMS data, and potential audiences for the findings.
- ii. Determine the variables needed for this particular analysis.
- iii. Create a subset of the master analysis data set that contains the selected variables. For further information, see the **Guidelines for State Analysis of PRAMS Data (Chapters 5-9)**.

- iv.* Recode variables as necessary.
- v.* In SAS, screen and assess accuracy of data:
 - a.* Note problems with the data or question identified in the review of the weighting documentation documents.
 - b.* Obtain frequencies of all categorical variables (birth certificate, questionnaire, operations, and required SUDAAN variables).
 - c.* Determine how missing values will be handled.
 - d.* Check adequacy of sample sizes of the stratum or domain being examined.
 - e.* Review associated comments and open-ended responses for PRAMS questions being analyzed. Recode values if needed.
- vi.* Perform analyses.

7.6c Qualitative Analysis. CDC provides the COMMENT data set to accompany the yearly weighted data set. This data set is available for qualitative analysis, and the comments may be linked to mothers' responses in the analysis data set. Because the comment data were not solicited in a standardized way, care should be taken in generalizing the findings. Also, comment data do not generally provide the in-depth description necessary for rigorous qualitative analysis.

7.7 Using and Disseminating Data from PRAMS

One of the primary goals of PRAMS is to provide information that can be used to improve understanding of health behaviors and to prevent adverse health events. The analysis plan is vital to this process as PRAMS data must be analyzed, published in some format, and distributed in order for it to be used.

7.7a Ways to Use PRAMS Data. PRAMS data can be used in many ways to improve maternal and child health. The following list gives some examples:

- Develop, modify, and evaluate public health programs and policies.
- Monitor progress toward the year 2020 health objectives or other state and local health goals.
- Support legislative proposals.
- Support needs assessment for Title V Block Grant.
- Support funding requests.

- Inform the health community about statewide prevalences of maternal behaviors.
- Determine patterns of using public health programs and clinical services.
- Educate women about healthy behaviors.

For detailed examples of how states have used their PRAMS data, see the **CDC PRAMS Web site** (<http://www.cdc.gov/prams/dta-successstories.html>)

7.7b Ways to Disseminate PRAMS Data. For PRAMS data to be used effectively, they must be provided to people in a form that those people can use. Early in the development of PRAMS, the <STATE> PRAMS team should identify the individuals and groups who should receive PRAMS data, the topics of interest to these groups, and the most effective media types and formats for reaching those groups. Some examples of potential audiences, media, and presentation formats are listed below:

<i>Audience</i>	<i>Media</i>	<i>Format</i>
General public	Press, television, radio, Internet	Media and press releases, short reports and graphs
Public health professionals	Conferences, journals, Internet	Presentations, posters, articles, reports, newsletters, fact sheets, brochures
Health care professionals	Conferences, journals, Internet	Presentations, posters, articles, reports, newsletters, fact sheets, brochures
Legislature	Short reports, briefings	Short reports and graphs, fact sheets, brochures

Some suggestions in preparing abstracts and manuscripts are provided in **Appendix T**.

7.7c Development of a Mailing List. Early in the project, the state should develop one or more lists of people who should receive information about and from PRAMS. It may be useful to develop different mailing lists based on the topic or purpose of the information. The list should include data users, program planners and policy developers, health care providers, university faculty, state provider associations, the PRAMS Steering Committee, and other interested parties and it should be expanded as necessary. Your CDC program manager should be placed on the list and CDC should be sent copies of presentations and

publications by state PRAMS teams or external researchers using <STATE> PRAMS data.

The first use of the mailing list will be to introduce the new PRAMS project in <STATE> to the health community, including its methods, data sources, and potential uses of findings.