

## Step 4: Monitor Survey Administration and Prepare for Data Analysis

Four steps are involved in ensuring the proper administration of the PHDS-PLUS and preparing your data files for analysis:

- 4.1 Monitor survey administration
- 4.2 Clean and code your PHDS-PLUS data set
- 4.3 Integrate updated child enrollment and utilization information
- 4.4 Weight your data set to represent your target population

Guidelines for each of these steps are provided below.

### STEP 4.1: Monitor Survey Administration



#### WHAT IS THE PURPOSE OF THIS STEP?

The purpose of this step is to closely track the quality of your survey administration process and to monitor and troubleshoot any problems with securing an optimal response rate and accurate data entry.

#### In this step you will:

- Receive and review weekly tracking reports from your survey vendor
- Assess and troubleshoot problems with response rates



#### GUIDELINES AND ISSUES TO CONSIDER

- Weekly survey tracking report

Survey responses should be tracked throughout the administration process. At a minimum, a variable indicating the survey disposition should be maintained by your survey vendor and reported on a weekly basis throughout the administration process. Values for this variable can include:

- *Interview conducted* – Parent was able to be contacted and the survey was administered
- *Complete* – At least 80% of the survey items for which all parents should answer were completed

- *Incomplete* – Less than 80% of the survey items were completed
- *Data error/bad contact info* – Bad addresses, disconnected phone, fax/computer line, etc.
- *Ineligible* – One or more of the eligibility criteria were not met (e.g., age >48 months or parent reports that child is not enrolled in the health plan)
- *Deceased* – Child has passed away
- *Language barrier* – Respondent speaks a language different from the survey administrator
- *Refused* - Parent refused to complete survey
- *No answer* - No answer/answering machine/screened by caller-ID

Calculate interim response rates and troubleshoot response problems

Your *raw response rate* is simply the number of questionnaires returned divided by the number of children in your starting sample. However, this rate is often misleading and not representative of the true response rate. Therefore, an *adjusted response rate* should be calculated on an interim basis throughout survey administration to track whether you are meeting response rate targets.

For the adjusted response rate, both the numerator (number of respondents) and the denominator (number sampled) are adjusted based on certain factors:

- First, the numerator (number of respondents) is adjusted from the total number of questionnaires returned to the number of complete questionnaires returned. This means that only surveys with at least 80% of the items for which all parents should have answered should be counted in the numerator of the response rate. Therefore, you should not include items that should have been appropriately skipped in your count of the number of items the parent answered. You can adjust this algorithm if you want to allow responses to be considered if less than 80% of items are completed.
- Second, the denominator (number sampled) is adjusted to include only those children who are eligible for the survey and for whom the telephone number you have is valid and therefore the parent could have been contacted to complete the survey. Thus, the denominator is adjusted by those who are ineligible, deceased, or have an unusable phone number in the data file. The formula to adjust your response rate is:

$\text{Response Rate} = \frac{\text{Number of 80\% or more completed surveys}}{\text{Number of children sampled} - (\text{deceased} + \text{ineligible} + \text{unusable phone number})}$
---

If your response rate falls short of projections you may need to troubleshoot with your survey vendor and consider adding and/or altering administration steps. For instance, you may decide to add an additional phone call, send a postcard, alter your introductory message left on voicemail systems, or introduce a small incentive for participation in the survey.

## **STEP 4.2: Clean and code your PHDS-PLUS data set**



### **WHAT IS THE PURPOSE OF THIS STEP?**

The purpose of this step is to obtain an accurate data set from your survey vendor, and this includes consistent and accurate cleaning and coding of the data set in preparation for constructing PHDS-PLUS quality measures and the analytic variables needed to report your PHDS-PLUS project findings.

#### **In this step you will:**

- Obtain and check interim data sets for accurate data labeling and entry
- Prepare data files for analysis



### **GUIDELINES AND ISSUES TO CONSIDER**

#### Obtain and check interim data sets for accuracy of data entry and survey administration

Your survey vendor should provide you with a test and interim data sets according to a predetermined schedule. For example, you should receive a data set once the first 100 survey responses have been received and entered into a database by your survey vendor.

You should ask your vendor to send a test data set that is based on a handful of “mock interviews” that they conduct. This test data set will test the CATI and the help you to check the data variables created by

#### **Quick Data Quality Tips:**

- Always be sure data variables are labeled
- Make sure your data dictionary is updated for any changes made to data labels or response codes. Good documentation is essential!
- Create a backup of your data set in case of emergency. Also, create temporary and permanent data sets wisely. Think about what you would need to do if you lost the data.
- Always keep a copy of your original data set.

the CATI. When you receive the test data set you should make sure that your vendor is using the data variable labels agreed upon and that responses to survey items are also assigned the agreed upon values (e.g., 1= "no"; 2 = "yes"). If errors are found, request that they be corrected immediately. In addition, you should examine whether any survey items appear to include an unusual number of missing values (more than 2% of respondents) or if survey skip patterns are not being followed (e.g., responses are recorded for individuals who should not have been asked a question because they did not qualify for that item).

Missing values and erroneous skip pattern data entries indicate problems with CATI and should be noted and addressed before the administration process begins. Corrections may involve conducting follow-up spot training with survey interviewers or adjusting the CATI set-up.

### Preparing the data files for analysis

Valid PHDS-PLUS findings require careful preparation of your data prior to starting your analysis. The following are necessary steps to prepare the data for analysis. They do not necessarily need to be conducted in the order presented.

#### **Data Prep Step #1:** Verify survey completeness

You should receive the data from the vendor for all of the interviews that they have conducted. However, for your analysis you should limit the data to include only surveys with at least 80% of the items completed.

#### **Data Prep Step #2:** Check for ineligible cases

Make sure parents who responded have children who meet the sampling criteria for age and continuous enrollment. (a) Run a frequency on the age variable from the survey responses. Compare responses of cases where parent-report indicates their child is outside the parameters of 48 months against the administrative data used to generate the sampling frame. Use the parent report as the "gold standard" and exclude cases in which the child was erroneously included in the sampling frame. (b) Remove records where the child was found not to be in the health plan, provider, or unit from which you are sampling.

### **Data Prep Step #3:** Check for duplicate data records

Make sure every record has a unique identifier.

### **Data Prep Step #4:** Check for out-of-range values

Run frequencies on all of your variables to check for out-of-range values or odd-looking distributions. For example, one item asks respondents to rate their health care on a scale from 0–10. In most cases, the distribution should cluster around seven, eight, or nine. If a large number of responses were “1” then you may have a data entry error where some of those responses should have been coded as 10. At this point, you may not be able to go back and correct the data error. If the error is random and affects only a few cases then you may want to exclude those cases. However, if the error seems to be systematic and impacts a large number of responses, it may be worth it to find the source of the error and correct it.

### **Data Step #5:** Identify problems with skip patterns

Run frequencies and cross-tabulations to verify that skip patterns were followed correctly. If errors seem to be random and affect only a small number of records (less than 2%), assume the item stem (question instructing the respondent to go to a different question) is accurate and then correct the response for the incorrect skip. Systematic errors or problems with a significant number of cases should be verified.

If you want to be absolutely certain that skip patterns were followed, you can make it a requisite that only the children of parents who responded appropriately to the filter question are included when you create the new variables.

### **Data Prep Step #6:** Assign missing values

Missing values should be recoded in some way so that you know not to include them in the analyses. If you recode missing responses to “0” be very careful not to include them in your analyses as this will affect your results. Designate missing values in the data set to ensure that they are omitted when calculating measures. Also, recode the response options of “refused” (primarily for telephone administration) to be a missing value. Examine the number of “I don’t know” responses that you get. If this total percentage is less than 2%, then you should recode them to be missing values.

## **STEP 4.3: Integrate updated child enrollment and utilization information**

### **WHAT IS THE PURPOSE OF THIS STEP?**

The purpose of this step is to make sure that data for each child are as current as possible at the time that you calculate your PHDS-PLUS quality indicators and create reports of your findings to key audiences.

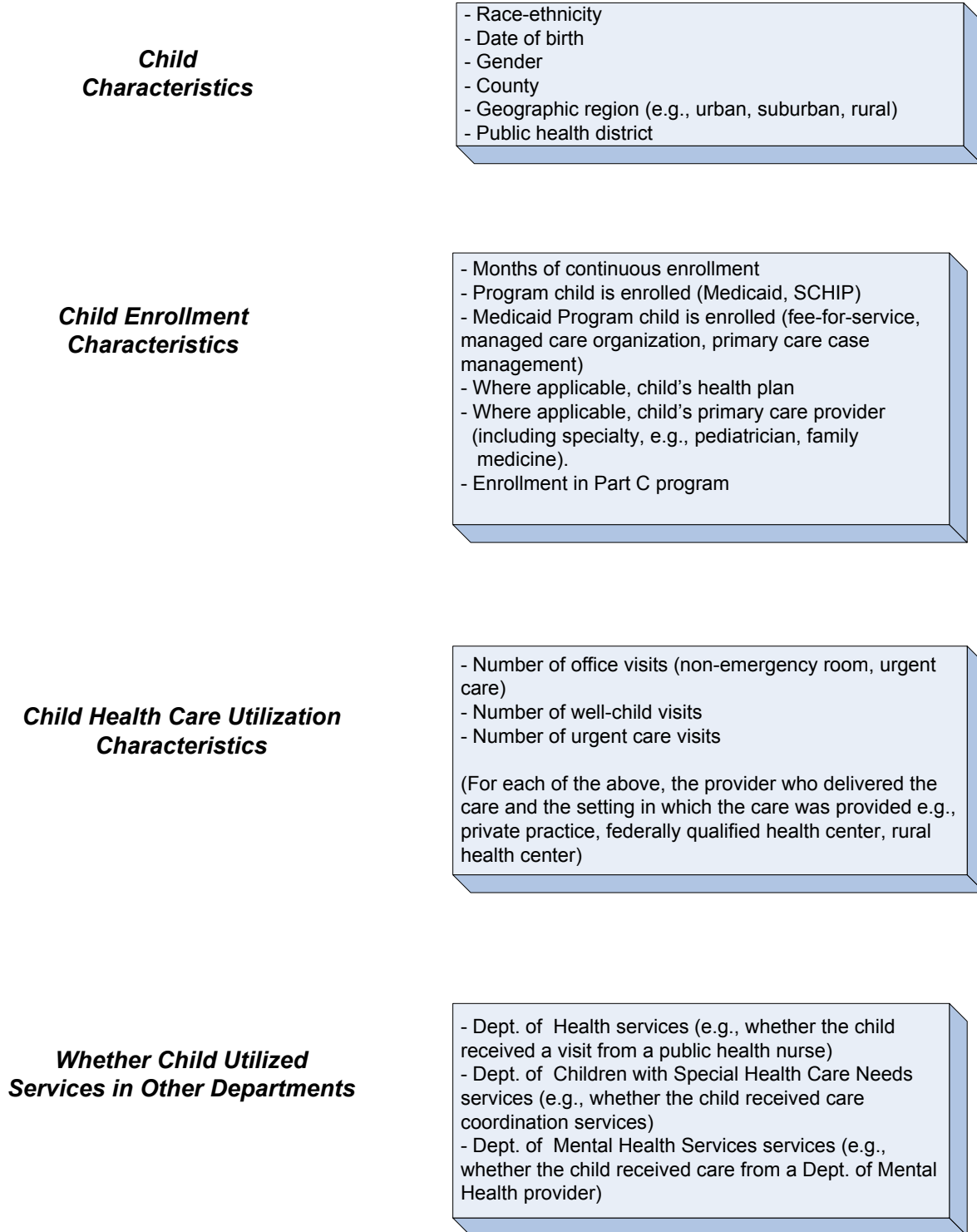
#### **In this step you will:**

- Identify variables that may need to be updated for each sampled child
- Obtain and merge updated data variables into your analytic variable data set

### **GUIDELINES AND ISSUES TO CONSIDER**

As discussed in Steps 2.5 and 3.2, many variables that are derived from enrollment and utilization data sets are valuable for purposes of analyzing and reporting PHDS-PLUS findings. Examples of these variables are listed in Figure 4.2 below and include information on utilization of services, program enrollment, geographic location, and so on. Some of these variables are expected to change over time. For each person in the starting sample, it is important to update these variables to incorporate new data entries or changes that may have occurred between the time that the initial PHDS-PLUS sample was pulled and the time that the survey was actually completed by parents of children included in your sample.

**Figure 4.1: Examples of Data Elements to Collect at the Time of Sampling**



## **STEP 4.4: Weight your data set to represent your target population**

### **WHAT IS THE PURPOSE OF THIS STEP?**

The purpose of this step is to be sure that your final data set is representative of the desired sample of children you would like to report at the end of your project. Weighting is a standard component of most survey projects, especially if you oversampled for certain population subgroups (e.g., racial/ethnic groups) or there are response biases (e.g., some populations subgroups are systematically less likely to respond to your survey).

#### **In this step you will:**

- ☑ Clarify your desired population of children to be sure your findings are representative and decide whether weighting your data is required
- ☑ Develop a weighting methodology and construct sampling weights to use during your data analysis



### **GUIDELINES AND ISSUES TO CONSIDER**

#### Determining whether weighting is required

The sampling and survey administration strategy specified throughout this manual is designed in the following manner: (1) To obtain sufficient samples of completed surveys for the three age groups that correspond to the age-specific sections of the PHDS-PLUS survey (3–9 months, 10–18 months, 19–48 months old) and (2) to target groups of particular interest in your state.

You will need to weight your data if:

1. You collect both a core and state-specific sample as suggested AND
2. You want to be able to combine the data from your core and state-specific samples in order to create reports that describe PHDS-PLUS findings across your state.



## Developing a weighting methodology

Your survey vendor should be able to construct valid weights once you clarify the population of children about which you would like to report PHDS-PLUS findings (e.g., all children enrolled in Medicaid). Below are issues and tips to consider as your weighting method is developed.

### 1. Clarify the Baseline Population to be Used in Calculating the Weights

Since the children eligible to be sampled for the PHDS-PLUS were ONLY those who had a phone number in the data file and who met the continuous enrollment requirement, data will ONLY be able to be weighted to represent this group of children. Therefore, the baseline populations to which you may refer for purposes of sampling is the population of continuously enrolled children in Medicaid who are 3–48 months old and who were eligible to be sampled for the PHDS-PLUS survey.

### 2. Specify the Core Factors to be Used in Calculating Weights

At a minimum, you will need to weight your data according to the child's age. This is because your PHDS-PLUS samples were initially stratified by age to ensure sufficient samples of children in the following age groups were obtained: 3–9 months (25% of starting/completed sample), 10–18 months 25% of starting/completed sample, 19–48 months (50% of the starting/completed sample). To make your PHDS-PLUS findings representative, you will need to determine the actual age distribution of the population of continuously enrolled children in Medicaid who are 3–48 months old and who were eligible to be sampled for the PHDS-PLUS survey according to these groups and weight the data accordingly. Other variables you may need to include in your weighting methodology are race, geographic location, type of health care provider or health plan, or any other variables for which you oversampled or are under- or over-represented in your completed sample due to a known response bias. An example weighting method is provided below in example 4.1.

## Example 4.1: Weighting from the Field

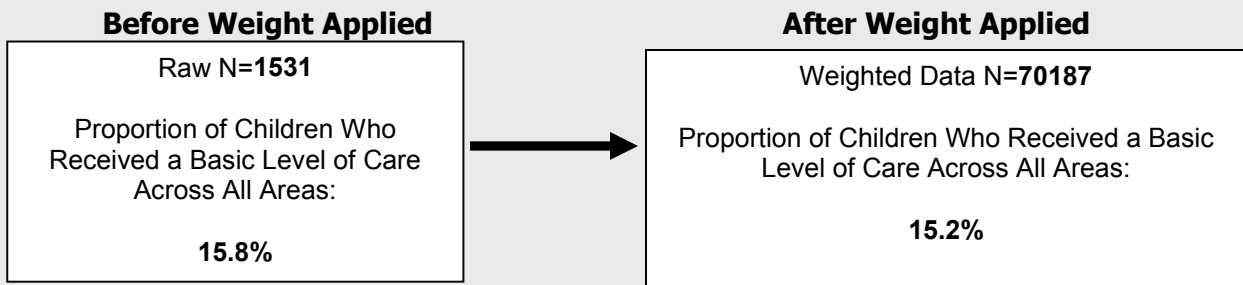
State X used the core and state-specific sampling strategy recommended by the CAHMI. For their state-specific sample, they decided to ONLY include children who had one or more well-child visits in the last 12 months.

Factors the state used in weighting their sample:

- ✓ Proportion of children in the three age groups who had a well-child visit in the last 12 months.
  - Group 1. Proportion of children 3–9 months old who had a well-child visit
  - Group 2. Proportion of children 10–18 months old who had a well-child visit
  - Group 3. Proportion of children 19–48 months old who had a well-child visit
- ✓ Proportion of children in the three age groups who did NOT have a well-child visit in the last 12 months.
  - Group 4. Proportion of children 3–9 months old who DID NOT have a well-child visit
  - Group 5. Proportion of children 10–18 months old who DID NOT have a well-child visit
  - Group 6. Proportion of children 19–48 months old who DID NOT have a well-child visit

Weighting Process Utilized

- ✓ The population of continuously enrolled children in Medicaid who were 3–48 months old and who were eligible to be sampled for the PHDS-PLUS survey was analyzed.
- ✓ The proportion of children in each of the six groups noted above was determined.
- ✓ A weight was created based on the figures obtained.
- ✓ Each child whose parents responded to the survey was categorized in one of these six age groups and a weight was assigned to each.





## ADDITIONAL TIPS AND RESOURCES

### TIPS FOR INCREASING YOUR RESPONSE RATE

- **Use letterhead for all correspondence.** This establishes trust with the respondent and adds credibility to the project.
- **Personalize the letters.** People are more likely to notice a letter addressed to them rather than a general greeting.
- **Use an express shipping service,** such as FedEx or UPS, for the pre-notification letter. Research has shown that people are more likely to respond to express mail than regular mail.
- **Provide an explanation** for how the survey will be used to improve health care for children.
- **Call at family-friendly times.** It is important to realize that the families you are calling have young children. Therefore, the call times should be respectful of the schedules of young children.
- **Utilize a survey vendor that has experience surveying parents of young children.** Whenever possible, contract with a survey vendor who has experience conducting surveys with parents of young children.
- **Provide a means to receive findings.** Let parents know how to get a copy of the findings of the study after it is complete.

### TIPS FOR PREPARING YOUR DATA FILES FOR ANALYSIS

While data preparation is often considered part of the analysis, this task is included in the data collection section because it is often completed by the vendor while administering the survey. Additional data preparations and cleaning are also included in the next section.

If you are using a vendor to administer the survey, the vendor should submit to you a data file that contains the following:

1. Coded responses for all items, including blank, do not know, refused to answer, and item skipped
2. Survey disposition, such as if the survey was completed, the reasons for incomplete surveys (see list above in *Step 4.1: Monitor Survey Administration*)
3. Other variables linked when the sampling frame was generated
4. Administrative data used for generating the sampling frame
5. Age of child in months
6. Months of continuous enrollment in the last 12 months
7. Any supplemental data linked prior to the removal of identifying information used for survey administration

## 8. A data dictionary for the file

Data entry or the coding process should be verified. You may also choose to audit the process by comparing the coded values against the original survey.

You also may choose to have your vendor conduct some initial data preparations, such as:

1. Verifying ineligible cases
2. Checking for duplicate data records
3. Running frequencies on all variables to check for values that are out of range
4. Identifying problems with skip patterns

You may want to have the vendor verify these with the original surveys to ensure that errors stem from the respondent and not from the administration process. Once these are identified, you will need to make decisions on how they will be handled for the analysis. Refer to the next section for more detail on analyzing the results.