

# Preparing and Analyzing Data From the CAHPS Clinician & Group Surveys

<b>Purpose of this Document</b> .....	<b>1</b>
<b>What the Analysis Program Does</b> .....	<b>1</b>
<b>Preparing Data for Analysis</b> .....	<b>1</b>
Task 1: Identify and Exclude Ineligible Cases.....	3
Task 2: Code and Enter the Data.....	3
Task 3: Clean the Data.....	5
Task 4: Conduct an Audit.....	6
<b>Example of a CAHPS Macro Call</b> .....	<b>6</b>

## Purpose of this Document

This document explains how to transform the raw data from the CAHPS Clinician & Group Surveys into data that the CAHPS Analysis Program can use, and how to use that program to analyze data from the Clinician & Group Surveys. Comprehensive information about using the Analysis Program for any of the CAHPS surveys is provided in *Instructions for Analyzing Data from CAHPS Surveys* at [https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Prep\\_Analyze/2015\\_Instructions\\_for\\_Analyzing\\_Data.pdf](https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Prep_Analyze/2015_Instructions_for_Analyzing_Data.pdf). Please read both documents carefully so that you are clear on how to use the macro for this survey.

## What the Analysis Program Does

The CAHPS Analysis Program generates output that you can use to calculate the performance of various entities—including medical groups, individual physicians, health plans, hospitals, counties, States, regions, or any other appropriate units of measure—to show how one entity's performance compares to the overall performance of all other entities. The program generates both text output and SAS data sets which provide information on the scores for global ratings, composite measures, and individual items.

## Preparing Data for Analysis<sup>1</sup>

Before conducting an analysis, you must carry out several tasks to prepare the data received from completed questionnaires. For information on determining a completed questionnaire, refer to the appendix in *Fielding the CAHPS Clinician & Group Surveys* at [https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Admin\\_Survey/1033\\_CG\\_Fielding\\_the\\_Survey.pdf](https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Admin_Survey/1033_CG_Fielding_the_Survey.pdf).

- Task 1: Identify and exclude ineligible cases.
- Task 2: Code and enter the data.
- Task 3: Clean the data.
- Task 4: Conduct an audit.

---

<sup>1</sup> This guidance on preparing data is also provided in the *Instructions for Analyzing Data from CAHPS Surveys*.

Many interim files will be created along the way. Before beginning this process, you must take steps to preserve the original data file created when the raw survey responses were entered. Any changes and corrections made during the cleaning and data preparation phase should be made on duplicate files. There are three reasons for this action:

1. The original data file is an important component of the complete record of the project.
2. Having an original file will allow you to correct data errors that were made during the cleaning process.
3. The existence of an original file is critical if the vendor or sponsor wants to go back later and conduct other analyses or tests, such as extent of error tests or tests of skip patterns.

### **Data File Specifications**

The data file contains the raw data from responses to the Clinician & Group Surveys. The short item handles of the items in each questionnaire can be found in the tables provided in *CAHPS Clinician & Group Surveys: Overview of the Questionnaires* at [https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Get\\_Surveys/1350\\_Overview\\_of\\_Questionnaires.pdf](https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Get_Surveys/1350_Overview_of_Questionnaires.pdf).

The responses to each question must use the code numbers, or precodes, contained in the questionnaires.

Users should construct a separate data file for each version of the survey. Do not include data from different survey instruments in the same data file. For example, do not include responses to the Adult 12-Month Survey and Child 12-Month Survey in the same data file.

If you are interested in submitting your data to the CAHPS Database, please refer to the data file specifications for submission at <http://www.cahps.ahrq.gov/CAHPS-Database/Submitting-Data/CG-Data.aspx>.

The number and scope of the data preparation tasks and the way they are carried out depend on the data collection protocol and the way in which the data were recorded. For example—

- If the vendor/sponsor collected data with a self-administered mailed questionnaire, did respondents record answers on optical scan forms<sup>2</sup> or record them directly on the CAHPS-formatted questionnaires?
- If the vendor/sponsor collected data through telephone interviews, did the interviewer use computer-assisted telephone interviewing (CATI) or paper-and-pencil forms?

### Task 1: Identify and Exclude Ineligible Cases

Several situations render a case ineligible for analysis. One common scenario that vendors must be prepared to handle occurs when the respondent reports that he or she has not visited the sampled physician or medical group. This might be indicated by a “no” response to Question 1 (“Our records show that you got care from the provider named below in the last 12 months. Is that[ right?]”).

Other questionnaires may be considered incomplete and excluded from analysis but are not excluded from the denominator used to calculate the response rate. For example:

- If someone else assisted the respondent or answered the questions (as a proxy) or
- If at least half of the key items on the questionnaire were not filled in. (For a list of key items, refer to Appendix B of *Fielding the CAHPS Clinician & Group Surveys* at [https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Admin\\_Survey/1033\\_CG\\_Fielding\\_the\\_Survey.pdf](https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Admin_Survey/1033_CG_Fielding_the_Survey.pdf).)

### Task 2: Code and Enter the Data

There are a variety of possible methods that can be used to enter data from CAHPS surveys. The exact level of coding required will depend on the method used to capture the data (e.g., questionnaires that require data entry versus questionnaires that are scanned by a computer). Your coding specialist should review each questionnaire to see whether the responses are legible and whether any responses need to be coded. Each item should have a corresponding code, even items that were not answered. The table below shows examples of recommended coding. If you are interested in submitting your data to the CAHPS Database, follow the data file and coding

---

<sup>2</sup> Optical scan forms are answer sheets in which respondents fill in the circle that corresponds to their answer choice. These forms are fed through an optical scanning machine, and the data are automatically captured by a computer. Standardized tests for students, such as the SAT, generally use optical scan forms.

specifications available at <https://www.cahps.ahrq.gov/CAHPS-Database/Submitting-Data/CG-Data.aspx>.

After coding is completed, enter the data into a computer file.

Response	Code
Actual response	Use the corresponding survey precode indicated beside the response option. Precodes typically begin with 1 and number consecutively for each response option.
Item was appropriately skipped	Code as 7 (or 77 if precodes exceed 7)
Item showed more than one response option completed when only one is appropriate (i.e., multiple marks)	Code as 8 (or 88 if precodes exceed 7)
Item was left blank and should not have been skipped	Code as 9 (or 99 if precodes exceed 7)

If you use **optical scan forms** for your mailed questionnaires, the scanning equipment automatically enters the data into a computer-readable file. If you do not use optical scan forms, the mail questionnaires are designed for direct data entry without the need for coding most respondent answers. However, if it is unclear which answer the respondent selected (e.g., the respondent's pencil mark does not neatly fit within a single answer category, or two responses are marked), then your coding specialist will have to make a decision about which response the respondent intended. If it is not readily apparent what the respondent intended, the coding specialist should indicate that the answer be entered as missing or multiple marks, as appropriate.

If you use a **CATI system** for a telephone survey, data are entered directly into a data file that has already been programmed to refuse unlikely and invalid responses. If you use paper questionnaires to record answers given in a telephone interview instead, the process for coding and data entry is the same as for the standard paper version of the mailed questionnaires.

To ensure quality, answers from paper-and-pencil questionnaires should be key entered by two separate data entry specialists. The results from the two should be compared to identify and correct data entry errors. At the end of the coding and data entry process, you will have an electronic data set of responses to all the questionnaire items.

Additional coding and recoding may be necessary prior to using the CAHPS Analysis Program. Refer to the SAS Data Set Requirements in the *Instructions for Analyzing Data from CAHPS Surveys* at [https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Prep\\_Analyze/2015\\_Instructions\\_for\\_Analyzing\\_Data.pdf](https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Prep_Analyze/2015_Instructions_for_Analyzing_Data.pdf) for recoding variables for use with the CAHPS Analysis Program.

### Task 3: Clean the Data

In many cases, the data set you have created will have imperfections. You will have to take several steps to fix these imperfections before any results are reported.

**Check for out-of-range values.** Out-of-range responses occur when respondents provide inappropriate responses for a particular question. For example, if the valid response choices for a question are 0 or 1, a value of 2 would be out of range. Similarly, if a respondent circled two categories when he or she was supposed to provide only one answer, the response is out of the acceptable range of the question.

To detect out-of-range values, you need to review question frequencies. This can be done either by visually scanning a report showing the item distributions or frequencies or by running the questionnaire data through a computer program. Both are often used to improve the quality of the data. If a value is found that is impossible (or unlikely) given the response options, the questionnaire should be reviewed and revisions made to the data. These revisions often involve setting the out-of-range values to “missing,” which drops them from the data analysis for that particular question. Carefully document the results from this review process, including any changes to the data set. Maintain an audit trail (electronically and on paper) so it is possible to go back to the original data file.

**Check for skip pattern problems.** Response inconsistencies generally arise when a respondent misunderstands a question or does not successfully follow instructions to skip questions. An example of a response inconsistency would be if a respondent answered that he or she had no doctor visits in the past 6 months, but then answered follow-up questions about visits in the past 6 months. If there are inconsistencies between the response to the screener question and the following question’s response, assume the screener response is correct.

**Check again for ineligible cases.** Identify any questionnaires that are not eligible for analysis and remove them from the data set used for the CAHPS SAS program. Questionnaires in which fewer than half of key items are answered should not be counted as completed surveys. (For a list of the key items in the survey, refer to Appendix B of *Fielding the CAHPS Clinician & Group Surveys* at [https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Admin\\_Survey/1033\\_CG\\_Fielding\\_the\\_Survey.pdf](https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Admin_Survey/1033_CG_Fielding_the_Survey.pdf).)

**Check for duplicates.** The number of records in the data file should match the number of completes and partials in the sample file. Duplicates can occur if the

vendor conducts a follow-up phone interview, if the mail questionnaire arrives at the same time or soon after and the case slips through the receipt control system, or if there are errors in data entry. Your policy should be to keep the first questionnaire that comes in.

#### **Task 4: Conduct an Audit**

Whether surveys are collected in standard paper format, as optically scanned forms, or as paper telephone questionnaires, a small random sample of the entered data should be audited by comparing hard-copy forms with the results of data entry. This enables you to catch any systematic errors. For example, if the optical scanning program was incorrect, Question 5 may be entered in the file where Question 6 was supposed to be. These types of systematic errors will show up consistently across all questionnaires.

#### **Example of a CAHPS Macro Call**

The CAHPS Analysis Program is designed to analyze both individual items and composites of several items that use the same response scale. Below is an example of a SAS data step and macro parameters (also referred to as macro calls) designed to prepare the raw data for analysis by the CAHPS macro. The item numbers in the example refer to questions in the Adult 12-Month Survey.

Once the data step is run and a new analysis data set is created, you can run multiple CAHPS macro calls using any combination of single items or composites based on recodings done in the data step.

This example shows that the CAHPS macro calls for the “Patients’ Rating of Provider” and the “Helpful, Courteous, and Respectful Office Staff” measures. Both measures are case-mix adjusted for age, education, and general health rating. The parameter for smoothing variances is set equal to 25 for the rating measure and 20 for the composite measure based on CAHPS Consortium recommendations. The minimum target sample size for items in the composite measure is set to 10—meaning that items in the composite that have 10 or fewer responses will be downweighted when analyzing the composite. These new parameters are especially relevant for analyzing data from Clinician & Group Surveys.

```
data runfinal( drop = i ) ;
  set in.test ;
```

```
* Recodes numeric site variables to character to simplify
  interpretation of the result tables. The CAHPS macro
  requires the variable PLAN be created. PLAN may be
  any entity, such as a hospital, clinic, state, gender, or
  whatever unit needs to be compared. The text output and
  variable names will still use the label PLAN, but you
  will know exactly what it is from this data step. ;
```

```

length plan $ 20 ;
if      region = 1 then plan = 'Clinic 1' ;
else if region = 2 then plan = 'Clinic 2' ;
else if region = 3 then plan = 'Clinic 3' ;
else
      plan = 'Clinic Other' ;

ghr = q26 ;
age = q28 ;
edu = q30 ;

run ;

* Run measures through the CAHPS macro
* for a single rating ;
%cahps(var      = q23,
      vartype   = 2,
      name      = Rating of Provider,
      adjuster  = ghr age edu,
      adultkid  = 3,
      smoothing = 25;
      dataset   = runfinal,
      outname   = rateprov ) ;

* Run measures through the CAHPS macro
* for a composite ;
%cahps(var      = q24 q25,
      vartype   = 3,
      name      = Office Staff,
      adjuster  = ghr age edu,
      adultkid  = 3,
      K         = 10,
      smoothing = 20;
      dataset   = runfinal,
      outname   = ofcstaff ) ;

```

Although not needed for the examples shown above, sometimes variables need to be recoded for use in the CAHPS Analysis Program. Below are example programming steps that can be used to recode variables prior to running the CAHPS macro calls. Recoding Yes/No response options is necessary if you are using the Visit Survey, and reverse coding is necessary for some supplemental items. Additional examples are listed in the *Instructions for Analyzing Data from CAHPS Surveys* at [https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Prep\\_Analyze/2015\\_Instructions\\_for\\_Analyzing\\_Data.pdf](https://www.cahps.ahrq.gov/~media/Files/SurveyDocuments/CG/Prep_Analyze/2015_Instructions_for_Analyzing_Data.pdf).

```

* Recode Yes/No response options from 1-2 to 0-1;
* Example item numbers from the Visit Survey;
array yn q15 q24;
do i = 1 to dim ( yn ) ;
  if yn [i] = 0 then yn [i] = . ;
  else if yn [i] = 2 then yn [i] = 0 ;
end ;

* Reverse code a how often response scale;
* Example item is CU1 from the supplemental items;
array rev q51 q54 ;
do i = 1 to dim ( rev ) ;
  if rev [i] in (1, 2, 3, 4) then rev [i] = 5 - rev [i] ;
  else
      rev [i] = . ;
end ;

```