

## Description of SEER-CAHPS Data Files

### 1) Cohort Files

#### A) **PEDSF** - Patient Entitlement and Diagnosis Summary File

Combines the EDB information and the SEER diagnosis information for the patient. We had 3,970,491 total people. 205,419 (5.2%) have PEDSF records. We removed people who died more than 90 days before the date received on any survey (72 in 2013; 8 in 2014). The final SEER-CAHPS PEDSF has 205,339 records.

The SEER-Medicare group constructs this file as part of their normal linkage processes.

We used the HIC to SEER crosswalk and BIC Equate file to select the records for our respondents. We dropped some fields we would not be using, especially fields with protected variables. The current file has data for diagnosis years 1973-2011.

**The PEDSF file contains one record per person for individuals in the SEER Program data base who have been matched with Medicare enrollment records and have completed a CAHPS survey in 1997-2013. Of persons who were reported by the SEER Registries to have been diagnosed with cancer at age 65 or older, 94 percent were matched with Medicare enrollment records. For persons appearing in the PEDSF file, basic SEER Program diagnostic information is available for up to 10 diagnosed cancer occurrences. Data also include Medicare entitlement and utilization data from 1996-2013.**

**NOTE:** In PEDSF, the cancers are stored in order of Sequence Number. Sequence numbers 00-59 are used for behaviors 2 and 3, while sequence numbers 60-87 are used for behaviors 0 and 1. Within each sequence number set, the cancers are in diagnosis date order, but overall the cancers MAY NOT be in diagnosis date order. This affects only 196 of the patients in PEDSF.

#### B) **SumDenom** - Medicare Surveillance Summarized Denominator File

Contains EDB information for all non-cancer respondents. We used the EDB file to construct these records for all respondents and then dropped the records for those in PEDSF.

We started with 3,970,491 respondents, including a 'blank' HIC found on several surveys.

We dropped the following:

1	Blank HIC
3,861	No EDB record could be obtained
12	Died prior to 1997 according to EDB

205,259 Found in PEDSF (160 people in PEDSF did not have an EDB in our file)

This left us with 3,761,358 viable respondents. We removed people who died more than 90 days before the date received on any survey (686 in 2013; 62 in 2014). The final SEER-CAHPS SumDenom has 3,760,610 records.

**This file contains respondents in the CAHPS who were not found in the SEER cancer database. Enrollment and entitlement data for these individuals is provided in the Summarized Denominator (SUMDENOM) file. This file contains information by calendar year for the months that the person was Medicare eligible, from 1996-2013. Variables include his/her unique Medicare health insurance claim number (the HIC is encrypted to protect confidentiality), date of birth, date of death (if any), sex, race, state of residence, enrollment in Part A and/or Part B, and enrollment in an HMO (if any) by month. This file can be used to identify persons to be included or excluded from an analysis, i.e. by sex, HMO enrollment, etc. The SUMDENOM file can be linked with the Medicare claims for the non-cancer cases by the HIC number (encrypted) which appears on all files.**

#### **C) Census files for ZIP code and Census Tract**

These files contain socio-economic indicators, such as % poverty by race, for each ZIP code and for each Census Tract. The Census Tract data is for 1990, 2000, and 2010. One has to link by ZIP code or by Census Tract to the PEDSF or SumDenom data to get the values for each person.

## **2) Claims Data**

We have claims data for 2002-2013. 2002-2004 claims are only available for those who answered a survey in 2011 or prior.

- A) MEDPAR: Inpatient Hospital claims data (There are some claims for earlier admission years)
- B) NCH: 100% Physician/Supplier Data
- C) Outpatient SAF
- D) Hospice SAF
- E) HHA SAF: Home Health Agency
- F) DME SAF: Durable Medical Equipment (This data is for 2003-2013)

Specific IDs, such as Provider ID, UPIN, NPI and Referring PINs are masked, as are ZIP codes. If you expect to need any of these fields, please discuss this with us as special permission for their release is needed. While the structure of these files is very similar to the SEER-Medicare versions, there are some differences. Most notable, all dates in the Claims data are YYYYMMDD. Some field names are different and in a few case, data items are in slightly different positions.

## **3) Survey Data**

Note that any given respondent may have more than one survey.

MA – Medicare Advantage

FFS – Fee For Service  
PDP – Prescription Drug Plan  
PPO – Preferred Provider Organization

- A) CAHPS.survey9705.pHIC.cpt.gz - cported SAS data set for 1997-2005 (MA,  
FFS)  
    (FFS is only for 2000-2004)
- B) CAHPS.survey0713.pHIC.cpt.gz - cported SAS data set for 2007-2013  
    (MA, MA+PDP, FFS, FFS+PDP, MA PPO)
- C) cahps\_survey\_formats\_2014.sas7bcat – SAS format library

## SEER-CAHPS Survey Data Information

### Files

The CAHPS Survey Data has 5 survey types

FFS – Fee For Service

FFS PDP – Fee For Service plus Prescription Drug Plan

MA – Medicare Advantage

MA PDP - Medicare Advantage plus Prescription Drug Plan

MA PPO - Medicare Advantage Preferred Provider Organization

MA data is available for 1997-2005, 2007-2013.

FFS data is available for 2000-2004, 2007-2013.

MA PDP and FFS PDP are available for 2007-2013.

MA PPO is available for 2008-2012.

These are stored in 2 files. The survey data division was based on the compatibility of the question data within the relevant surveys.

1997-2005 data for MA and FFS

2007-2013 data for all survey types

### Death before Date Survey Received

When the files were originally constructed, people who died prior to 1996 were removed as they could not have answered the surveys, which started in 1997. There were still 7665 surveys that had a Date Survey Received after the Date of Death in the 2013 linkage, 7423 in the 2014 linkage.

<= 30 days (KEEP)	3875 (55.55%)	4123 (55.54%)
31-90 days (KEEP)	2951 (38.50%)	3217 (43.34%)
>= 91 days (DROP)	839 (10.95%)	83 (1.12%)

We believe respondents whose surveys fall in the 30-day window may actually have answered the survey themselves, but due to time in transit or disruption in the household due to their demise, the receipt of the survey could reasonable have been delayed.

We believe surveys in the 31-90-day window may have been answered by proxies with the addressee in mind. 1815 of the 3217 surveys in this window (56.4%) did have the PROXY flag set to YES. Also, survey received date was estimated for earlier years, and may be up to 82 days off (see svy\_dt\_rcv below).

For those surveys received more than 90 after days, we felt

a) the responses were less likely to be an accurate reflection of the addressee's experience

b) for the longer windows, the addressee was not alive during the 6 months in question.

Therefore, we made the decision to remove all survey data where the Date Survey Received was more than 90 days after Date of Death. We also removed the PEDSF or SumDenom record for those respondents that had no survey data left after this removal.

Researchers should consider this issue carefully. One may wish to exclude all surveys where the respondent died prior to the date the survey was received. Alternately, one may wish to consider the timing, whether or not the survey was completed with PROXY help, what specific kind of proxy help is noted, how well the sampling demographics match the responses, or some combination of these factors to decide which entries to retain and which to discard.

### **Missing Values**

There are several missing values in the CAHPS survey data. For most researchers, this may not be important and you can run SAS analyses as you would normally. If you wish to go into more detail with respect to missing entries or conduct analyses of missing data patterns, it is important to review the various missing categories.

.	Question Not on Survey
.G	Good Skip based on Skip Pattern
.V	Valid Answer of 'Does not apply'
.D	Don't Know
.N	Not Answered, on Survey
.R	Refused
.A	Answered-Should have Skipped
.S	Skipped-Should have Answered
.I	Inconsistent Response (to previous questions)
.O	Out of Range (Invalid value coded)
.M	Multiple Response
.Z	Provider Doesn't Match Survey Type

#### **. Question Not on Survey**

Since multiple survey types and survey years are found within a single file and since the surveys were not consistent over time or type, some questions are not found on every survey. If a question was not asked at all, then for those years and types the missing value will be a simple . value.

For example, the question about getting a flu shot this year was only asked in 1998 and 1999 Medicare Advantage (MA) surveys. It would have a . value for 1997 and 2000-2005 MA surveys, which are stored in the same file structure.

If you are calculating percentages missing for a question or percent complete for a respondent, this value should not be included in the Numerator or Denominator.

**.G Good Skip based on Skip Pattern**

Some questions have leading Skip Pattern questions and the respondent is instructed to skip the question if they answered No.

For example, the question “Were you seen for an illness or injury?” is a Yes/No question. If the respondent answered NO, he should have skipped the question “How often did you get care for an illness or injury as soon as you wanted?” We will use this example again.

If the respondent answered NO to being seen for an illness or injury and skipped “How often did you get care as soon as you wanted”, then it is a good skip based on the skip pattern question.

If you are calculating percent complete for a respondent, this value should not be counted against them.

**.V Valid Answer of 'Does not apply'**

Some questions provided a valid response which effectively means ‘This doesn’t apply to me.’

For example, in the 2009 Medicare Advantage + Prescription Drug Plan (MA PDP) survey, question 37 is “In the last 6 months, how often did the PDP’s customer service give you the information/help you needed about prescription drugs?” The last choice is “I did not try to get information or help from my health plan’s customer service in the last 6 months”. This is a valid response, but was recoded to .V as it does not affect calculations of how satisfied people were in this area.

If you are calculating percentages missing for a question or percent complete for a respondent, this value should not be included in the Numerator or Denominator.

**.D Don't Know**

Some questions provide a valid response which effectively means “I don’t know.” However, these surveys were also given by phone and sometimes the respondent said they didn’t know. Both these types of responses are classified as .D values.

An example of the survey based response can be found in the 2009 MA PDP survey, question 34, “Have you ever asked anyone at your health plan to reconsider a decision not to provide or pay for health care or services?” The last choice is “Don’t know”. This was coded to .D Don’t know, but is a valid response.

If you are calculating percentages missing for a question or percent complete for a respondent, you should handle this value with care. See the portal for the surveys if you need to check a particular question’s valid responses.

[https://appliedresearch.cancer.gov/seer\\_cahps/study\\_documents/final-surveys](https://appliedresearch.cancer.gov/seer_cahps/study_documents/final-surveys)

**.N Not Answered, on Survey**

When a question is on the survey, but the respondent just didn’t answer it, the field will get a .N value. This is used for questions outside of the skip pattern set. For example, if gender was not answered, it would be set to .N value.

**.R Refused**

This was used for phone based surveys when the respondent refused to answer the question.

**.A Answered-Should have Skipped**

As with .G, this value is used for questions associated with a skip pattern question when the respondent answered a question that they should have skipped.

For example, if the respondent answered NO to being seen for an illness or injury and then said he always got care as soon as wanted, then he answered a question that should have been skipped.

**.S Skipped-Should have Answered**

As with .G, this value is used for questions associated with a skip pattern question when the respondent answered a question that they should have skipped.

For example, if the respondent answered YES to being seen for an illness or injury and then skipped "How often did you get care as soon as wanted", then he skipped a question that should have been answered.

**.I Inconsistent Response (to previous questions)**

If the respondent provided inconsistent responses outside of a skip pattern, then the 2<sup>nd</sup> question was set to .I inconsistent response.

For example, in the 2010 MA PDP survey, question 15 is "Do you have a personal Doctor?" If the respondent responded NO to that question, but in question 28, "How often did your personal doctor seem informed about the care you got from specialists?" he answered Never, Sometimes, Usually or Always, that is inconsistent with the original information that he didn't have a personal doctor, and was coded as .I value.

**.O Out of Range (Invalid value coded)**

If a question had 3 valid responses, but the coded value was 4, the value is out of range. This would be the result of bad coding, but the value is not useable and was reset to be .O out of range.

**.M Multiple Response**

If a question had multiple responses, for example the respondent answered that he 'Sometimes' and 'Always' got care as soon as wanted it, then it is coded as a .M, multiple responses given.

**.Z Provider Doesn't Match Survey Type**

For the 2012 MA PPO data, some people were sent this survey type even though the Part D contract was not of the correct type. These respondents had Part D that were known to be from PDP or had Part D where it was unclear whether the contract was from MAPD or PDP. There were 4660 such respondents. All their responses to Prescription Drug related questions were masked with .Z as they were not providing information on the same type of plan as the other respondents of this survey type.

**GENERAL**

These surveys were coded by different vendors over time. The earlier survey years (1997-2005) did not have as many missing categories. When the SEER-CAHPS data was cleaned for public use, we attempted to make missing characterizations consistent over time. Specifically, we reviewed the 1997-2005 data and used these missing values where appropriate. However, the user should be aware of this limitation if they wish to do analyses using distinct missing categories.

**Constructed fields in the CAHPS Survey data files**

There are some fields that were constructed as part of the CAHPS Survey data harmonization process. Some fields were initially present on only one set of files and had to be created for the other set. The others were constructed for both sets of data.

sa\_agecalc – the age calculated based on date of birth and November 30 of the svy\_year. This was present in 1997-2005; it was added to 2007-2013.

sc\_agercvd – the age calculated based on date of birth and svy\_dt\_rcv (described below).

**NOTE:** Both ages above were calculated using `yrdif(Start,End,'AGE')`. Sometimes, when the month and day of the dates were within one day, the resulting age was slightly off. As an example, if date of birth = 6/1/1970 and survey received date = 5/31/2010, if the calculated age was 40 even though the person does not turn 40 until the next day, this was considered to be slightly off. In these cases, age was recalculated to be  $(\text{End}-\text{Start})/365.25$ . Neither calculation method always results in the correct age because of leap year issues, and this adjustment was performed to ensure the `floor(age)` values were logical.

agecat – 7 age groups based on age question or sa\_agecalc if the question was missing or not distinct enough. This was present in 97-05; it was added to 07-13.

cmp\_xxxx – the 7 composite fields – Get Care Quickly, Get Needed Care, Doctor Communication, Customer Service, PDP Customer Service, Get Needed Drugs, Care Coordination. The questions mentioned below were converted to a scale of 1 to 100 (worst to best) and then the mean of all non-missing values was calculated for each survey.

cmp\_getcareqck: illasaw, rtnasaw.

cmp\_getndcare: sp\_getappt, pl\_getcare for 2007-2011



sp\_prob, sp\_probref, sp\_vst\_ofst, pl\_probcare,  
 pl\_probcare\_ofst,  
 md\_happy, md\_easyfind, pl\_delays for 1997-2005  
 cmp\_drcomm: md\_explain, md\_listen, md\_respect, md\_enghtime.  
 cmp\_cstsvr: cs\_csgetinfo, cs\_csrespect, pl\_ezpaper for 2007-2011  
 pl\_probinfo, cs\_allinfo, pl\_probpape, pl\_probpape\_ofst,  
 cs\_hlpfl for  
 1997-2005  
 cmp\_pdcstsvr: pd\_csgetinfo, pd\_csrespect, pd\_getcovinf, pd\_getcostinf  
 2007-2011 FFS+PDP, MA+PDP, MA+PPO only  
 cmp\_getnddrg: rx\_ezmeds, pd\_ezrxmeds, pd\_ezrxpharm, pd\_ezrxmail  
 for 2007-2011  
 rx\_get\_need, rx\_prob\_need for 1997-2005  
 cmp\_carecoord: 2012+ only (missing for all FFS-PDP)  
 md\_medrecs, mdtalkmeds, md\_getmngca, sp\_mdinformd,  
 Combination of (md\_testfup, md\_testasan)  
 See MCAHPS.composites.Summary for programmers.docx for additional

information.

cnd\_heart – Yes/No flag for heart was created for the 2007-2011 data based on heart\_attack

and angina. If either was set to yes, then cnd\_heart = yes. If at least one was set to no, then cnd\_heart was set to no regardless of whether both were filled in. cnd\_heart was left blank only if both were blank.

cnd\_comorb\_cnt – Count of comorbid conditions present, it excludes cancer. The 2 heart conditions are counted 1 time for the 2007-2011 data.

sa\_contract, sa\_contract\_AB, sa\_contract\_D - Masked version of the contract numbers for the plan, for Part A&B and for Part D.

If you expect to need the true ID, please discuss this with us as special permission for their release is needed.

sc\_cancer\_conflict – SEER-CAHPS field based on SumDenom, PEDSF, 1<sup>st</sup> DX date and cnd\_cancer

In the 2008 and forward surveys, a set of questions was asked, have you ever been told you have.... One of these was cancer (cnd\_cancer). We also have information about cancer by virtue the respondent's inclusion in SumDenom or PEDSF and the earliest diagnosis date in PEDSF. We added a flag to indicate when there were conflicts between these two types of information.

In order to determine the first month of diagnosis (because entries in PEDSF are sorted by sequence number), we compared the dates of diagnosis. If one date had an unknown month but the same year as

another, we used the sequence order. If the first date of diagnosis had an unknown month, we use January to determine this flag.

Since we only have month for diagnosis, if the survey was received in the same month as diagnosis, we considered this as the survey being POST diagnosis.

	.	<b>Question not on Survey</b> (any survey that didn't ask the question)
	.N	<b>Not Answered, on Survey</b> (all variations of missing for cnd_cancer were collapsed)
<b>PEDSF pre dx)</b>	0	<b>No Conflict: YES + PEDSF post dx; NO + (SumDenom or</b>  (cnd_cancer = YES and in PEDSF and svy_dt_rcv >= Diagnosis OR cnd_cancer = NO and in PEDSF and svy_dt_rcv < Diagnosis OR cnd_cancer = NO and in SumDenom)
	1	<b>Conflict (Can Ignore): NO + PEDSF (post/dx)</b> (cnd_cancer = NO and in PEDSF and svy_dt_rcv >= Diagnosis We think SEER is more accurate)
	2	<b>Conflict (Timing): YES + PEDSF pre dx</b> (cnd_cancer = YES and in PEDSF and svy_dt_rcv < Diagnosis)
	3	<b>Conflict (cohort): YES + SumDenom</b> (cnd_cancer = YES and in SumDenom)

sc\_dual\_status - SEER-CAHPS field indicating Medicaid Dual Eligible status based on logic

provided by CMS.

Different fields were provided to capture this concept in the different survey years.

The two main variables (and supplemental information from a third) are used to calculate a single field for analytic use.

sa\_mdcd\_stus - Sample Members who are Dual Eligible: missing 2007, 2008

sa\_dual\_stus - Dual Status Description: missing for 2010.

sa\_deemd\_lis - Deemed For Low Income Subsidy

sc\_dual\_status = 1 if

2007: sa\_dual\_stus = 02, 04, 08

2008: (Survey is not FFS+PDP and sa\_dual\_stus = 01-06, 08, 09)

OR

(Survey is FFS+PDP and sa\_deemd\_lis = Y or 0)

\* If Survey is FFS+PDP and sa\_deemd\_lis is not Y or 0,  
sc\_dual\_status is missing.

2009-2013: sa\_mdcd\_stus = Y  
 Otherwise sc\_dual\_status = 0

sc\_lis - SEER-CAHPS field indicating Low Income Subsidy status based on logic provided by CMS.

Two fields are related to low income subsidy in the original file:

sa\_lis: Low Income Subsidy

sa\_deemd\_lis: Deemed for Low Income Subsidy

However the relationship between the two is not easily discernible. CMS provided information that informed our decision to consider a case as LIS if either sa\_lis = 'Y' or sa\_deemd\_lis = 'Y'.

sc\_pcs\_sf12\_1

sc\_mcs\_sf12\_1

sc\_pcs\_vr12

sc\_mcs\_vr12

Physical condition score and Mental condition scores.

SF12\_1: based on the SF12 version 1 question style, available on FFS 2000-2004.

VR12: based on the VR 12 question style, available on 2007 data.

These were constructed based on the standard algorithms.

sc\_race – SEER-CAHPS Race field based on the responses to the Hispanic and Race questions.

If Hispanic = YES sc\_race = Hispanic

Else if one single race

was selected

sc\_race = select race (White, Black, Other,

Asian, Native Amer)

Else if multiple other

racess were selected

sc\_race = Multiple Races

svy\_Yearsxxx – 3 different field names for MA, FFS, All. This indicates when the same

person has surveys in multiple years of the survey for the different sets. MA 1997-2005; FFS 2000-2004; All 2007-2013.

These exist on 1997-2005 MA and 2000-2004 FFS, but they were reconstructed to ensure the data matches the flag. A field was also constructed for All 2007-2013.

svy\_dt\_rcv

Not all surveys had a date received. We imputed the svy\_dt\_rcv using the date in the middle of the start and end of survey collection. The information below was obtained during discussions with CMS.

The survey collection start and end dates are:

2000 - October 9, 2000 - February 1, 2001  
 2001 - September 12, 2001 - February 1, 2002  
 2002 - September 10, 2002 - February 7, 2003  
 2003 - September 10, 2003 - February 21, 2004  
 2004 - September 8, 2004 - February 7, 2005  
 2005 - Medicare Advantage ONLY (no FFS) - January 6 -April 30,  
 2006

For 2006 forward, data collection shifted to the Winter/Spring of the year following the reference year and exact dates are less documented due to the many surveys fielded after the introduction of the Medicare Prescription Drug Plan, i.e. the surveys were for the MA, MA PD, PDP, PPO and Medicare FFS. For the reference years 2006-2011 data collection, each of the above surveys was attempted to begin and end in the same timeframe, but exact dates for each survey differed by a few days due to production schedules and mailings:

2007 All Medicare CAHPS - early February – mid-June 2007  
 2008 All Medicare CAHPS - early February – mid-June 2008  
 2009 All Medicare CAHPS - early February – mid-June 2009  
 2010 All Medicare CAHPS - second week of February – mid-June  
 2010  
 2011 All Medicare CAHPS - third February – third week of-June  
 2011  
 2012 All Medicare CAHPS - third February – third week of-June  
 2012  
 2013 All Medicare CAHPS - end February – early June 2013

- Early: end of first week of the month
- Week of the Month: last date of those weeks.
- Mid: 15th. In 2010, use June 20.

The “1997” survey we believe is a misnomer in that although it was required by the 1997 Balanced Budget Amendment, all reports on it show that it was not fielded until early 1998 – and then a second survey was fielded in the fall of 1998.

For 1997 and 1998 the dates below were the beginning of the surveys. <CMS was> fairly certain that data collection used the same protocols as in later years and that would lead me to say that a mid-point of data collection for “1997” would be April 8, 1998; and for 1998 November 10, 1998.

1999 M+C CAHPS - September 1999 and December 1999  
 (The report does not include the exact dates for this year. M+C was predecessor for MA.)

Thus, the final dates used for Date Received, if NO date was specified, are:

For 1997:	start = 4/8/1998	end = 4/8/1998	4/8/1998
For 1998:	start = 11/10/1998	end = 11/10/1998	11/10/1998
For 1999:	start = 9/15/1999	end = 12/15/1999	10/31/1999
For 2000:	start = 10/9/2000	end = 2/1/2001	12/06/2000
For 2001:	start = 9/12/2001	end = 2/1/2002	11/22/2001
For 2002:	start = 9/10/2002	end = 2/7/2003	11/24/2002
For 2003:	start = 9/10/2003	end = 2/21/2004	12/01/2003
For 2004:	start = 9/8/2004	end = 2/7/2005	11/23/2004
For 2005:	start = 1/6/2006	end = 4/30/2006	03/04/2006
For 2007:	start = 2/9/2007	end = 6/15/2007	04/13/2007 /* we have exact date */
For 2008:	start = 2/8/2008	end = 6/15/2008	04/12/2008 /* we have exact date */
For 2009:	start = 2/6/2009	end = 6/15/2009	04/12/2009 /* we have exact date */
For 2010:	start = 2/12/2010	end = 6/20/2010	04/17/2010 /* we have exact date */
For 2011:	start = 2/18/2011	end = 6/24/2011	04/22/2011 /* we have exact date */
For 2012:	start = 2/18/2012	end = 6/24/2012	04/22/2012 /* we have exact date */
For 2013:	start = 2/28/2013	end = 6/2/2013	04/16/2013 /* we have exact date */

### **Weights in the CAHPS Survey data**

One of the features of the SEER-CAHPS database is the availability of data on both Medicare fee-for-service (FFS) and Medicare Advantage (MA) enrollees. However, each MA plan has to survey a representative sample of its insurees, so the MA population is over-sampled relative to those with FFS Medicare. In order to produce estimates that better represent the distribution of FFS and MA enrollees in the Medicare population, the SEER-CAHPS data provides two different weight variables.

There are two types of weights found in the CAHPS survey files.

WGT\_SIMPLE is a base weight calculated to make the sample representative of the beneficiary populations in the units in the original design. All years and survey types have this type of weight. Using the base weight variable (wgt\_simple) allows the analyst to produce estimates that are representative of the beneficiary populations in the units of the original design. For the MA and standalone PDP sample, these units were contracts; for the FFS sample, these were states.

WGT\_RAKED was constructed after, using a raking weighting procedure (loglinear weights calculated by iterative proportional fitting) to weight the respondents to match the control distributions estimated from the first round sample (with base weights). In some cases small cells were collapsed with adjacent cells, to avoid extreme weights. This corrected for biases due to differential non-response associated with beneficiary characteristics as well as reducing the effects of random variation in non-response. MA and FFS 2000-2004 does not have this type of weight as the group calculating the weights were unable to get data on non-respondents from that period. Using the raked weight variable (wgt\_raked) allows the analyst to correct for biases arising from differential nonresponse associated with beneficiary characteristics as well as reducing the effects of random variation in nonresponse. Currently, raked weights are only available for respondents with surveys in 2011 or later.

Additional information can be found in the yearly Medicare CAHPS Technical Reports, in the Appendix related to Weighting. Some examples of recent reports are referenced here:

2008 Medicare CAHPS Technical Report, Appendix 2.01: Weighting of CAHPS-Medicare data, 2008, version 1.0

2010 MCAHPS Tech Report, Appendix 2.04: Individual-Level Weight Construction

2012 Medicare CAHPS Technical Report, Appendix 2.04: Individual-Level Weight Construction

**The following text may be used in describing the weights briefly in manuscripts:**

**“Data were weighted to represent the enrolled population of contract by county combinations, followed by applying a raking procedure (loglinear**

**weights by iterative proportional fitting) to respondents to match weighted sample distributions within each contract (or state, for FFS beneficiaries unenrolled in a PDP) of gender, age, race/ethnicity, Medicaid and low income supplement eligibility, Special Needs Plan status, PD enrollment, and zip-code level distributions of income, education, and race/ethnicity.”**