MA & PDP CAHPS
Individual-Level Weight Construction

OVERVIEW

Base weights were calculated to make the sample representative of the strata in the original design. For the PDP sample, these strata were defined as contracts; for the MA sample, as survey type (MA-only and MA-PD) within contracts; and for the FFS sample, they were defined as states. We then used a raking procedure (loglinear weights calculated by iterative proportional fitting) to weight the respondents to match distributions estimated from the sample (with base weights). In some cases, small cells were collapsed with adjacent cells, to avoid extreme weights, for example highest income quartile with second highest if there were very few in the top quartile. This corrected for biases due to differential nonresponse associated with beneficiary characteristics as well as reducing the effects of random variation in nonresponse.

BASE WEIGHTS

We used final enrollment numbers to determine the first round sampling plan for the CAHPS sample which contained about three times as many beneficiaries as were required in the final sample. Survey data was collected on individuals based on the original sampling plan.[1] Weights were computed for each contract or state stratum by dividing the enrollment for that stratum by the number of respondents in that stratum (WS_WES_COMP = 1).[2] In the case of MA, the MA-Only and MA-PD weights were calculated separately by survey type in each contract. For the calculation of weights in the case of MA-Only enrollees in strata for which there were no MA-Only surveys (or responses), these non-surveyed enrollees are collapsed into the complementary MA- PD survey-type within contract.

NONRESPONSE/POST-STRATIFICATION WEIGHTING

We weighted the responses obtained to represent all persons in their state (FFS) or contract (MA and PDP), thus correcting for biases due to associations of nonresponse with variables available in the sampling frame or via geocoding. To do this, we weighted the data to make weighted distributions of these variables match unbiased estimates of their population totals. For this step, all first round sampled observations were used. These included those cases that were not chosen in the second round of sampling (in which only one-third of those initially sampled were chosen to receive a survey). From that list of cases, we further excluded those lacking both zip code and state information and those from outside the U.S. and Puerto Rico). Use of this enlarged sample for estimating control totals or ‘targets’ improves the accuracy of these estimates.

Base weights by contract were calculated from the enrollment tables from CMS for the entire first round sample. For most respondents in FFS, we calculated base weights in a similar

---

[1] When we designed the sample, we used only the enrollees in the 50 states, Washington, D.C., and Puerto Rico.
[2] The survey and respondent counts entering the weighting do not include Institutionalized or Deceased Individuals or others who were excluded from the survey.

fashion, but by state instead. Using these base weights, we obtained control totals or targets by tabulating base-weighted distributions for variables available in the first round sample file; these variables included age, race, sex, dual-eligible status, and type of survey for MA (MA-PD or MA-Only). Zip-code level variables for racial/ethnic composition and socioeconomic characteristics of the beneficiary's zip code tabulation area (ZCTA) were obtained from the five year estimates from the American Community Survey.

Implementation of weighting was immensely facilitated by use of a suite of programs in the R programming language we developed specifically for weighting of the Medicare CAHPS surveys. The software incorporates the following convenient features:

- Marginal weighting cell definitions are specified in a simple formula notation, requiring no additional coding.
- The software accommodates two computational approaches for fitting steps: proportional fitting for one-way layouts with many categories (such as race by contract interaction) and Newton-Raphson estimation for multiple variables that are not fully interacted.
- Collapsing of constraints (collapsing of cells in one-way layouts, and removal of inconsistent constraints in multi-way fitting steps) is pre-specified and automated.
- Tools are provided for monitoring and reporting convergence and consistency with constraints.
- Setting up the fitting sequence and iteration requires only a few lines of easily modified code.

Our weighting programs examined the individual-level respondent data to identify cells corresponding to marginal controls for which there were very few cases sampled or responses collected, making it more likely that sampling variability in those categories would contribute to extreme weights. In such a case, cells were collapsed according to pre-specified sequences to remove these potentially extreme cells. For example, large numbers of race×contract cells were very small because there were very few members of that race/ethnicity in the contract's area; in such case the smaller race/ethnicity groups might be collapsed into a combined category.

We then use an iterative proportional fitting process to rake the weights of the responses (Deming and Stephan 1940; Purcell and Kish 1980) to match the marginal control counts so the weighted distribution of respondents better represented the full population. We also forced all respondent weights to be at least 1, since every respondent represents her/himself. The iteration proceeded until weights converged to values that were negligibly changed by additional iterations, and all the marginal targets were met within close tolerances. For MA, PDP, and FFS, this change was well below one percent. Those respondents with larger relative changes typically fell into cells with low numbers of responses (many near our cutoff value of eleven responses) relative to the original sample design.

Section A lists the control variables or combinations of variables whose distributions or cross-tabulations in the adjusted sample data were matched by the weighting procedure. A large collection of demographic variables, including both those drawn directly from the CMS enrollment files (such as age, SSA race, dual eligible, and LIS) and those drawn from linked census data (characteristics of the Zip Code Tabulation Area such as median income or racial composition) were used in the definition of control variables. Quartiles were derived for these

Zip Code variables computed within the respective state. Main effects were generally interacted with contract to improve the representativeness of data on each contract. Most interactions of individual-level variables were not further interacted with contract but were interacted only at the national level, to avoid non-convergence or extreme weights due to many very small cell counts. Thus these interactions were controlled to make national-level analyses representative of national distributions on a finer set of variables.

We produce a summary of the distributions of the final post-stratification/nonresponse-weights for each of the three main segments of the Medicare population and compare them to the plan-level (state-level for FFS) base weights used in the initial analyses for contract-level reporting. The coefficient of variation (CV) is the standard deviation of the weights divided by their mean. The CV is related to the amount of inflation of variance due to unequal weighting for national estimates of a characteristic not related to the weights by the variance inflation factor ($1+CV^2$). The most important observation is that the CV of final weights is only slightly larger (in each sector) than that of the contract-level (or state-level, for FFS) weights. Thus, most of the inefficiency of the design for national analytic estimates is due to the requirement to collect a reportable sample size for every contract, and very little additional inefficiency is driven by nonresponse/post-stratification weighting.

Section A: Control Variables

FFS Model for states other than Puerto Rico and Virgin Islands

Control variables in National (Overall) FFS Model, all formulae include control for State

- Dual Eligibility
- Gender
- Age
- Race
- Age and Gender
- Median Household Income Zip Code Census data (Quartiles)
- Percent Hispanic Zip Code Census data (Quartiles)
- Percent Black Zip Code Census data (Quartiles)
- Median High School Graduation Zip Code Census data (Quartile)
- Age, Median Household Income (above/below median), and Quintile of state ranking with respect to Median Household Income.
- Percent Black Zip Code Census data (above/below Median), Race, and Quintile of state ranking with respect to Percent Black.
- Percent Hispanic Zip Code Census data (above/below Median), Race, and Quintile of state ranking with respect to Percent Hispanic.
- Percent HS Graduation Zip Code Census data (above/below Median), Dual Eligibility, and Quintile of state ranking with respect to Percent HS Graduation.

PDP Model

Control variables in PDP Model

- Contract
- Contract and Dual Eligibility
- Contract and Gender
- Contract and Age
- Contract and Race
- Contract, Age and Gender
- State and Race
- State and Median Household Income Census Quartile
- State and Percent Hispanic Zip Code Census Quartile
- State and Percent Black Zip Code Census Quartile
- State and Percent High School Graduate Census Quartile
- State and Dual Eligibility
- Age, Median HH Income (Above/Below Median), Quintile of state ranking on Median Household Income, and State.
- Race, Percent Hispanic Zip Code Census data (above/below Median), Quintile of state ranking with respect to Percent Hispanic, and State.
- Race, Percent Black Zip Code Census data (above/below Median), Quintile of state ranking with respect to Percent Black, and State.

- Percent HS Graduation Zip Code Census data (above/below Median), Dual Eligibility, Quintile of state ranking with respect to Percent HS Graduation, and State.

MA Model

Control variables in MA Model

- Contract and Survey Type
- Contract and Dual Eligibility
- Contract and Gender
- Contract and Age
- Contract and Race
- Contract, Age and Gender
- State and Race
- State and Median Household Income Census Quartile
- State and Percent Hispanic Zip Code Census Quartile
- State and Percent Black Zip Code Census Quartile
- State and Percent High School Graduate Census Quartile
- State and Dual Eligibility
- Age, Median HH Income (Above/Below Median), Quintile of state ranking on Median Household Income, and State.
- Race, Percent Hispanic Zip Code Census data (above/below Median), Quintile of state ranking with respect to Percent Hispanic, and State.
- Race, Percent Black Zip Code Census data (above/below Median), Quintile of state ranking with respect to Percent Black, and State.
- Percent HS Graduation Zip Code Census data (above/below Median), Dual Eligibility, Quintile of state ranking with respect to Percent HS Graduation, and State.