

NATIONAL HEALTH AND AGING TRENDS STUDY (NHATS)
Development of Round 1 Survey Weights

November 29, 2012

Suggested Citation: Montaquila, Jill, Freedman, Vicki A., Spillman, Brenda, and Kasper, Judith D. 2012. National Health and Aging Trends Study Development of Round 1 Survey Weights. NHATS Technical Paper #2. Baltimore: Johns Hopkins University School of Public Health. Available at www.NHATS.org. We acknowledge the valuable contributions of Graham Kalton of who led the NHATS sample design and provided helpful comments, along with Brad Edwards, on earlier versions of this paper. We also thank David Ferraro, who played an instrumental role in the development of the Round 1 weights and produced several tabulations that appear in this paper. This technical paper was prepared with funding from the National Institute on Aging (U01AG032947).

1. Introduction

The NHATS public use data support weighted analysis of Medicare beneficiaries ages 65 and older living in the contiguous United States on September 30, 2010. The survey weights included with the public use file account for differential probabilities of selection and adjust for potential bias related to unit nonresponse.

For Round 1 of NHATS, two types of sampling weights have been produced: a tracker weight (on the Tracker file with the variable name w1trfinwgt0) and an analytic weight (on the Sample Person file with the variable name w1anfinwgt0). For variance estimation (see Section 7), NHATS has also included replicate versions of these weights (w1trfinwgt1-w1trfinwgt56 and w1anfinwgt1-w1anfinwgt56).

The methodology that was used to develop these weights and appropriate uses of each of these weights are discussed in the following sections. The next section provides an overview of how cases were classified for purposes of weight development. Sections 3 and 4 detail the creation of the tracker and analytic weights, respectively. Section 5 reports on the effect of weighting adjustments on the precision of NHATS survey estimates. Section 6 provides guidance on the use of the tracker and analytic weights. A final section provides information on the proper calculation of variance estimates to account for the complex design and estimation procedures used in NHATS.

2. Definition of Respondent

In the development of survey weights, an important first step is the classification of cases into groups based on eligibility and response status. For Round 1 of NHATS, the following disposition codes map into respondent, ineligible, and nonrespondent statuses:

Table 1. Classification of Round 1 NHATS Sample for Weight Development Purposes

Disposition code	n	Classification for Tracker Weight	Classification for Analytic Weight
60 Complete	7,550	Respondent	Respondent
61 Complete, NH facility	468	Respondent	Respondent
63 Complete SP, FQ not complete	59	Respondent	Respondent
64 Complete FQ, SP not complete	168	Respondent	Nonrespondent
75 Physically/mentally unable to participate, no proxy	67	Nonrespondent	Nonrespondent
76 Too ill to participate, no proxy	83	Nonrespondent	Nonrespondent
77 Refusal, Sample Person	2,733	Nonrespondent	Nonrespondent
78 Language barrier	61	Nonrespondent	Nonrespondent
79 Unable to locate	205	Eligibility unknown	Eligibility unknown
80 Unavailable during field period	13	Nonrespondent	Nonrespondent
82 Outside of Primary Sampling Unit	78	Nonrespondent	Nonrespondent
83 Ineligible (moved out of contiguous US)	77	Ineligible*	Ineligible*
85 Refusal, facility	6	Nonrespondent	Nonrespondent
86 Deceased	697	Ineligible*	Ineligible*
87 Refusal, proxy	142	Nonrespondent	Nonrespondent
88 Work stopped	4	Nonrespondent	Nonrespondent
Total and Number Assigned Weight	12,411	9,019	8,077

* These cases were eligible at the time of sample selection (based on information provided on the 20% HISKEW file that was used for beneficiary selection) but were found to be ineligible by the time of the data collection period.

SP=Sample Person interview; FQ=Facility Questionnaire

For the Tracker weight, only cases classified as Respondents and Ineligible are assigned a positive weight; for the Analytic weight, only Respondents are assigned a positive weight. Cases for which at least one survey component is available (codes 60, 61, 63 and 64) are considered respondents for purposes of the tracker weight. Those who became ineligible for the sample after they were selected, either because they died or moved out of the contiguous U.S. by the time of the fieldwork, have positive tracker weights, but their analytic weights are valued zero. For the analytic weight, cases residing in a nursing home (code 61) or with a completed Sample Person (SP) interview (60, 63) were considered respondents (n=8,077). For the SP interview, cases were required to have completed the self-reported disability protocol (through the section on Participation; PA) to be considered complete.

3. Computation of Tracker Weights

The computation of the tracker weight began with the base weight and included two adjustments—a weighting class adjustment for nonresponse and a raking adjustment to estimated population totals from the HISKEW file.

The base weight was computed as:

$$w_{Bi} = \frac{1}{p_{1i}p_{2i}p_{3i}},$$

where p_{1i} is the PSU selection probability for the PSU from which SP i was selected; p_{2i} is the ZIP cluster selection probability for the ZIP cluster from which SP i was selected, conditional on having selected the PSU; and p_{3i} is the selection probability of SP i , conditional on having selected the PSU and ZIP cluster. (For ease of discussion here, we are not using notation for the specific PSU and ZIP cluster.) See Montaquila, Freedman, Edwards, and Kasper (2012) for details of the computation of probabilities of selection.

The next step was to develop and apply to the base weights a sample-based weighting class adjustment (Kalton and Flores Cervantes 2003). For this adjustment, each SP was classified as either a respondent (R), a nonrespondent (NR), an ineligible case (I), or a case with unknown eligibility (U) (see Table 1 above). This step involved adjusting weights within weighting classes, also referred to as *nonresponse adjustment cells*. The cells were formed using variables that were available for both respondents and nonrespondents.

For the adjustment to be effective in reducing nonresponse bias in key estimates from the survey, the variables selected to form the cells should be associated with both the response propensities (the probabilities of response) and those key survey items. For NHATS, predictors likely to be related to late-life disability were considered. Potential variables for creating non-response cells came from five sources:

- Beneficiary information from the sampling frame (the 20% HISKEW File), including demographic characteristics of the beneficiary (e.g., age as of September 30, 2010, gender) and geographic information (e.g., census division, metro and micropolitan status) based on the beneficiary's address in CMS' Medicare Enrollment Database (EDB) and an indicator of sample release group (see Montaquila, Freedman, Edwards, and Kasper (2012) for details of the sample release process);
- County-level demographic information based on the 5% HISKEW file (e.g., percent of beneficiaries in the county who are Black; percent of beneficiaries in the county who are Hispanic) for the county linked to the beneficiary's address from the EDB;

- Census tract-level information based on the 2006-2010 5-year American Community Survey (e.g. tract-level demographic information), based on linkages to the beneficiary's address from the EDB;
- An indicator that the beneficiary's address from the EDB matches an address on a national 2009 list of licensed assisted living facilities¹; and
- An indicator of whether the beneficiary could be considered a nursing home resident based on a match to records from the Minimum Data Set (MDS), which contains periodic assessments for all Medicare or Medicaid certified nursing homes.

The latter indicator was based on an algorithm developed by Kasper, Edwards, and Freedman to identify beneficiaries who had a pattern of records in the MDS from January 1, 2011-October 31, 2011 consistent with a long-term resident rather than short-term skilled nursing stays. (See Appendix A for details).

Appendix Table B1 provides (base weighted) response rates (taking into account differential probabilities of selection only) by categories of the various indicators. We used these variables as input to a classification tree analysis to determine which of these variables were associated with nonresponse. This approach uses a search algorithm to identify variables associated with response propensities. At each step in the process, chi-square tests were performed to determine the most significant predictor of response, given the set of conditions already specified in the particular "branch." We also set a minimum cell size of 50.²

We forced the first branch to be nursing home resident (yes/no) because nursing home residents were not required to complete an SP Interview and therefore the underlying nonresponse process differed from (and was much lower than that for) sampled persons in the community and in other residential care settings. We included all variables in both branches, except the assisted living match indicator, which was limited to the non-nursing home branch.

We considered imposing a similar break for residents of licensed assistant living facilities (because a small number of individuals in non-nursing home residential care settings who had only a facility interview were retained in the tracker file), but opted not to for two reasons. First, the list did not cover all residential care places and second, our investigations suggested that the EDB address and actual address where the sampled persons lived differed at a much higher rate among those in assisted living than for those in other community settings. Thus, although the list was complete for 2009, the match of EDB address to the list did not provide a good representation of residential care settings broadly defined as in NHATS.

¹The list was compiled by the "Shaping Long Term Care in America Project" at Brown University funded in part by the National Institute on Aging (1P01AG027296)." See Appendix A for details.

²The classification tree analysis is designed to work with categorical predictor variables. Alternatives to this approach are propensity modeling based on logistic regression and Cartesian product cross-classification. The logistic regression approach uses a parametric model to identify predictors of response. When the pool of potential predictors includes continuous variables and categorizing the continuous variables would result in substantial losses of information, logistic regression modeling would be preferred over classification tree analysis. The Cartesian product cross-classification approach involves specifying a set of adjustment cell variables based on prior experience (generally, (1) prior analyses that identified predictors associated with response propensities; and/or (2) predictors associated with response and/or subject matter expertise that informs the choice of variables).

For respondents living in the community and other residential settings (not nursing homes), final non-response cells included 15 indicators (indicated in Appendix Table A1 with a * for the non-nursing home branch and ^ for the nursing home branch). Combinations of these variables created 26 unique nonresponse cells among the non-nursing home group and 8 nonresponse cells among the nursing home residents (See Appendix B Figures 1 and 2). Note that although the second release (see Replicate Release Group Appendix Table B1) had a marginally lower response rate, the release group indicator was not identified by the classification tree algorithm as a significant predictor of response.

Once the cells were constructed, the base weights of the respondents within a nonresponse adjustment cell were inflated to compensate for the base weights of nonrespondents in that cell, as well as a proportion of the cases with unknown eligibility estimated to be eligible. We assumed that the proportion of the unknowns who would be eligible is the same as the proportion among cases with known eligibility. That is, the nonresponse adjustment factor, f_{1b} , applied to each respondent in adjustment cell b was:

$$f_{1b} = \frac{\sum_{j \in R_b \cup NR_b} w_{Bj} + \sum_{j \in U_b} \left(\frac{\sum_{k \in R_b \cup NR_b} w_{Bk}}{\sum_{k \in R_b \cup NR_b \cup I_b} w_{Bk}} \right) w_{Bj}}{\sum_{j \in R_b} w_{Bj}}$$

where R_b , NR_b , U_b , and I_b are the sets of respondents, nonrespondents, cases with unknown eligibility, and ineligible, respectively, in adjustment cell b . For SP i who falls in adjustment cell b , the nonresponse adjusted weight, w_{Ni} , is

$$w_{Ni} = \begin{cases} w_{Bi} f_{1b}, & i \in R_b \\ w_{Bi}, & i \in I_b \\ 0, & i \in NR_b \cup U_b \end{cases}$$

Note that in this adjustment, the weights of ineligible are not affected. It is not appropriate to adjust the weights of these cases at this stage, but it is important to carry their base weights forward for use in the raking adjustment (described below).

The final step in creating the tracker weight involved raking the nonresponse adjusted weights, w_{Ni} , to control totals developed from the 5% HISKEW (September 2010 HISKEW) that was used for sampling. Raking involves adjusting the weights to match a set of population totals or proportions (or reliable estimates thereof) for two or more dimensions. The external population totals are referred to as control totals. The raking procedure is carried out in a sequence of adjustments: first, the weights are adjusted to the control totals for one marginal distribution (or dimension), then to the control totals for the second marginal distribution, and so on. One sequence of adjustments to the marginal distributions is known as a cycle or iteration. The procedure is repeated over several iterations until simultaneous convergence of the weighted totals to all sets of marginal distributions is achieved. Raking may be used to improve the precision of estimates, particularly estimates of totals; to adjust for “residual”

nonresponse bias (bias not already adjusted for in the separate nonresponse adjustment); to adjust for differential frame coverage of subgroups of the target population; and to ensure that survey estimates of the sizes of key subgroups match reliable external estimates.

For consistency, the raking adjustment also included the ineligible (primarily deaths), since the frame that served as the source of the control totals also includes beneficiaries who were ineligible for NHATS. These ineligible beneficiaries on the frame cannot be identified as such when the sample is initially drawn, but once identified are excluded from the computation of the control totals.

Four dimensions were used in this raking adjustment:

- (1) Age category (65-69, 70-74, 75-79, 80-84, 85-89, 90+) by sex by race from the EDB (Black; non-Black);
- (2) Age category (65-69, 70-74, 75-79, 80-84, 85-89, 90+) by Census region;
- (3) Age category (65-69, 70-74, 75-79, 80-84, 85-89, 90+) by MSA status (from the HISKEW); and
- (4) Age category (65-69, 70-74, 75-79, 80-84, 85-89, 90+) by a binary indicator of whether the SP was enrolled in Medicare prior to age 65.

The final tracker weight for respondent i was then calculated as:

$$w_{Fi} = w_{Ni}f_{2c},$$

where f_{2c} is the raking adjustment factor for raking cell c (where the raking “cell” is defined as the full cross-classification of the raking dimensions), where SP i has the attributes corresponding to the levels of the dimensions that define raking cell c .

4. Computation of Analytic Weights

The computation of the analytic weights begins with the final tracker weight. A weighting class adjustment was developed for the class of nonrespondents living in residential care other than nursing homes who had completed a facility interview but not a Sample Person interview ($n=168$; designated as code 64). The approach was designed to preserve the tracker weight distributions by residence type (nursing home, other residential care, and community). That is, we allowed the weights of residential care non-nursing home cases ($n=353$) to be adjusted to account for similar cases missing the SP Interview.

Because the sample size is much smaller for this nonresponse adjustment, only a subset of variables used in tracker weight classification tree analysis was considered for the analytic weight nonresponse adjustments (see Appendix Table B2). Five variables (designated with * in Table B2) were retained in the final classification tree, forming 7 cells (see Appendix B Figure 3).

As a final step, we applied a raking procedure so that marginal totals based on the tracker and analytic weights would match by 5-year age groups, sex, race, region, micro/metropolitan status, and whether Medicare was received before age 65.

5. Design Effects Related to Weighting

Although weighting adjustments are aimed at reducing bias, increased variation in weights generally increases the variances of survey estimates (Kish, 1965). Thus, in the development and implementation of the weighting methodology for NHATS, care was taken to balance the bias reductions against the potential increases in variance.

The estimated overall design effect due to variation in the base weights was 1.28. After applying nonresponse adjustments within cells determined by the classification tree results, the estimated overall design effect due to unequal weighting increased to 1.31. We also investigated the need for trimming and found no extreme outlier weights (the ratio of maximum weight to the mean weight is well below 3), so did not pursue trimming. The raking step did not change the design effect (remains 1.31) or generate any influential outlier weights.

The additional steps involved in creating the analytic weight (nonresponse adjustment and raking) had minimal effect on the estimated overall design effect (remains 1.31 for the 8,077 sample) and did not introduce any influential outlier weights.

6. Use of the Tracker vs. Analytic Weight

When using the tracker weight, respondents are weighted up to represent all Medicare beneficiaries ages 65 and older who were alive on September 30, 2010 and residing in the contiguous United States. In contrast, the analytic weight reproduces only those alive and eligible for NHATS during the fieldwork period, which began May 2011.

The only other difference between the two sets of weights is the treatment respondents who live in residential care settings other than nursing homes. In cases where an FQ interview was completed but an SP interview was not completed, a positive weight sits in the tracker file and a zero weight in the analytic file. The analytic weights of individuals with both an SP and FQ interview have been adjusted to represent these cases (persons assigned both an SP and FQ interview but with only an FQ). For all other respondents the analytic and tracker weights are equal.

Most often analyses will use the analytic weight. The tracker weight is appropriate for making national estimates using the FQ information (e.g. for services available to older adults living in residential care settings) and for investigating the role of mortality on Round 1 disability estimates and successive cross-sections.

7. Variance Estimation

Two broad classes of methods have been developed for computation of standard errors of estimates from complex sample surveys: (1) Taylor series linearization and (2) replication methods. The NHATS data files contain the information necessary for analysts to use either of these approaches to compute standard errors.

The Taylor series linearization approach uses a mathematical technique to approximate a nonlinear statistic with a linear form. The variance of the nonlinear statistic is approximated by the variance of the linear function. Software packages that use Taylor series linearization to estimate variances of statistics from complex sample surveys require the user to specify design information including “stratum” and “cluster” variables. Unlike those based on replication methods, variance estimates based on Taylor series linearization do not fully capture the effects of all of the weight adjustments; however, in most

cases, the differences between Taylor series standard errors and replication-based standard errors are not large. The “stratum” and “cluster” variables that allow users to compute variance estimates using Taylor series linearization are provided on the NHATS tracker and SP files as the variables w1varstrat and w1varunit, respectively.

Replication methods provide a relatively simple way of estimating variances and have some advantages over other variance estimation methods (e.g., linearization approach). The basic idea behind the replication approach is to identify subsets of the full sample (“replicates”), to calculate the estimate of interest for each replicate, and then to use the variability among these replicate estimates to estimate the variance of the full sample statistics. Different approaches can be used to create these replicates.

For NHATS, the replication approach that was used is the modified balanced repeated replication (BRR) method suggested by Fay (Judkins 1990). When estimating the variance of ratios of rare subsets, one problem that occasionally arises from standard BRR is that one or more replicate estimates will be undefined due to zero denominators. Instead of increasing the weights of one half-sample by 100 percent and decreasing the weights of the other half-sample to zero as in standard BRR, Fay’s method perturbs the weights by $\pm 100(1-K)$ percent where K is referred to as “Fay’s factor.” The perturbation factor for standard BRR is 100 percent, or $K=0$. For NHATS, $K = 0.3$ was used.

In order to create the replicates (and also to define the variables needed for Taylor series linearization variance estimates), it is necessary to define variance strata and variance units. Each first-stage noncertainty stratum (i.e., each noncertainty stratum defined for PSU selection) constituted a separate variance stratum. Noncertainty PSUs were selected in pairs with two PSUs from each stratum, so the first PSU in the stratum will form the first variance unit and the second PSU will form the second variance unit. The remaining variance strata and variance units were formed by combining the second-stage sampling units (ZIP clusters) within certainty PSUs. Each resulting variance stratum contained two variance units.

Nonresponse adjustment and raking were repeated for each of the replicates. The final tracker replicate weights are provided in the variables w1trfinwgt1-w1trfinwgt56, and the analytic replicate weights are provided in the variables w1anfinwgt1-w1anfinwgt56. Through the creation of person-level replicate weights, Fay’s method approximately reflects the contribution of variance due to nonresponse adjustments, calibration adjustments (e.g., poststratification or raking), and other weight adjustment factors that are dependent on the observed sample.

References

Judkins DR. (1990). Fay's method for variance estimation. *Journal of Official Statistics*, 6(3), 223-239.

Kalton G & Flores Cervantes I. (2003). Weighting methods. *Journal of Official Statistics*, 19(2), 81-97.

Kish L. (1965). *Survey sampling*. New York: John Wiley and Sons.

Montaquila J, Freedman VA, Edwards, B, & Kasper JD. 2012. *National Health and Aging Trends Study Round 1 Sample Design and Selection. NHATS Technical Paper #1*. Baltimore: Johns Hopkins University School of Public Health. Available at www.NHATS.org.

Appendix A. Description of Residential Setting Indicators Used for Round 1 NHATS Weighting

This appendix describes and provides results for two variables that were used in the weighting process as proxies for residential setting: (1) an indicator of whether administrative records from the Center for Medicare and Medicaid Services' Minimum Data Set (MDS) suggested the case was a non-acute nursing home stay and (2) an indicator of whether the sampled person's address in the Interview Management System (IMS) could be matched to a list of addresses of assisted living facilities compiled by the "Shaping Long Term Care in America" Assisted Living Census: 2009. (Hereafter, this list is referred to as the "assisted living list.") The assisted living list contained addresses of all licensed, certified or approved Assisted Living vendors of either the Medicaid agency or other state agencies responsible for residential services for the aged.

1. MDS Match

We matched all cases from the NHATS sample other than the ineligible to MDS records for the period January 2011–October 2011. Our aim was to develop an algorithm that was consistent with our data collection protocol to identify residents of nursing home facilities and to distinguish them from individuals having a temporary skilled nursing facility (SNF) stay.

Section A of the MDS 3.0 has several items that describe reason for assessment. We used combinations of items A0310B (PPS Assessment), A0310F (Entry/Discharge Reporting) and A2100 (Discharge status), A1600 (Entry date) and A2000 (Exit date) to identify cases with "short-term" stays and remaining cases with an MDS were considered to be consistent with our definition of "resident."

A0310B. Prospective Payment System (PPS) Assessment	A0310F. Entry/Discharge Reporting	A2100. Discharge status
1. 5-day	1. Entry	1. Community (private home/apt, board/care, assisted living, group home)
2. 14-day	10. Discharge-return not anticipated	2. Another nursing home or swing bed
3. 30-day	11. Discharge-return anticipated	3. Acute hospital
4. 60-day	12. Death in Facility	4. Psychiatric hospital
5. 90-day	99. Not entry/discharge	5. Inpatient rehab facility
6. Readmission/return		6. MR/DD facility
7. Unscheduled PPS		7. Hospice
99. Not PPS		8. Deceased
		9. Other

The following three sets of cases were classified as SNF (non-residential) stays:

1. All MDS records have ($1 \leq A0310B \leq 7$) or ($A0310B=99$ and $A0310F=(1, 10 \text{ or } 11)$) (PPS alone, entry/discharge alone, or combination of PPS with entry/discharge)
2. Any MDS record $A0310F = 10$ (discharge return not anticipated) and $A2000$ (discharge date) between Jan and May 1 2011 (start of data collection)
3. If no record with $A0310F=1$ (entry record) with entry date after last discharge date, where discharge date ($A2000$) is for latest MDS record with $A0310F=10$ (discharge return not anticipated) and $A2100 = 1$ (discharge to community) (discharged to community and not readmitted before Oct 2011)

Table A1 provides results of the match and application of this algorithm to MDS records by NHATS Round 1 disposition codes for the NHATS sample excluding ineligible. Using the algorithm described above, MDS records suggested a pattern consistent with nursing home residence for 538 SPs. Another 460 had one or more MDS record that suggested the case was a short-term (PPS) nursing home stay.

The sensitivity and specificity of the MDS-based residence measure is very high. 423 out of the 468 cases in NHATS that were considered nursing home residents were identified as such using this MDS methodology (90.4%). The majority of misclassified cases ($n=31$) did not match any MDS record and another 14 were identified using this algorithm as a short stay. Out of the 7,777 cases thought to be not nursing home residents according to NHATS, all but 0.7% (all but 58) were correctly classified.

Using this matched indicator to partition nonrespondents, we are able to calculate response rates for those thought to be nursing home residents (see last row of Table A1). We found that the response rate was 89.4% for nursing home residents in NHATS Round 1 (vs. 70.1% for the rest of the eligible cases), a difference large enough to lead us to impose nursing home residence as the first branch in the classification tree analysis for nonresponse adjustment.

Table A1. Distribution of final disposition by available MDS for the NHATS sample excluding ineligible

Final Round 1 NHATS Disposition	NHATS-MDS Match, (Jan-Oct 2011 MDS)			Total
	No MDS Match	MDS suggests nursing home resident*	MDS suggests short-term stay	
Respondents:				
60 Complete (not nursing home)	7,206	52	292	7,550
61 Complete, Nursing home resident	31	423	14	468
63 Residential care, not nursing home, FQ not complete	57	0	2	59
64 Residential care, not nursing home, SP not complete	137	6	25	168
Nonrespondents:				
75 Final, physically/mentally unable to participate, no proxy	57	3	7	67
76 Final too ill to participate, no proxy	69	2	12	83
77 Final refusal, SP	2,641	18	74	2,733
78 Final language barrier	61	0	0	61
79 Final unable to locate	175	20	10	205
80 Final unavailable during field period	12	0	1	13
82 Final outside of PSU	67	7	4	78
85 Final refusal, facility	2	3	1	6
87 Final refusal, proxy	120	4	18	142
88 Final work stopped	4	0	0	4
Total NHATS sample excluding ineligible	10,639	538	460	11,637
Response rate	69.8%	89.4%	72.4%	70.9%

2. Assisted Living List Match

A list of assisted living facilities in the US was compiled for 2009 by the “Shaping Long Term Care in America Project” at Brown University funded in part by the National Institute on Aging (1P01AG027296).” The project obtained names and addresses of all licensed, certified or approved vendors of either the Medicaid agency or other state agencies responsible for residential services for the aged. Whenever possible, excel spreadsheets or data bases were downloaded or provided directly by states. In some cases researchers converted files or manually re-entered the names and addresses.

The IMS address is the most current address available for the SP prior to contact. It is initialized with the SP’s address from the Medicare EDB, but may be updated as a result of information obtained while attempting to contact the SP. (The IMS address does not, however, contain updates based on the address reported during the SP interview.)

The IMS address was matched to the assisted living list using the AutoMatch software for probabilistic matching. Address matching is prone to error due to inconsistencies in the formatting and specification of address fields, and due to typographical errors. Rather than requiring an exact match of all address fields, probabilistic matching was used to identify combinations of IMS addresses and addresses from the assisted living list that had high probabilities of being true matches, based on specified parameters.

The probabilistic matching process was done in three “passes”. The parameters that drove the identification of likely matches were specified most stringently for pass 1 and were successively loosened for passes 2 and 3. Following manual review of output from each of the passes, it was determined that the matches identified on passes 1 and 2 could generally be considered to be valid matches, but the matches identified only on pass 3 should not be regarded as true matches.

Table A2. Results of Round 1 NHATS match to assisted living list for the NHATS sample excluding ineligible

Final Disposition	Number of Cases, by Status of Match to Assisted Living List			
	No match	Match, Passes 1 or 2	Match, Pass 3	Total
Respondents				
60 Complete (not nursing home)	7,435	96	19	7,550
61 Complete, Nursing home resident	428	34	6	468
63 Residential care, not nursing home, FQ not complete	54	5	0	59
64 Residential care, not nursing home, SP not complete	115	49	4	168
Nonrespondents				
75 Final, physically/mentally incompetent, no proxy	66	1	0	67
76 Final too ill, no proxy	83	0	0	83
77 Final refusal, SP	2,724	8	1	2,733
78 Final language barrier	61	0	0	61
79 Final unable to locate	205	0	0	205
80 Final unavailable during field period	13	0	0	13
82 Final outside of PSU	78	0	0	78
85 Final refusal, facility	5	1	0	6
87 Final refusal, proxy	141	1	0	142
88 Final work stopped	4	0	0	4
Total NHATS sample excluding ineligible	11,412	195	30	11,637
Response rate	70.4%	94.4%	96.7%	70.9%

Table A2 provides a summary of results of the address matching process. A total of 195 cases (about 2 percent of fielded cases) were matched to the assisted living list based on the IMS address in passes 1 and 2. An additional 30 cases matched with the looser criteria (3rd pass).

The field codes do not allow direct assessment of the sensitivity and specificity of this match. We therefore developed a classification of the type of facility based on responses to select items from the facility questionnaire (FQ). For respondents with a final fieldwork disposition code of 60 (complete) who completed an FQ and those with a final code of 64 (completed an FQ only), we classified respondents as living in assisted living, special care unit, nursing home, independent living, or unknown based on place type responses to FQ6 (with upcodes from other specify) and FQ10.

For FQ6, the interviewer hands the staff person a show card and asks: “Which of these BEST describes {FACILITY NAME}? Again, if there are different parts or levels of care in this place, please tell me about the whole place.” Responses are:

1. Free-standing nursing home
2. Free-standing assisted living facility
3. Nursing home and assisted living facility
4. Continuing care retirement facility (CCRC)
5. Adult family care home
6. Group home
7. Board and care home
8. Retirement community or senior housing (not CCRC)
91. Other (Specify)

FQ10 is asked of respondents who answer 2-4, 8 or 91. “Is the place where {SP} lives considered independent living, assisted living, a special care unit, a nursing home care unit, or something else?”

- 1 Independent living
- 2 Assisted living
- 3 Special care, memory care, or Alzheimers unit
- 4 Nursing home
- 91 Other (SPECIFY)

Respondents who were living in a free-standing assisted living facility, an adult family care home, a group home, a board and care home, or in the assisted living part of a CCRC or of a combined nursing home/assisted living facility were classified as assisted living. Among the 228 cases identified as such, the IMS address was matched to the assisted living list on passes 1 or 2 for 85 cases (37%) (see Table A3). Among 39 cases identified as living in a special care unit, 7 were matched (18%). Among those in all other settings including nursing home settings, independent living, unknown facilities, and the community 92 (1%) matched.

Table A3. Results of Round 1 NHATS match to assisted living list, by facility type

Facility Type	Number of Cases, by Status of Match to Assisted Living List			
	No match	Match, Passes 1 or 2	Match, Pass 3	Total
Assisted Living	127	85	16	228
Special care/memory care/Alzheimers unit	32	7	0	39
Nursing home	421	31	8	460
Independent living/other	194	43	1	238
Facility type not reported	71	11	1	83
Not in facility	7,187	7	3	7,197
Total	8,032	184	29	8,245

Appendix B: Variables Used in Nonresponse Adjustment for Round 1 NHATS Weights

Appendix Table B1. Response Rates by Various Frame Indicators: NHATS Round 1

Variable & Values	Weighted Response Rate	Variable & Values	Weighted Response Rate
BENEFICIARY INDICATORS		TRACT-LEVEL INDICATORS (Quartiles)	
Age¹ *	(H_AGECAT)	Household Income³ *	(C_AGG_HH_INC)
65-69	72.4%	1	76.0%
70-74	70.2%	2	72.5%
75-79	69.5%	3	70.6%
80-84	69.5%	4	69.1%
85- 89	75.3%	Median Household Income³ *	(C_MED_HH_INC)
90+	75.7%	1	76.7%
Enhanced Race Indicator¹ *	(H_ENHRACEETH)	2	72.2%
Non-Hispanic Black	76.4%	3	69.4%
Hispanic	75.5%	4	68.1%
White/Other	70.5%	Missing	75.6%
		Median Household Income 65+³ *	(C_MED_HH_INC_65)
Gender¹	(H_SEX)	1	75.4%
Male	70.8%	2	71.1%
Female	71.6%	3	70.3%
Census Region¹	(S_REGION)	4	69.1%
Northeast	67.2%	Missing	76.6%
Midwest	72.7%	% Households with Adult 65+³ *	(C_PCT_HH_65)
South	72.0%	1	70.3%
West	72.4%	2	71.7%
Census Division¹ *^	(DIVISION)	3	72.1%
New England	61.3%	4	70.8%
Middle Atlantic	69.8%	% Households in Poverty³ *	(C_PCT_HH_POV)
East North Central	69.1%	1	70.4%
West North Central	78.0%	2	71.2%
South Atlantic	70.4%	3	76.0%
East South Central	75.1%	4	71.3%
West South Central	73.1%	% Households Reporting Public Assistance³ *	(C_PCT_HH_PUBASST)
Mountain	76.8%	1	69.2%
Pacific	71.8%	2	70.3%
Census Metro/Micro Area Designation (2008)¹ ^	(METMICRO)	3	72.3%
Metropolitan area	70.1%	4	73.7%
Micropolitan area	73.7%	% Households Reporting Retirement Income³ *	(C_PCT_HH_RETIREINC)
Non-metro	80.8%	1	71.5%
Health Maintenance Organization Beneficiary¹ *	(HMOTYPE)	2	71.9%
Yes	73.3%	3	71.7%
No	70.5%	4	70.3%
Age First Enrolled in Medicare¹ *	(MEDIC_BEG)	% Households Reporting Social Security³	(C_PCT_HH_SOCSEC)
Prior to age 65	77.0%	1	69.8%
At or after age 65	70.7%	2	71.2%
Replicate Release Group	(RLSGRP2)	3	71.2%
Release 1 (May 2011)	71.4%	4	72.1%
Release 2 (Sept 2011)	68.8%	% Households Reporting SSI³	(C_PCT_HH_SSS)
COUNTY LEVEL INDICATORS		1	70.5%
% Black 65+ (deciles)² *	(PCTBLK)	2	70.2%

0	81.7%	3	71.5%
1	70.1%	4	73.1%
2	69.6%	% Households Owning Their Home³	
3	70.3%	(C_PCT_OWNSHOME)	
4	71.5%	1	71.6%
5	69.7%	2	71.8%
6	68.8%	3	72.7%
7	65.4%	4	69.4%
8	69.7%	% Households 65+ Owning Their Home³*	
9	72.1%	(C_PCT_OWNSHOME_65)	
% Hispanic 65+ (deciles)² * ^	(PCTHISP)	1	70.1%
0	73.6%	2	73.2%
1	73.5%	3	71.5%
2	74.8%	4	70.2%
3	74.0%	% Households 65+ Below Poverty³ *	
4	70.5%	(C_PCT_POV_65)	
5	70.6%	1	69.9%
6	68.3%	2	70.3%
7	67.8%	3	71.7%
8	69.9%	4	72.9%
9	68.9%	Per Capita Income³	
% Poverty (deciles)² ^	(POVERTY_PCT)	(C_PER_CAP_INC)	
0	73.3%	1	75.0%
1	66.7%	2	72.5%
2	69.9%	3	70.9%
3	75.7%	4	68.0%
4	67.1%	Missing	71.4%
5	73.1%	OTHER INDICATORS	
6	68.4%	MDS Match Algorithm Indicator (MDSMATCH)	
7	70.5%	NH Resident 92.8%	
8	74.6%	Not NH Resident 70.6%	
9	75.2%	Licensed Assisted Living Match Indicator	
		(ALADDRMATCH)	
		AL Resident 94.5%	
		Not AL Resident 71.0%	

¹Based on Information on the September 30, 2010 CMS 20% Health Insurance Skeleton Eligibility Write-Off (HISKEW) file.

²Based on county-level information from the CMS 5% HISKEW File linked to the beneficiary's EDB address

³Based on tract-level information from the 2006-2019 5-year American Community Survey file linked to the beneficiary's EDB address

*=retained in classification tree analysis for non-nursing home branch

^=retained in classification tree analysis for nursing home branch

N=11,637 (8245 respondents and 3392 non-respondents)

Variable names used in classification trees shown parenthetically.

Appendix Table B2. Sampled Person Interview Response Rates Among Cases with Completed Facility Questionnaires, by Various Indicators: NHATS Round 1

Variable & Values		Weighted Response Rate	Variable & Values		Weighted Response Rate
OVERALL		67.8%			
BENEFICIARY INDICATORS			COUNTY LEVEL INDICATORS		
Age^{1*}	(H_AGECAT)		% Black 65+ (deciles)²	(PCTBLK)	
65-69		87.7%	0		67.2%
70-74		78.0%	1		70.5%
75-79		75.6%	2		65.1%
80-84		69.7%	3		62.1%
85- 89		53.5%	4		63.6%
90+		67.8%	5		72.0%
Enhanced Race Indicator¹	(H_ENHRACEETH)		6		61.2%
Non-Hispanic Black		87.6%	7		77.5%
Hispanic		90.8%	8		58.6%
White/Other		66.4%	9		90.7%
Gender¹	(H_SEX)		%Hispanic 65+ (deciles)^{2*}	(PCTHISP)	
Male		74.2%	0		66.1%
Female		65.4%	1		71.6%
Census Region^{1*}	(S_REGION)		2		79.1%
Northeast		66.0%	3		59.1%
Midwest		72.3%	4		72.4%
South		66.1%	5		63.0%
West		66.9%	6		67.2%
Census Division^{1*}	(DIVISION)		7		67.9%
New England		80.0%	8		68.4%
Middle Atlantic		62.2%	9		66.3%
East North Central		67.8%	% Poverty (deciles)²	(POVERTY_PCT)	
West North Central		76.8%	0		52.2%
South Atlantic		66.8%	1		68.1%
East South Central		71.9%	2		67.3%
West South Central		60.9%	3		75.3%
Mountain		83.1%	4		75.4%
Pacific		62.3%	5		65.2%
Census Metro/Micro Area Designation (2008)¹	(METMICRO)		6		64.7%
Metropolitan area		67.2%	7		70.5%
Micropolitan area		69.3%	8		68.1%
Non-metro		72.5%	9		78.9%
Health Maintenance Organization Beneficiary¹	(HMOTYPE)		OTHER INDICATOR		
Yes		72.4%	Facility Type Indicator^{3*} (FQ1DLOCSP)		
No		66.4%	Independent living/other		79.6%
Age First Enrolled in Medicare¹	(MEDIC_BEG)		Assisted Living		58.6%
Prior to age 65		70.1%	Special care/memory care/Alzheimers unit		38.1%
At or after age 65		67.5%	Facility type not reported		49.0%
Replicate Release Group	(RLSGRP2)				
Release 1 (May 2011)		68.0%			
Release 2 (Sept 2011)		62.4%			

¹Based on Information on the September 30, 2010 CMS 20% Health Insurance Skeleton Eligibility Write-Off (HISKEW) file.

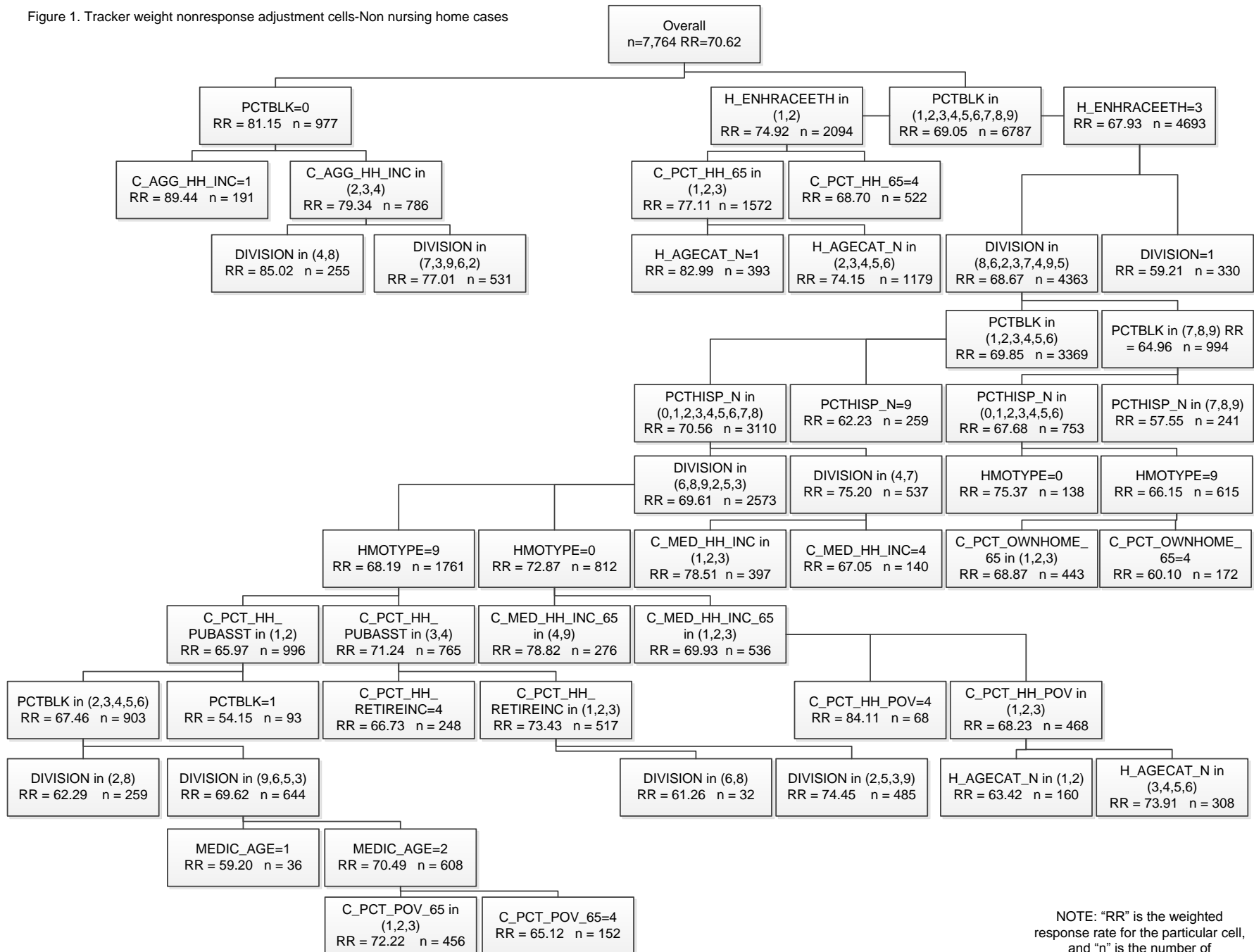
²Based on county-level information from the CMS 5% HISKEW File linked to the beneficiary's EDB address

³Indicator (fq1dlocsp) is based on items from the FQ: FQ6 (fq1facdescri); FQ6A (fq1dosfacd); FQ10 (fq1faaretype); FQ10A (fq1dosfaca). All variables are available on the SP file.

*=retained in classification tree analysis for adjustment of missing SP interview

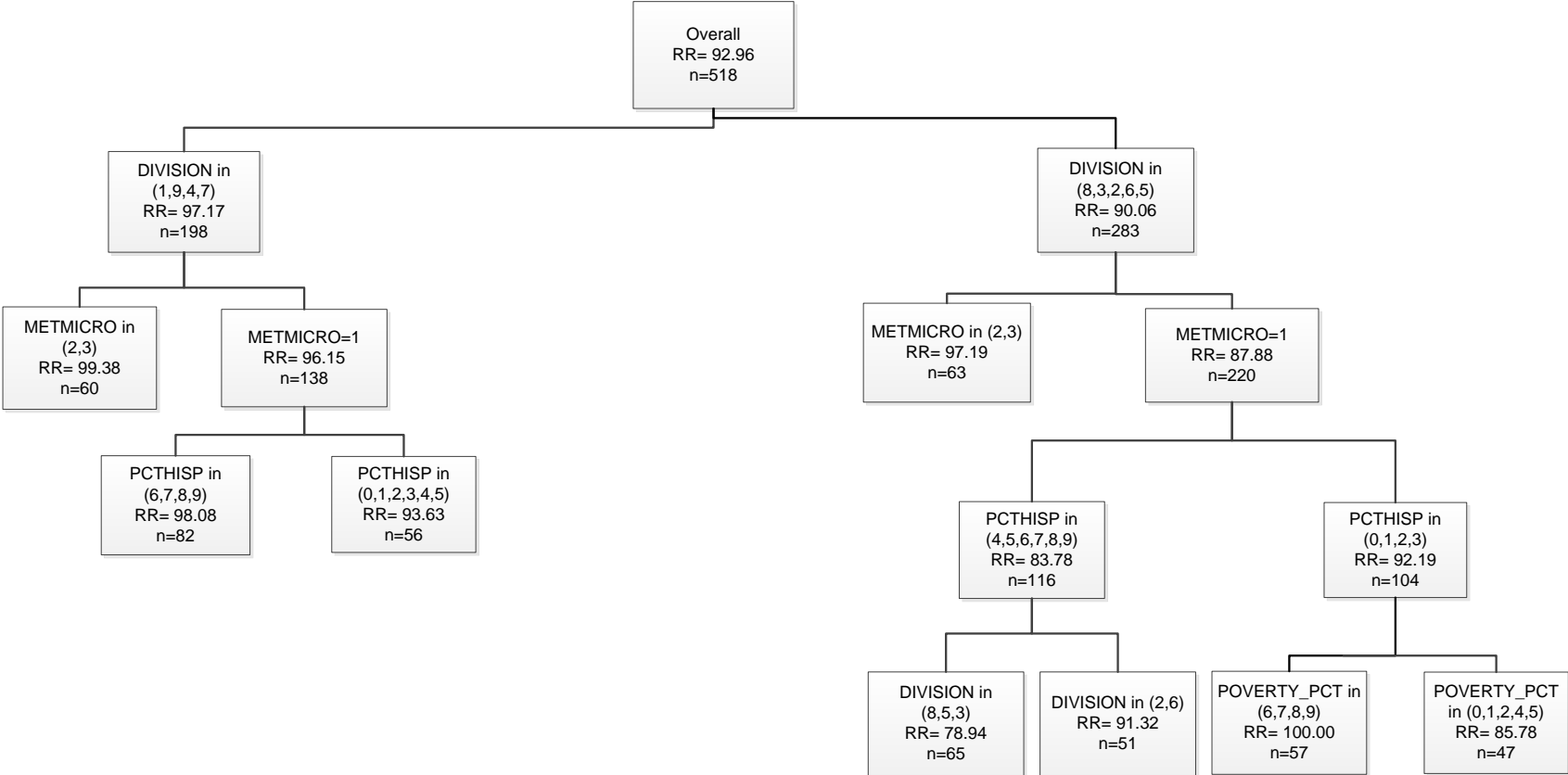
N=521 (323 respondents and 168 nonrespondents); Variable names used in classification trees shown parenthetically.

Figure 1. Tracker weight nonresponse adjustment cells-Non nursing home cases



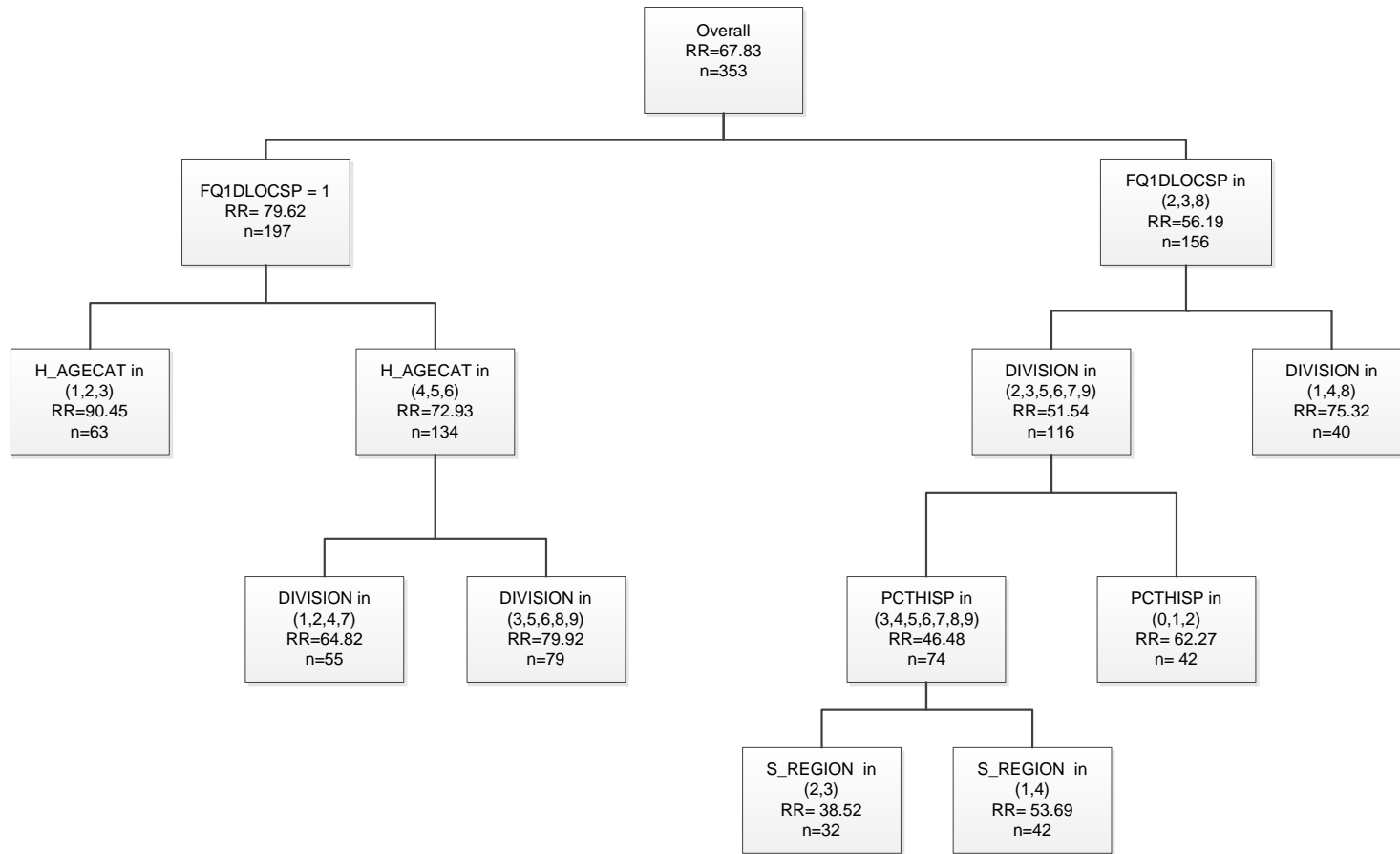
NOTE: "RR" is the weighted response rate for the particular cell, and "n" is the number of respondents in the cell.

Figure 2: Tracker weight nonresponse adjustment cells-Nursing home cases



Note: "RR" is the weighted response rate for the particular cell, and "n" is the number of respondents in the cell.

Figure 3: Analytic weight nonresponse adjustment cells-SPs in residential care only



NOTE: "RR" is the weighted response rate for the particular cell, and "n" is the number of respondents in the cell.