NHATS Technical Paper #29

NATIONAL HEALTH AND AGING TRENDS STUDY (NHATS)
COVID-19 Sample Person (SP) and Family Members and Friends (FF) Supplemental Survey Weights

August 2021
Release 1.0

Suggested Citation: Hu, Mengyao, Freedman, Vicki A., Ferraro, David, Jiao, Rui, and DeMatteis, Jill. 2021. National Health and Aging Trends Study COVID-19 Sample Person (SP) and Family Members and Friends (FF) Supplemental Survey Weights. NHATS Technical Paper #29 Baltimore: Johns Hopkins University School of Public Health. Available at www.NHATS.org. This technical paper was prepared with funding from the National Institute on Aging (U01AG032947).

NHATS COVID-19 Survey Design and Need for Weights

The NHATS COVID-19 Sample Persons (SP) file is designed to represent Medicare beneficiaries ages 70 and older living in the contiguous United States in 2020.

The NHATS COVID-19 Family Members and Friends (FF) file is designed to represent family and unpaid caregivers to a representative sample of living Medicare beneficiaries ages 70 and older in 2020.

In order to be able to make statements about population estimates for these groups, the COVID-19 SP and FF samples must be weighted to account for different probabilities of selection and response to NHATS as well as differential probabilities of response to the COVID-19 SP and FF supplements. In order to make accurate statements about the variance of those estimates, design variables must also be used to account for the complex design of NHATS.

This technical paper describes the development of the NHATS COVID-19 weights and design variables for the NHATS COVID-19 SP and FF Files.

Overview of Weight and Design Variables

Each COVID-19 data file contains a set of weights (full sample weight, replicate weights) adjusted for nonresponse and design variables (stratum, cluster) that allow for proper variance estimation.

File	Full sample weight	Replicate weights	Stratum	Cluster
COVID-19 SP	w10cvspfinwgt0	w10cvspfinwgt1-w10cvspfinwgt56	w10varstrat	w10varunit
COVID-19 FF	w10cvfffinwgt0	w10cvfffinwgt1-w10cvfffinwgt56	w10varstrat	w10varunit

When the unit of analysis is the sample person, SP weights should be used. When the unit of analysis is caregiver, FF weights should be used. The design variables (stratum and cluster) should be specified when using software that uses Taylor series linearization to estimate the variances of estimates from complex sample surveys.

Replicate weights are also provided and may be used with software that uses replication methods to estimate the variances of estimates from complex sample surveys. The replication approach that was used is the modified balanced repeated replication (BRR) method suggested by Fay (Judkins 1990). Fay's method perturbs the weights by ±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, NSOC and COVID-19 supplements, K = 0.3 was used.

How to Use Full Sample Weights and Design Variables

Stata Commands. In Stata, users should specify the following svyset command for COVID-19 SP and FF files.

```
/*COVID-19 SP file*/
        svyset w10varunit [pweight=w10cvspfinwgt0], strata(w10varstrat)
        svy: [stata procedures]
/*COVID-19 FF file*/
        svyset w10varunit [pweight=w10cvfffinwgt0], strata(w10varstrat)
        svy: [stata procedures]
```

SAS Commands. In SAS, users should specify command for COVID-19 SP and FF files.

```
/*COVID-19 SP file*/
        [sas procedure]
        weight w10cvspfinwgt0;
        cluster w10varunit;
        strata w10varstrat;
        [model or other statement]
        run;
/*COVID-19 FF file*/
        [sas procedure]
        weight w10cvfffinwgt0;
        cluster w10varunit;
        strata w10varstrat;
        [model or other statement]
        run;
```

For more information about how to account for sample design in NHATS, please refer to "Accounting for Sample Design in NHATS and NSOC Analyses: Frequently Asked Questions" (Freedman et al. 2020), available at www.NHATS.org.

Calculation of Weights

COVID-19 SP weights

The computation of the COVID-19 SP weights begins with the final Round 10 NHATS analytic weight for the 2015 cohort (DeMatteis et al. 2021). The final Round 10 analytic weight had undergone a raking adjustment to align weighted totals with the NHATS control totals for the 2015 Cohort.

A single stage nonresponse adjustment was then made to the Round 10 final analytic weight. In the adjustment, a set of variables were input into a classification tree analysis to determine which variables were associated with nonresponse. This approach uses a classification tree algorithm to identify variables associated with response propensities. 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 set a minimum cell size of 30 respondents.

Weighted distributions for demographic variables using the nonresponse adjusted COVID-19 SP weight resulted in similar distributions based on the Round 10 analytic weight. Consequently, we did not opt to incorporate an additional raking adjustment for the COVID-19 SP weight.

COVID-19 FF weights

Because completion of the COVID-19 FF questionnaire was contingent on the SP having completed the COVID-19 SP questionnaire, the calculation of the COVID-19 FF weights began with the NHATS COVID-19 SP weight. A 2-step weighting class adjustment for nonresponse was applied because there were two opportunities for FF nonresponse with potentially different mechanisms—1) the NHATS Sample Person who completed the COVID-19 booklet reported having helpers that they did not list or did not provide sufficient contact information for the listed FF, and 2) non-response at the FF level (for FF for whom the COVID-19 SP respondents provided contact information).

At each stage of adjustment, similar to non-response adjustments for COVID-19 SP weights, a set of variables were input into a classification tree analysis to determine which variables were associated with nonresponse. 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." The same minimum cell size of 30 respondents was set.

Because the caregivers providing the most support during COVID-19 were not always identified by the SP, and basic demographic information was sometimes missing for caregivers who were identified, no raking adjustment was applied in computing the COVID-19 FF weights.

COVID-19 Sample Person (SP) Survey Nonresponse Adjustments

Appendix Table 1 shows variables considered for the adjustments in COVID-19 SP data, along with weighted response rates for each level of each variable. Final non-response cells included 16 indicators (indicated in Appendix Table 1 with an "a"). Combinations of these variables created 24 unique nonresponse cells (see Appendix Figure 1).

The design effect due to variation in the COVID-19 SP base weights (before nonresponse adjustment) was 1.64. The design effect due to variation in the COVID-19 SP analytic weights (after nonresponse adjustment was 1.63. The steps involved in creating the COVID-19 SP weight did not introduce any influential outlier weights.

COVID-19 Family Members and Friends (FF) Survey Nonresponse Adjustments

Appendix Table 2 shows variables considered for the first and second adjustments in COVID-19 FF data for caregivers to SPs, along with weighted response rates for each level of each variable. Final non-response cells included a total of 12 indicators (indicated in Appendix Table 2 with an "a" for the first step and a "b" for the second step). Combinations of these variables created 26 unique nonresponse cells for the first adjustment and 22 nonresponse cells for the second adjustment (see Appendix Figures 2 and 3).

The design effect due to variation in the COVID-19 FF base weights (before nonresponse adjustment) was 1.63. The design effect due to variation in the COVID-19 FF analytic weights (after nonresponse adjustment) was 1.81.

The steps involved in creating the COVID-19 FF weight did not introduce any influential outlier weights (using 10% as a cutoff). There was one case identified that accounts for 9% of the weighted sum of the domain "relationship to SP"; users may choose to evaluate outlier weights for their own applications and trim as needed.

REFERENCES

DeMatteis, Jill M., Freedman, Vicki A., Jiao, Rui, and Kasper, Judith D. 2021. National Health and Aging Trends Study Development of Round 10 Survey Weights. NHATS Technical Paper #28. Baltimore: Johns Hopkins University School of Public Health. Available at www.NHATS.org.

Freedman, V. A., Hu, M., DeMatteis, J., & Kasper, J. D. 2020. Accounting for sample design in NHATS and NSOC analyses: Frequently asked questions. NHATS Technical Paper #23. Baltimore: Johns Hopkins University School of Public Health. Available at www.NHATS.org.

Table 1. Weighted Responses Rates for Variables used in Nonresponse Adjustment for COVID-19 Sample Person Weights

	Completion of the Covid booklet (Weighted Response Rate)
OVERALL	84.7
Age - (H_AGECAT_R5) [1]	
1: 65-69	85.8
2: 70-74	86.4
3: 75-79	83.5
4: 80-84	80.3
5: 85- 89	81.6
6: 90+	75.3
Gender – (H_SEX) [1]	
1: Male	85.3
2: Female	84.1
Census Region – (S_REGION)	
1: Northeast	82.3
2: Midwest	87.6
3: South	84.6
4: West	83.7
Census Division – (DIVISION) [a]	
1: New England	84.2
2: Middle Atlantic	81.4
3: East North Central	87.0
4: West North Central	88.5
5: South Atlantic	85.1
6: East South Central	80.3
7: West South Central	86.1
8: Mountain	85.0
9: Pacific	83.5
Census Metro/Micro Area Designation (2013) – (S_METMICRO) [a]	
1: Metropolitan area	84.2
2: Micropolitan area	85.4
3: Non-metro	90.2
Health Maintenance Organization Beneficiary – (HMOTYPE) [1]	
0: Yes	84.6
9: No	84.8
Enrollment Age – (ENROLL_AGE) [1]	
1: Before 65	74.0
2: Reaching 65	85.6
R1/R5 SP Race/Ethnicity – (RL5DRACEHISP_R) [a]	
1: White, non-Hispanic	88.7
2: Black, non-Hispanic	69.4
3: Other (Am Indian/Asian/Native Hawaiian /Pacific Islander/other specify), non-Hispanic	73.9
4: Hispanic	65.0
5: DK/RF	77.9

(Continued next page)

	Completion of the Covid booklet (Weighted Response Rate)
R1/R5 SP Education– (ELHIGSTSCHL_R) [a]	
1: DK / RF	72.7
2: Below high school	68.9
3: High school	83.8
4: Above High school	89.1
% Black 65+ (deciles) – (PCTBLK) [2, a]	
0: 1st decile	90.8
1: 2nd decile	81.9
2: 3rd decile	88.3
3: 4th decile	84.1
4: 5th decile	84.1
5: 6th decile	89.4
6: 7th decile	81.8
7: 8th decile	83.6
8: 9th decile	83.4
9: 10th decile	75.8
% Hispanic 65+ (deciles) – (PCTHISP) [2, a]	
0: 1st decile	89.0
1: 2nd decile	84.3
2: 3rd decile	85.0
3: 4th decile	85.6
4: 5th decile	86.0
5: 6th decile	83.3
6: 7th decile	88.4
7: 8th decile	85.8
8: 9th decile	83.1
9: 10th decile	74.5
% Poverty (deciles) – (PCTPOV) [2, a]	
0:1st decile	89.3
1: 2nd decile	87.3
2: 3rd decile	85.1
3: 4th decile	86.0
4: 5th decile	87.5
5: 6th decile	86.6
6: 7th decile	83.6
7: 8th decile	80.0
8: 9th decile	82.0
9: 10th decile	74.2
Household Income – (C_AGG_HH_INC)[3]	
1: 1st quartile	75.7
2: 2nd quartile	83.8
3: 3rd quartile	85.3
4: 4th quartile	88.1

(Continued next page)

	Completion of the Covid booklet (Weighted Response Rate)
Median Household Income – (C_MED_HH_INC) [3]	
1: 1st quartile	77.1
2: 2nd quartile	82.5
3: 3rd quartile	86.9
4: 4th quartile	89.5
Median Household Income 65+ – (C_MED_HH_INC_65) [3]	
1: 1st quartile	78.7
2: 2nd quartile	84.5
3: 3rd quartile	84.4
4: 4th quartile	89.6
% Households with Adult 65+ – (C_PCT_HH_65) [3, a]	
1: 1st quartile	81.1
2: 2nd quartile	82.9
3: 3rd quartile	85.5
4: 4th quartile	87.4
% Households in Poverty– (C_PCT_HH_POV) [3, a]	
1: 1st quartile	90.4
2: 2nd quartile	85.8
3: 3rd quartile	84.2
4: 4th quartile	74.1
% Households Reporting Public Assistance – (C_PCT_HH_PUBASST) [3, a]	
1: 1st quartile	87.6
2: 2nd quartile	86.4
3: 3rd quartile	84.1
4: 4th quartile	79.6
% Households Reporting Retirement Income– (C_PCT_HH_RETIREINC) [3, a]	
1: 1st quartile	76.1
2: 2nd quartile	81.9
3: 3rd quartile	87.7
4: 4th quartile	88.4
% Households Reporting Social Security – (C_PCT_HH_SOCSEC) [3, a]	
1: 1st quartile	82.9
2: 2nd quartile	83.3
3: 3rd quartile	84.8
4: 4th quartile	86.6
% Households Reporting SSI– (C_PCT_HH_SSS) [3, a]	
1: 1st quartile	89.6
2: 2nd quartile	87.0
3: 3rd quartile	84.4
4: 4th quartile	77.1
% Households Owning Their Home – (C_PCT_OWNHOME) [3]	
1: 1st quartile	76.0
2: 2nd quartile	82.6
3: 3rd quartile	86.1
4: 4th quartile	89.1
% Households 65+ Owning Their Home – (C_PCT_OWNHOME_65) [3]	
1: 1st quartile	76.2
2: 2nd quartile	83.9
3: 3rd quartile	86.4
4: 4th quartile	88.8

(Continued next page)

	Completion of the Covid booklet (Weighted Response Rate)
% Households 65+ Below Poverty – (C_PCT_POV_65) [3]	
1: 1st quartile	87.5
2: 2nd quartile	85.1
3: 3rd quartile	85.3
4: 4th quartile	81.1
Per Capita Income – (C_PER_CAP_INC) [3, a]	
1: 1st quartile	74.7
2: 2nd quartile	82.7
3: 3rd quartile	87.0
4: 4th quartile	90.1
SP Residence – (R10DRESID) [a]	
1: Community	85.6
2: Residential care resident, not nursing home (SP interview complete)	79.3
4: Nursing home (SP interview complete)	51.4
SP Booklet type and language – (CV10DBOOKLANG) [a]	
1: Self-administered booklet English	86.8
2: Proxy booklet English	63.9
3: Self-administered booklet Spanish	66.7

[1]Based on Information either on the September 30, 2010 CMS 20% Health Insurance Skeleton Eligibility Write-Off (HISKEW) file if the case is in the original sample, or on the September 30, 2014 CMS 20% Enrollment Database (EDB) extract if the case is in the replenishment sample.

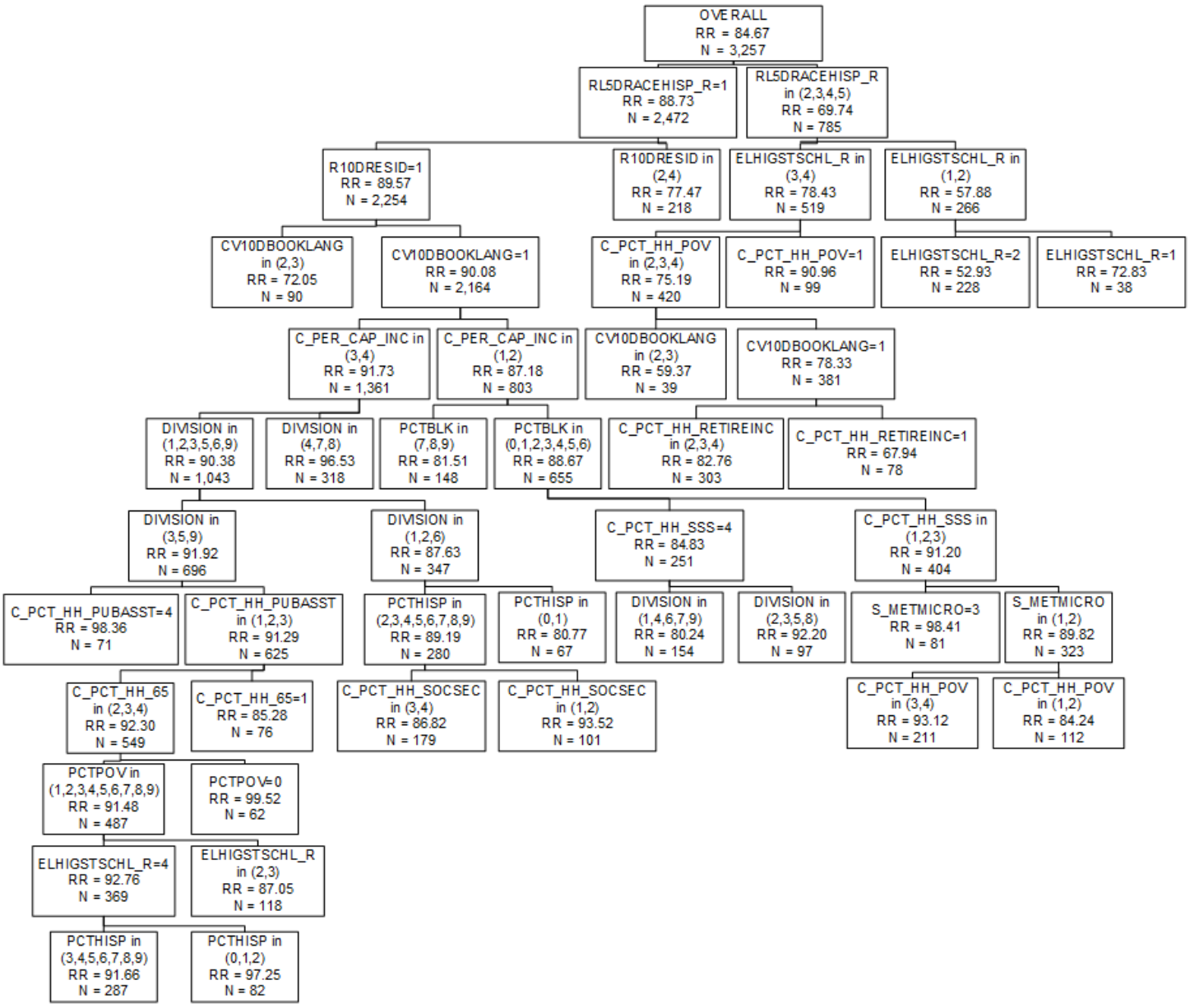
[2]Based on county-level information from the September 30, 2014 CMS 5% EDB extract linked to the beneficiary's EDB address.

[3]Based on tract-level information from the 2009-2013 5-year American Community Survey file linked to the beneficiary's EDB address.

a=retained in classification tree analysis for nonresponse adjustment for SP completion of Covid booklet

Variable names used in classification trees shown in parentheses.

Figure 1. Covid-19 SP Weight nonresponse adjustment cells – SP Refusal



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

Table 2. Weighted Responses Rates for Variables used in Nonresponse Adjustment for Covid-19 Family Members and Friends Weights

	Contact info and consent provided for FF member (Weighted Response Rate)	FF interview complete, given contact info and consent provided (Weighted Response Rate)
OVERALL	86.2	75.3
Age - (H_AGECAT_R5) [1, a, b]		
1: 65-69	85.7	73.7
2: 70-74	85.2	74.2
3: 75-79	89.1	76.3
4: 80-84	88.8	73.3
5: 85- 89	78.9	86.6
6: 90+	87.3	84.4
Gender – (H_SEX) [1, a, b]		
1: Male	84.7	81.2
2: Female	87.1	71.3
Census Region – (S_REGION) [a, b]		
1: Northeast	87.5	71.1
2: Midwest	87.8	76.8
3: South	85.9	77.3
4: West	84.1	73.8
Census Division – (DIVISION) [a, b]		
1: New England	83.7	67.5
2: Middle Atlantic	89.2	72.8
3: East North Central	88.2	81.6
4: West North Central	87.1	68.8
5: South Atlantic	86.7	80.0
6: East South Central	85.3	73.9
7: West South Central	84.9	74.8
8: Mountain	92.8	63.5
9: Pacific	82.9	75.2
Census Metro/Micro Area Designation (2013) – (S_METMICRO) [a]		
1: Metropolitan area	86.1	74.5
2: Micropolitan area	84.1	82.9
3: Non-metro	93.1	70.3
Health Maintenance Organization Beneficiary – (HMOTYPE) [1, a, b]		
0: Yes	84.1	78.2
9: No	88.3	71.0
Enrollment Age – (ENROLL_AGE) [1]		
1: Before 65	83.8	72.8
2: Reaching 65	86.4	75.5
SP Race/Ethnicity – (RL5DRACEHISP_R) [a, b]		
1: White, non-Hispanic	86.9	78.7
2: Black, non-Hispanic	84.8	67.8
3: Other (Am Indian/Asian/Native Hawaiian /Pacific Islander/other specify), non-Hispanic	86.1	70.0
4: Hispanic	85.5	53.0
5: DK/RF	69.4	75.2

(Continued next page)

	Contact info and consent provided for FF member (Weighted Response Rate)	FF Interview complete, given contact info and consent provided (Weighted Response Rate)
SP Education– (ELHIGSTSCHL_R) [a, b]		
1: DK / RF	71.7	75.0
2: Below high school	83.0	67.8
3: High school	88.5	75.5
4: Above high school	86.7	77.3
SP Residence – (R10DRESID) [a]		
1: Community	85.8	74.8
2: Residential care resident, not nursing home (SP interview complete)	87.5	81.3
4: Nursing home (SP interview complete)	98.0	77.8
SP Booklet type and language – (CV10DBOOKLANG) [a, b]		
1: Self-administered booklet English	85.6	75.7
2: Proxy booklet English	94.7	81.9
3: Self-administered booklet Spanish	77.4	46.2
FF Relationship to SP – (FF10DRELSHPCAT) [2, b]		
1: Spouse/partner	NA	90.6
2: Adult child	NA	66.9
3: Other relatives	NA	67.9
4: Non-relatives	NA	78.3
9: Inapplicable	NA	64.0
Whether FF is a proxy respondent – (FF10DCASETYPE) [2, b]		
1: Proxy FF	NA	97.3
2: Non-proxy FF listed first	NA	77.1
3: Non-proxy FF listed second	NA	66.9

[1]Based on Information either on the September 30, 2010 CMS 20% Health Insurance Skeleton Eligibility Write-Off (HISKEW) file if the case is in the original sample, or on the September 30, 2014 CMS 20% Enrollment Database (EDB) extract if the case is in the replenishment sample.

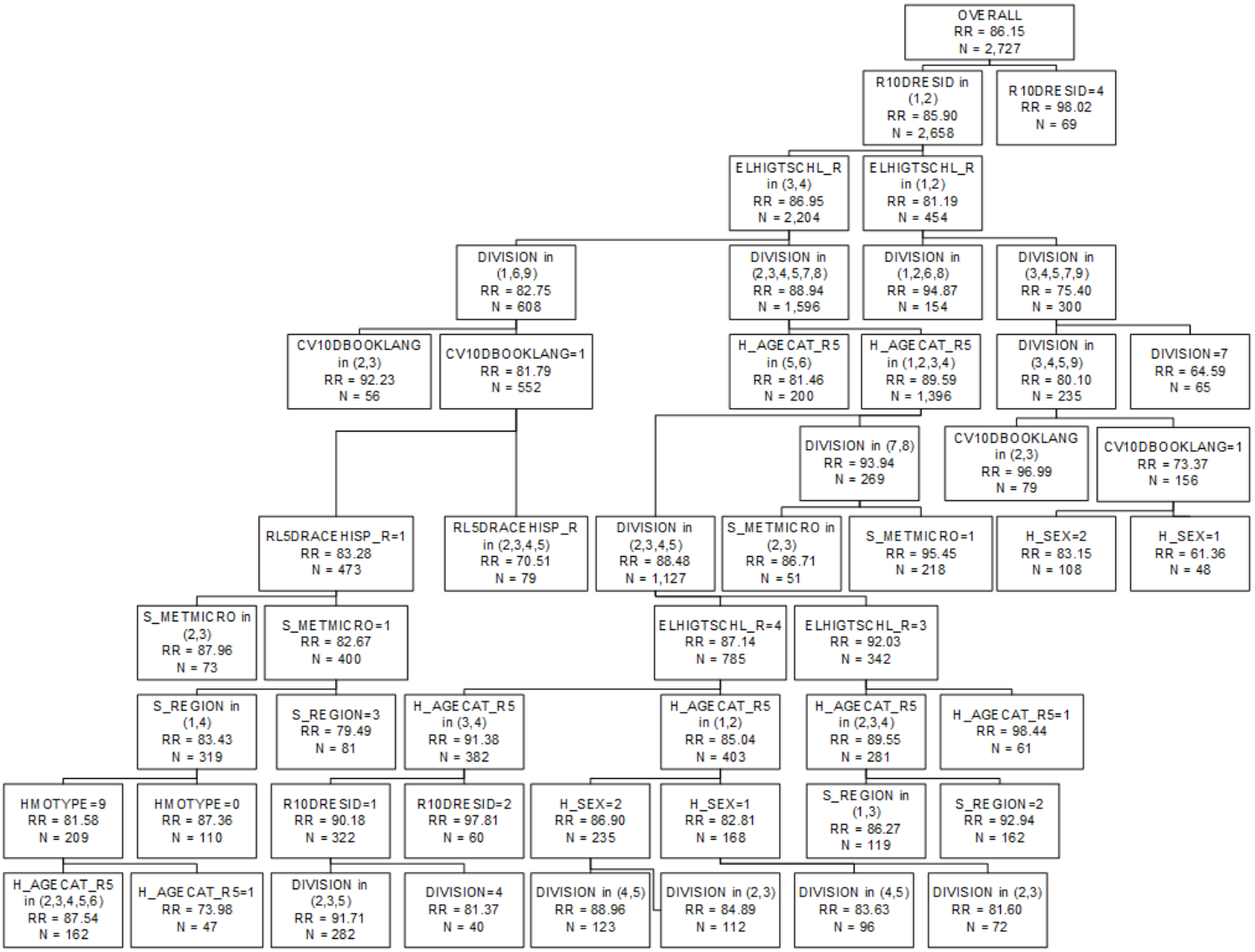
[2]The variable was only used in nonresponse modeling for nonresponse adjustment for nonresponse on the part of a FF member for whom contact information and consent were provided by the SP

a=retained in classification tree analysis for nonresponse adjustment for SP refusal to provide contact information and consent for the FF member

b=retained in classification tree analysis for nonresponse adjustment for FF member failure to complete interview, given that contact information and consent was provided by SP

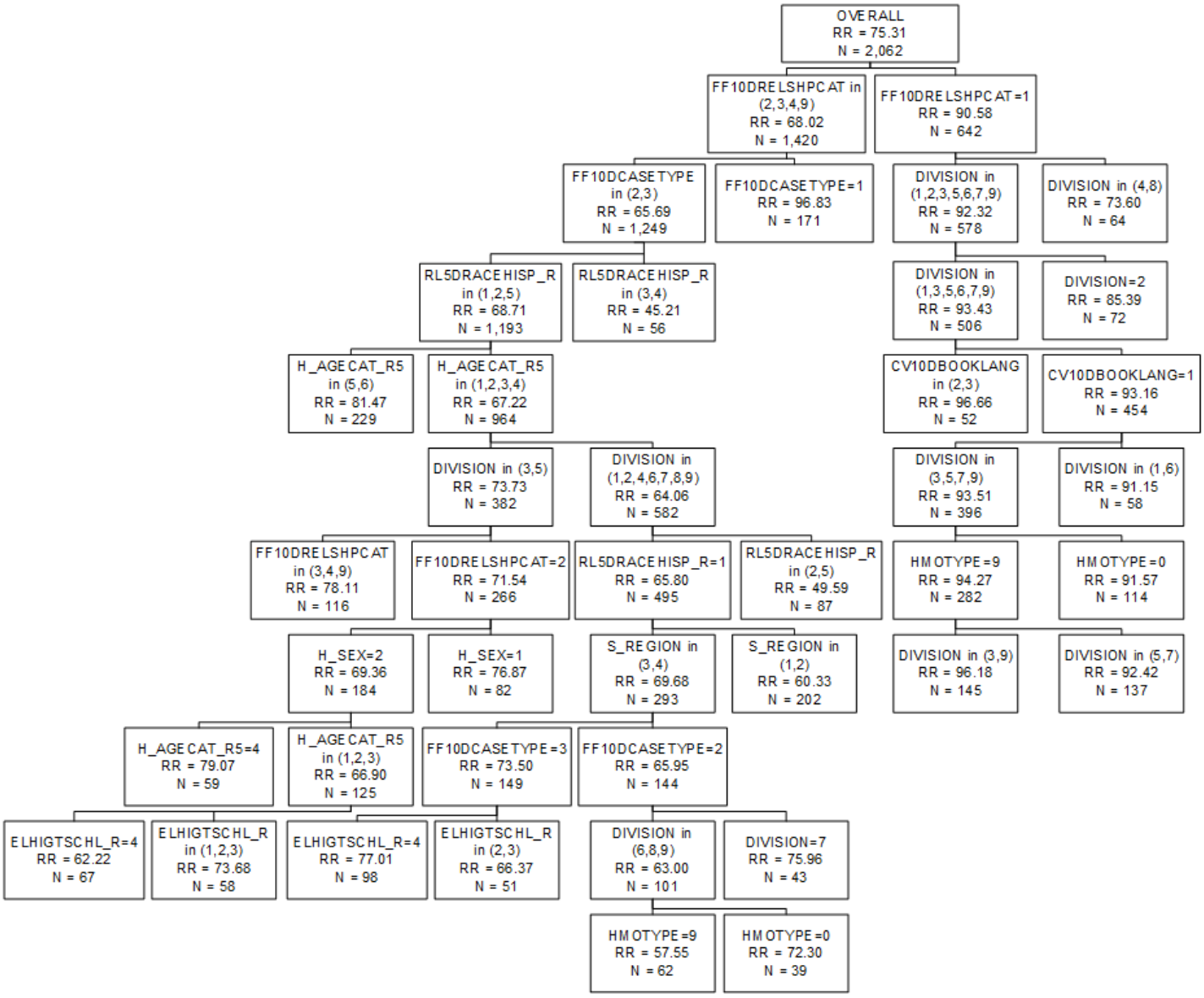
Variable names used in classification trees shown in parentheses

Figure 2. Stage 1 Covid-19 FF weight nonresponse adjustment cells – SP Refusal



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

Figure 3. Stage 2 Covid-19 FF weight nonresponse adjustment cells - FF nonresponse



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