# NATIONAL STUDY OF CAREGIVING (NSOC) IV
# WEIGHTING GUIDE

November 2022
Release 1.0

Suggested Citation:  Hu, Mengyao, Rui Jiao, & Vicki A. Freedman.  2022. National Study of Caregiving (NSOC) IV Weighting Guide. NHATS Technical Paper #33. Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.nhats.org.

## NSOC Design and Need for Weights

The National Study of Caregiving (NSOC) is designed to represent family and unpaid caregivers to a representative sample of living Medicare beneficiaries. NSOC IV Round 11 included family and unpaid caregivers to a representative sample of living and recently deceased Medicare beneficiaries who were age 65 or older as of October 1, 2014.[1].

In order to be able to make statements about estimates for this population group, NSOC IV Round 11 sample must be weighted to account for caregivers having different probabilities of selection and different probabilities of responding to NSOC. In order to make accurate statements about the variance of those estimates, design variables must also be used to account for NSOC's complex sample design.

This technical paper describes the NSOC weights and design variables for NSOC IV. Separate technical papers describe weights and design variables for NSOC I-III (Freedman et al. 2019a; Freedman et al. 2019b).

## Overview of Weight and Design Variables

NSOC IV Round 11 contains a set of weights (full sample weight, replicate weights) adjusted for nonresponse and design variables (stratum, cluster) that allow for proper variance estimation.

| NSOC IV | File | Full sample weight | Replicate weights | Stratum | Cluster |
|---------|------|--------------------|--------------------|---------|---------|
| Round 11 | Cross-sectional | w11cgfinwgt0 | w11cgfinwgt1-w11cgfinwgt56 | c11varstrat | c11varunit |

Analyses in which the caregiver is the unit of analysis should use the NSOC weight. (NHATS weights should be used when the unit of analysis is the care recipient (NHATS Sample person)). 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 to 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 and NSOC, K = 0.3 was used.

**How to Use NSOC Sample Weights and Design Variables**

Below we provide users with examples of how to use NSOC sample weights and design variables in Stata, SAS, and R. For each, we provide an example using full sample weights with design variables (Taylor Series Linearization) and using replicate weights.

*Stata*. In Stata, users should specify the following svyset command. Note that the svy prefix and

---

[1] All but 2 living and 3 deceased Sample Persons receiving assistance from NSOC participants were age 72 or older.

subpop() option should be used when performing subpopulation estimation.

```
/* Full Sample Weights and Design Variables */
        /*Caregivers to living sample persons*/
        svyset c11varunit [pweight=w11cgfinwgt0], strata(c11varstrat)
        svy, subpop(if fl11spdied==-1): [Stata procedures]

        /*Caregivers to deceased sample persons*/
        svyset c11varunit [pweight=w11cglmlfinwgt0], strata(c11varstrat)
        svy, subpop(if fl11spdied==1): [Stata procedures]

/*Replicate Weights */
        /*Caregivers to living sample persons*/
        svyset [pweight=w11cgfinwgt0], brrweight(w11cgfinwgt1 - w11cgfinwgt56) fay(.3)
        vce(brr) mse
        svy, subpop(if fl11spdied==-1): [stata procedures]

        /*Caregivers to deceased sample persons*/
        svyset [pweight=w11cglmlfinwgt0], brrweight(w11cglmlfinwgt1 - w11cglmlfinwgt56)
        fay(.3) vce(brr) mse
        svy, subpop(if fl11spdied==1): [stata procedures]
```

**SAS Commands**.  In SAS, users should specific domain, weight, cluster and strata statements.

```
/* Full Sample Weights and Design Variables */
        /*Caregivers to living sample persons*/
        [SAS procedure];
                domain fl11spdied;
                weight w11cgfinwgt0;
                cluster c11varunit;
                strata c11varstrat;
                [model or other statement];
        run;

        /*Caregivers to deceased sample persons*/
        [SAS procedure];
                domain fl11spdied;
                weight w11cglmlfinwgt0;
                cluster c11varunit;
                strata c11varstrat;
                [model or other statement];
        run;


/*Replicate Weights */
        /*Caregivers to living sample persons*/
```

```
proc [SAS survey procedure] data = [data name] varmethod = brr (fay = 0.3);
        domain fl11spdied;
        weight w11cgfinwgt0;
        repweights w11cgfinwgt1- w11cgfinwgt56;
        [model or other statement];
run;


/*Caregivers to deceased sample persons*/
proc [SAS survey procedure] data = [data name] varmethod = brr (fay = 0.3);
        domain fl11spdied;
        weight w11cglmlfinwgt0;
        repweights w11cglmlfinwgt1- w11cglmlfinwgt56;
        [model or other statement];
run;
```

**R Commands**.  In R, users can use the svydesign: commands with id, strata and weights options.

```
/* Full Sample Weights and Design Variables */
        /*Caregivers to living sample persons*/
        library(survey) #need this line only once per session
        nsoc.dsgn.l <- svydesign(id=~c11varunit, strata=~c11varstrat, weights=~w11cgfinwgt0,
        data= [data frame name], nest=TRUE)
        nsoc.subsetdsgn.l <- subset(nsoc.dsgn.l, fl11spdied == -1)
        [model or other statement]

        /*Caregivers to deceased sample persons*/
        library(survey) #need this line only once per session
        nsoc.dsgn.d <- svydesign(id=~c11varunit, strata=~c11varstrat,
        weights=~w11cglmlfinwgt0, data= [data frame name], nest=TRUE)
        nsoc.subsetdsgn.d <- subset(nsoc.dsgn.d, fl11spdied == 1)
        [model or other statement]

/*Replicate Weights */
        /*Caregivers to living sample persons*/
        library(survey) #need this line only once per session
        nsoc.dsgn.l <-svrepdesign(weights=~w11cgfinwgt0, data=[data frame name],
        type="Fay", rho = 0.3, repweights=" w11cgfinwgt[1-56]+")
        nsoc.subsetdsgn.l <- subset(nsoc.dsgn.l, fl11spdied == -1)
        [model or other statement]

        /*Caregivers to deceased sample persons*/
        library(survey) #need this line only once per session
        nsoc.dsgn.d <-svrepdesign(weights=~w11cglmlfinwgt0, data=[data frame name],
        type="Fay", rho = 0.3, repweights="w11cglmlfinwgt[1-56]+")
        nsoc.subsetdsgn.d <- subset(nsoc.dsgn.d, fl11spdied == 1)
```

[model or other statement]

More details about how to account for complex survey design in NSOC can be found in the technical paper on Accounting for Sample Design in NHATS and NSOC Analyses: Frequently Asked Questions (Freedman et al. 2022).


## Calculation of Weights

In each round, the calculation of the NSOC (cross-sectional file) weights began with the final NHATS analytic weight (see Montaquila, Freedman, Edwards and Kasper, 2012; DeMatteis, Freedman, and Kasper, 2016). For caregivers in networks with more than 5 eligible caregivers, the NHATS analytic weight was adjusted for caregiver subsampling (by multiplying by the number of eligible caregivers divided by 5). In NSOC IV Round 11, separate weights were constructed for caregivers to living and deceased SPs.[2]

For each weight, a 2-step weighting class adjustment for nonresponse was applied (Kalton and Flores-Cervantes 2003). A 2-step adjustment was used because there were two opportunities for NSOC interview nonresponse with potentially different mechanisms for nonresponse—refusal to provide contact information by the NHATS Sample Person) and other non-response at the caregiver level (for caregivers for whom the NHATS Sample Person did not refuse).

At each stage of adjustment, a set of variables were input into a classification tree analysis to determine which variables were associated with nonresponse. This approach uses SAS procedure HPSPLIT 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." For the cross-sectional weights, we set a minimum cell size of 30 respondents and a minimum weighted response rate for the cell of 0.5.

Finally, a raking adjustment was imposed to align weighted totals with those computed from the NHATS OP file (using NHATS analytic weights). The raking adjustment consisted of two dimensions: (1) the number of caregivers in the NHATS sample person's network and (2) the relationship of the caregiver to the sampled person.

---

[2] Caregivers of sample persons who died between NHATS and NSOC were treated as caregivers to living SPs for purposes of constructing NSOC IV cross-sectional sample weights. However, the value of their weight was assigned to w11cglmlfinwgt0.

**NSOC IV Round 11 Nonresponse Adjustments**

**Cross-sectional Files**
Appendix Table 1 shows variables considered for the first and second adjustments in NSOC IV Round 11 for caregivers to living and deceased SPs, along with weighted response rates for each level of each variable. Final non-response cells included a total of 10 indicators for caregivers of living SPs and 3 indicators for caregivers of deceased SPs (indicated in Appendix Table 1 with "a" and "b" for the first and second steps, respectively, for living SPs, and "c" and "d" for the first and second steps, respectively, for deceased SPs). Combinations of these variables created 25 unique nonresponse cells for the first adjustment and 7 nonresponse cells for the second adjustment for caregivers to living SPs (see Appendix Figures 1 and 2) and 6 unique nonresponse cells for the first adjustment and 5 nonresponse cells for the second adjustment for caregivers to deceased SPs (see Appendix Figures 3 and 4).

Overall, for both weights combined, the design effect due to variation in the NSOC IV Round 11 base weights (before nonresponse adjustment and raking) was 1.65. The design effect due to variation in the NSOC IV Round 11 analytic weights (after nonresponse adjustment and raking) was 1.68. The steps involved in creating the analytic weight did not introduce any influential outlier weights.

# Reference

DeMatteis, Jill, Freedman, Vicki A., and Kasper, Judith D. 2016. National Health and Aging Trends Study Round 5 Sample Design and Selection. NHATS Technical Paper #16. Baltimore: Johns Hopkins University School of Public Health. Available at www.NHATS.org.

Freedman, Vicki A., DeMatteis, Jill, Kasper, Judith D. 2019a. National Study of Caregiving (NSOC) I-III Weighting Guide. NHATS Technical Paper #24. Baltimore: Johns Hopkins University Bloomberg School of Public Health. Available at www.nhats.org.

Freedman, Vicki A. and Jennifer C. Cornman. 2019b. National Study of Caregiving III Time Diary User Guide (Final Release). Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.nhats.org.

Freedman, Vicki A., Mengyao Hu, Jill DeMatteis, Judith D. Kasper. 2022. Accounting for Sample Design in NHATS and NSOC Analyses: Frequently Asked Questions. NHATS Technical Paper #23 v2. Johns Hopkins University School of Public Health. Available at www.NHATS.org.

Kalton, G., and Flores-Cervantes, I. (2003). Weighting methods. *Journal of official statistics*, 19(2), 81.

Montaquila, Jill, Freedman, Vicki A., Edwards, Brad, and Kasper, Judith D. 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. Nonresponse Adjustments for NSOC IV Weights

**Table 1. Weighted Responses Rates for Variables used in Nonresponse Adjustment for NSOC IV Round 11 Cross-sectional NSOC Weights**

| | Contact info not refused by SP (Weighted Response Rate) | Interview complete, given contact info not refused (Weighted Response Rate) |
|---|---|---|
| OVERALL | 93.8 | 62.5 |
| **Hours of Help SP received last month [4][a] – (HOURSMONTH)** | | |
| 0: Not codeable (<1) | 91.3 | 59.6 |
| 1:1-<10 | 90.2 | 57.3 |
| 2:10-<20 | 94.9 | 63.5 |
| 3:20-<30 | 99.4 | 63.1 |
| 4: 30-<40 | 96.1 | 81.4 |
| 5: 40-<60 | 98.3 | 71.2 |
| 6: 60-<120 | 98.8 | 67.8 |
| 7: 120-<180 | 100.0 | 66.6 |
| 8: 180-744 (24/7) | 99.1 | 72.1 |
| 9:Missing / Inapplicable | 88.5 | 51.7 |
| **Relationship to SP [a][b][c][d] – (RELATION)** | | |
| 1:Spouse/Partner | 97.6 | 78.9 |
| 2:Son | 96.4 | 52.8 |
| 3:Daughter | 97.4 | 65.5 |
| 4:Sibling | 91.4 | 58.8 |
| 5:In-law | 87.1 | 59.4 |
| 6:Other relative | 90.9 | 48.9 |
| 7:Nonrelative | 80.5 | 59.0 |
| **Census Division[1][a][b][c][d] – (DIVISION)** | | |
| 1:New England | 95.9 | 59.2 |
| 2:Middle Atlantic | 94.6 | 63.8 |
| 3:East North Central | 96.7 | 67.3 |
| 4:West North Central | 92.2 | 73.8 |
| 5:South Atlantic | 95.0 | 57.0 |
| 6:East South Central | 88.8 | 63.9 |
| 7:West South Central | 92.6 | 54.3 |
| 8:Mountain | 95.0 | 57.0 |
| 9:Pacific | 92.6 | 67.2 |
| **SP Race/Ethnicity [a][b] – (RL5DRACEHISP)** | | |
| 1:White, non-Hispanic | 92.5 | 66.2 |
| 2:Black, non-Hispanic | 95.6 | 59.6 |
| 3:Other (Am Indian/Asian/Native Hawaiian /Pacific Islander/other specify), non-Hispanic | 97.8 | 62.8 |
| 4:Hispanic | 98.7 | 45.6 |
| 5:More than one | 100.0 | 0 |
| 6:DK/RF | 96.0 | 48.3 |
| **R11 SP Age at interview [4][a] – (R11D2INTVRAGE)** | | |
| 2:70-74 | 95.2 | 62.0 |
| 3:75-79 | 91.0 | 58.4 |
| 4:80-84 | 92.3 | 62.2 |
| 5:8-89 | 95.0 | 60.4 |
| 6:90 + | 94.6 | 66.3 |

(Continued next page)

| | Contact info not refused by SP (Weighted Response Rate) | Interview complete, given contact info not refused (Weighted Response Rate) |
|---|---|---|
| **Number of NSOC selected eligible helpers [a][c] – ( SAMPLECG_CNT)** | | |
| 1:1 Selected eligible helper | 94.3 | 78.2 |
| 2:2  Selected eligible helpers | 95.1 | 61.1 |
| 3:3  Selected eligible helpers | 95.2 | 59.2 |
| 4:4  Selected eligible helpers | 92.5 | 63.1 |
| 5: 5  Selected eligible helpers | 88.3 | 46.2 |
| **SP Education [a][b] – ( EL5DHIGSTSCHL )** | | |
| 0:DK / RF | 96.4 | 44.9 |
| 1:No schooling completed | 100.0 | 47.2 |
| 2:1st - 8th grade | 98.0 | 54.7 |
| 3: 9th - 12th grade | 93.1 | 59.8 |
| 4:High school graduate | 96.0 | 64.6 |
| 5:Vocational, technical, business, or trade school certificate or diploma | 92.6 | 61.6 |
| 6: Some college but no degree | 91.1 | 61.4 |
| 7:Associate's degree | 92.4 | 72.1 |
| 8: Bachelor's degree | 91.5 | 69.1 |
| 9: Master's, professional, or doctoral degree | 92.1 | 63.0 |
| **SP gender  – (R5DGENDER)** | | |
| 1:Male | 96.4 | 64.6 |
| 2:Female | 92.6 | 61.5 |
| **Reason for Proxy is Dementia[2][4][a] – (IS1REASNPRX)** | | |
| 0-inapplicable | 93.0 | NA |
| 1=Yes | 94.3 | NA |
| 2=No | 97.9 | NA |
| **NSOC Brochure accepted by SP[3] – (CC17)** | | |
| 0:No | NA | 66.7 |
| 1:Yes | NA | 61.8 |
| **Helper Gender[3] – (OP11DGENDER)** | | |
| 1:Male | NA | 57.7 |
| 2:Female | NA | 65.3 |
| **Helper lives with SP[3][b] – (OP11PRSNINHH)** | | |
| 0:Inapplicable | NA | 58.5 |
| 1:Yes | NA | 68.7 |
| 2:No | NA | 64.4 |
| **SP Residence[4][b] – ( R11DRESID)** | | |
| 1:Community | 93.4 | 60.9 |
| 2:Residential care resident, not nursing home (SP interview complete) | 92.5 | 69.1 |
| 4: Nursing home (SP interview complete) | 95.9 | 51.3 |
| **Census Metro/Micro Area Designation (2013)[1] – (S_METMICRO)** | | |
| 1:Metropolitan area | 93.3 | 61.9 |
| 2:Micropolitan area | 96.8 | 64.6 |
| 3:Non-metro | 95.4 | 66.3 |

[1]Based on county-level information from the September 30, 2014 CMS 5% EDB extract linked to the beneficiary's EDB address.
[2]The variable was only used in nonresponse modeling for nonresponse adjustment for SP refusal to provide contact information
[3]The variables were only used in nonresponse modeling for nonresponse adjustment for caregiver interview, given no SP refusal
[4]The variables were only used in nonresponse modeling for caregivers with a living SP
a=retained in classification tree analysis for nonresponse adjustment for SP refusal to provide contact information among caregivers with a living SP
b=retained in classification tree analysis for nonresponse adjustment for caregiver interview, given no SP refusal, among caregivers with a living SP

c=retained in classification tree analysis for nonresponse adjustment for SP refusal to provide contact information among caregivers with a deceased SP

d=retained in classification tree analysis for nonresponse adjustment for caregiver interview, given no SP refusal, among caregivers with a deceased SP
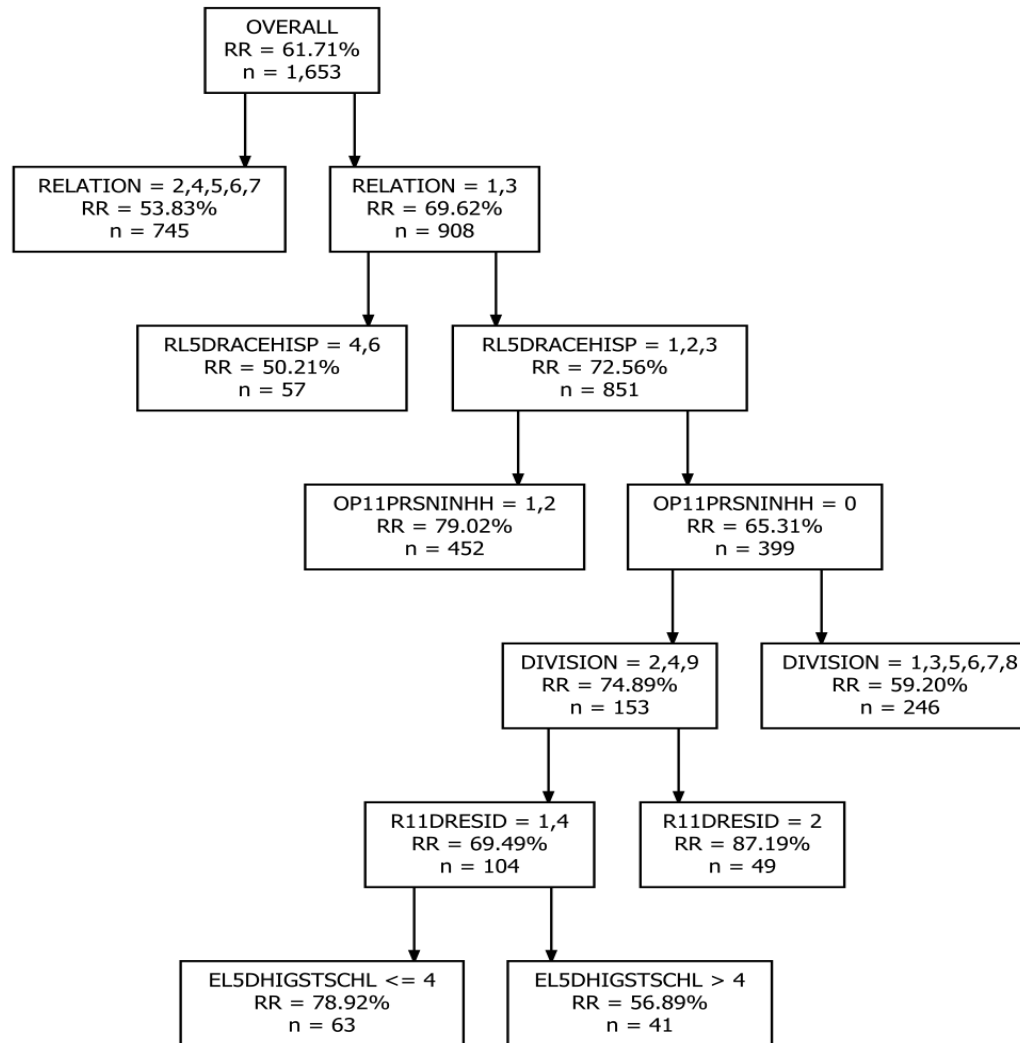
Variable names used in classification trees shown parenthetic

**Figure 1: Stage 1 NSOC IV cross-sectional weight nonresponse adjustment cells – SP Refusal (living SP)**
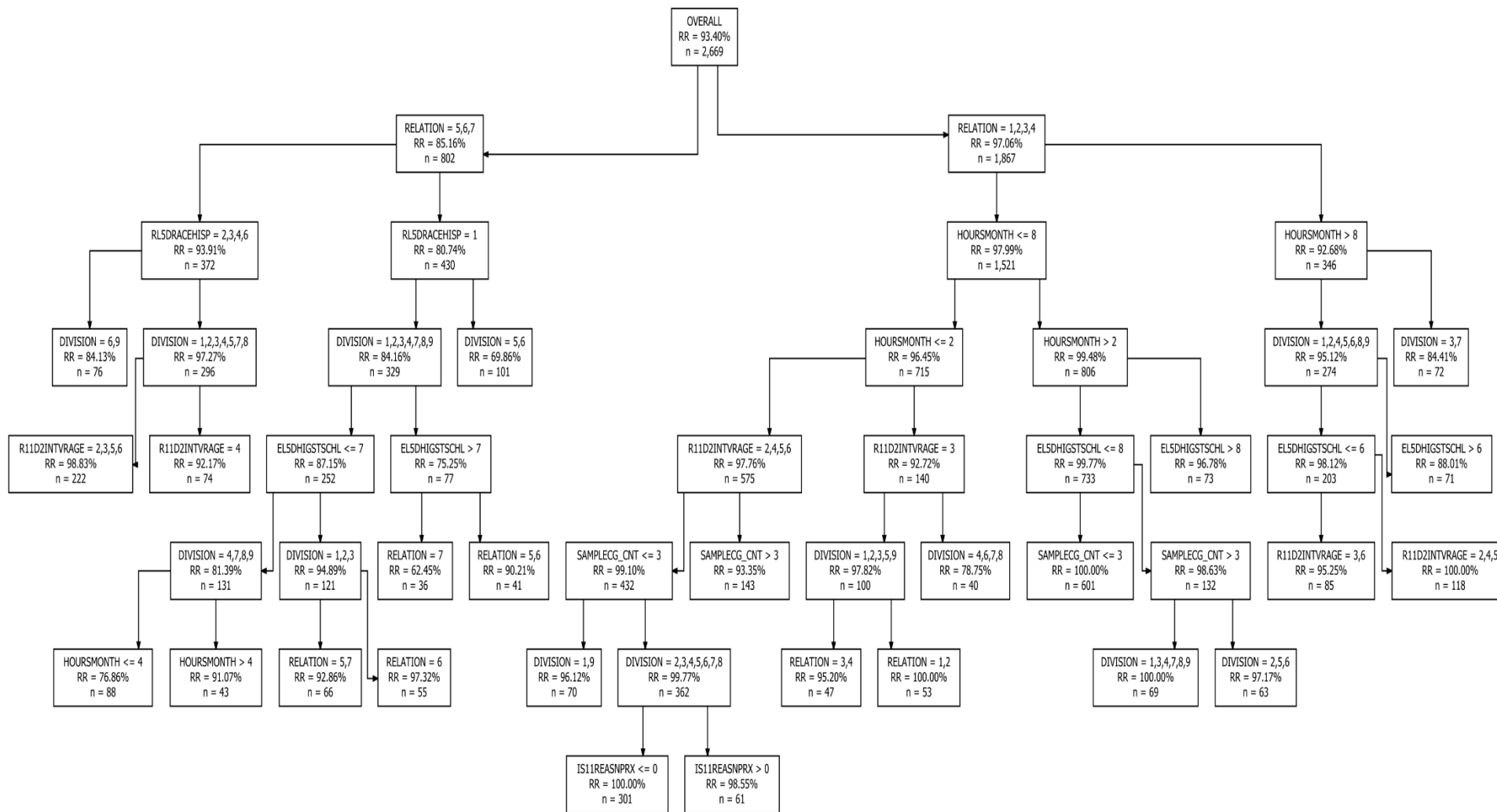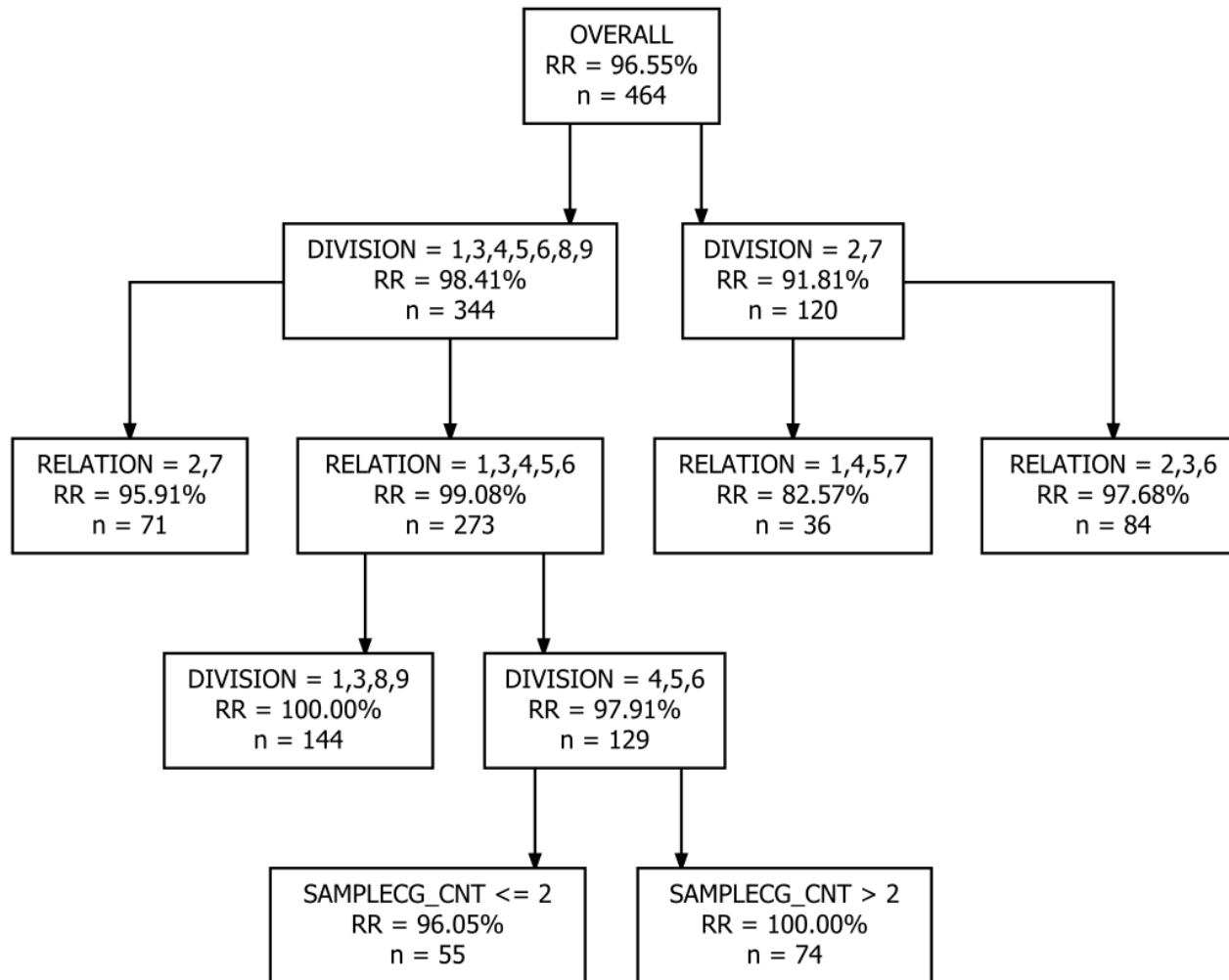


NOTE: "RR" is the weighted response rate for the particular cell, and "n" is the number of respondents in the cell.

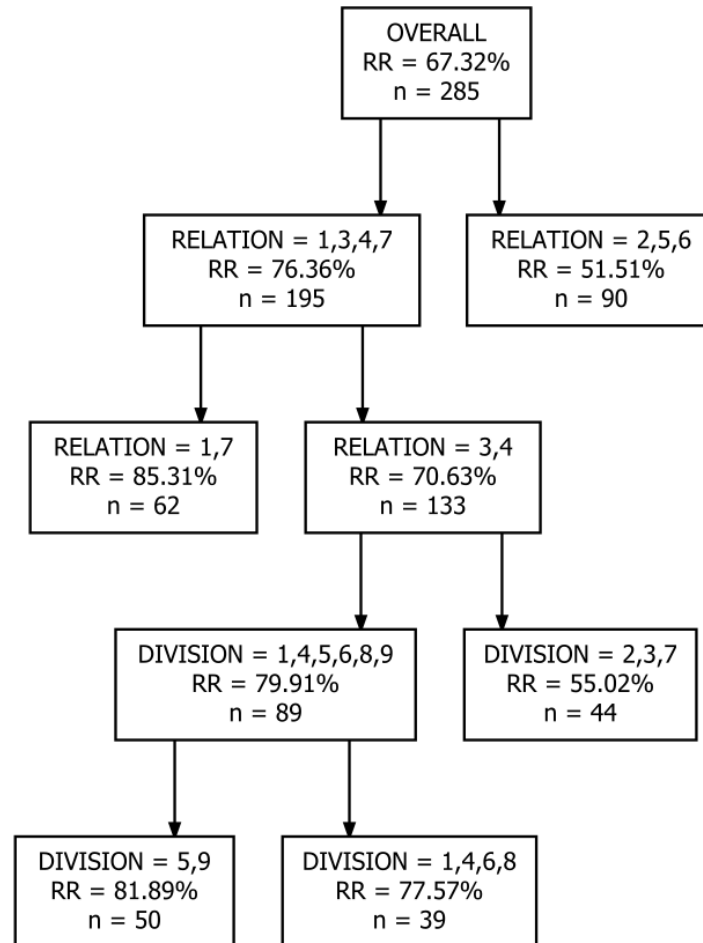**Figure 2: Stage 2 NSOC IV cross-sectional weight nonresponse adjustment cells – Caregiver to living SP**

```
                              ┌─────────────────┐
                              │    OVERALL      │
                              │  RR = 61.71%    │
                              │  n = 1,653      │
                              └────────┬────────┘
                    ┌──────────────────┴──────────────────┐
                    ▼                                      ▼
    ┌───────────────────────────┐          ┌───────────────────────────┐
    │ RELATION = 2,4,5,6,7      │          │    RELATION = 1,3         │
    │ RR = 53.83%               │          │    RR = 69.62%            │
    │ n = 745                   │          │    n = 908                │
    └───────────────────────────┘          └────────────┬──────────────┘
                                    ┌───────────────────┴──────────────────┐
                                    ▼                                       ▼
                  ┌───────────────────────────┐         ┌───────────────────────────┐
                  │ RL5DRACEHISP = 4,6        │         │ RL5DRACEHISP = 1,2,3      │
                  │ RR = 50.21%               │         │ RR = 72.56%               │
                  │ n = 57                    │         │ n = 851                   │
                  └───────────────────────────┘         └────────────┬──────────────┘
                                              ┌────────────────────┴──────────────────┐
                                              ▼                                        ▼
                            ┌───────────────────────────┐          ┌───────────────────────────┐
                            │ OP11PRSNINHH = 1,2        │          │ OP11PRSNINHH = 0          │
                            │ RR = 79.02%               │          │ RR = 65.31%               │
                            │ n = 452                   │          │ n = 399                   │
                            └───────────────────────────┘          └────────────┬──────────────┘
                                                          ┌────────────────────┴──────────────────┐
                                                          ▼                                        ▼
                                        ┌───────────────────────────┐          ┌───────────────────────────┐
                                        │ DIVISION = 2,4,9          │          │ DIVISION = 1,3,5,6,7,8     │
                                        │ RR = 74.89%               │          │ RR = 59.20%               │
                                        │ n = 153                   │          │ n = 246                   │
                                        └────────────┬──────────────┘          └───────────────────────────┘
                                    ┌────────────────┴──────────────────┐
                                    ▼                                    ▼
                  ┌───────────────────────────┐          ┌───────────────────────────┐
                  │ R11DRESID = 1,4           │          │ R11DRESID = 2             │
                  │ RR = 69.49%               │          │ RR = 87.19%               │
                  │ n = 104                   │          │ n = 49                    │
                  └────────────┬──────────────┘          └───────────────────────────┘
          ┌────────────────────┴──────────────────┐
          ▼                                        ▼
┌───────────────────────────┐          ┌───────────────────────────┐
│ EL5DHIGSTSCHL <= 4        │          │ EL5DHIGSTSCHL > 4         │
│ RR = 78.92%               │          │ RR = 56.89%               │
│ n = 63                    │          │ n = 41                    │
└───────────────────────────┘          └───────────────────────────┘
```

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

**Figure 3: Stage 1 NSOC IV cross-sectional weight nonresponse adjustment cells – SP Refusal (deceased SP)**



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

**Figure 4: Stage 2 NSOC IV cross-sectional weight nonresponse adjustment cells – Caregiver to deceased SP**



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