

National Study of Caregiving I - III
Twenty Interval Regression Income Imputations:

Rounds 1, 5 and 7

April 2024

Suggested citation: Hu, Mengyao & Vicki A. Freedman. 2024. National Study of Caregiving I-III
Twenty Interval Regression Income Imputations: Rounds 1, 5 and 7. NHATS Technical Paper
#44. Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.nhats.org.
This technical paper was prepared with funding from the National Institute on Aging
(U01AG032947; R01AG062477)

Contents

Overview	3
Auxiliary Imputed Income Files and Variables.....	3
Using Imputed Income in Analyses.....	4
Extent of Missing Data for Total Income	5
Imputation Approach.....	6
Imputation Groups.....	7
Variables Used in Imputation.....	7
Income Imputation Procedure.....	8
Descriptive Statistics for Total Income	8
Comparison with Previous Interval Regression Imputation Approach.....	9
References	11
Appendix	12

Overview

The National Study of Caregiving (NSOC) asks caregivers to report their total income for the prior calendar year. For those who do not report an exact value for total annual income, information is collected in broad categories. For single caregivers, individual income is reported; for those with a spouse or partner, couple income is reported. Caregivers who are spouses/partners of NHATS participants skip these questions in NSOC I-III, but have their income collected or imputed in NHATS (Freedman et al. 2023; Hu & Freedman 2024a).

In order to facilitate the use of these income data by the research community, we have developed multiple total income variables that include imputed values for those with missing income information. For NHATS and NSOC through Round 11, we initially released 5 sets of imputed values. For NSOC these values were based on interval regression models and were released in a set of enhanced files.

We recently released 20 imputed values of income using an interval regression methodology for all rounds of NHATS (Hu & Freedman 2024a). To ensure more reliable standard error estimation and consistency with NHATS, we now provide 20 imputed values for all available rounds of NSOC. The 20 imputations are provided in a set of auxiliary files and were prepared using an interval regression methodology.

This technical paper describes the updated income imputation methodology using interval regression in NSOC I-III (Rounds 1, 5 and 7). Comparisons with the initial release of five sets of imputed values are provided. A separate technical paper describes the methodology for NSOC IV (Rounds 11 forward) (Hu & Freedman 2024b).

Although this technical paper demonstrates that percentile distributions for the two sets of estimates are close, means and standard deviations are lower for estimates based upon the updated methodology. We therefore recommend that researchers use the auxiliary set of 20 imputations.

Auxiliary Imputed Income Files and Variables

The NSOC Imputed Income file includes one observation for each NSOC participant in the given round. In Rounds 1 and 5, all caregivers have reported or imputed values of total income. In Round 7 only, following NSOC missing data conventions, those who helped in the last year but not the last month (fl7helpyear=1 yes) were set to -4 on these variables.

The income imputation file for Rounds 1, 5 and 7 of NSOC includes 20 values of total income. In addition, a flag variable indicating which cases have imputed (vs. reported) values in NSOC and NHATS is included along with a derived variable indicating the reason for imputation.

Table 1. Total Income Imputation Variables Included in NSOC I-III Auxiliary Files (Rounds 1, 5, 7)

Variable name	Label	Description
chi#toincimif	C# F IMPUTED TOTAL INC FLG INTERVAL	Flag indicating imputation: 1. Reported in NSOC 2. Reported in NHATS (Spouse/partner of SP) 3. Imputed in NSOC 4. Imputed in NHATS (Spouse/partner of SP)
chi#dtoincimi1 - chi#dtoincimi20	C# D HI10 IMPUTED TOTAL INC1-20 INTERVAL	Includes reported and imputed amounts
chi#dtoincimreas	C# D IMPUTED TOTAL INC REASON	Derived variable indicating reason for imputation: 1. Not imputed: exact value reported 2. Imputed: bracket response only 3. Imputed: missing exact value and bracket response 4. Imputed: other

Using Imputed Income in Analyses

We recommend that researchers use the auxiliary set of 20 imputations, since they yield more reliable estimates with narrower standard errors than the 5 imputations previously provided.

To adjust coefficients and standard errors for the variability between imputations following the combination rules by Rubin (1987), users can use the “mi estimate: svy” command in Stata to run estimations on the imputed dataset. To use this procedure, researchers must first generate an income variable that only includes non-imputed income values (that is, they must replace outliers that were imputed with a missing value. We have included code below for merging the auxiliary file, generating a non-imputed version of total income, and estimating weighted mean total income.

Below is an example of Stata code that may be used to estimate unweighted and weighted mean income using cross sectional NSOC Rounds 1, 5 or 7. The # stands for round number. Round 7 requires some extra steps to create a weight and to subset to those helping in the last month.

```
*merge income imputation file with NSOC file
use “[path]\NSOC_R#_Interval_Inc_Imp_File.dta”, clear
merge 1:1 spid opid using “[path]\NSOC_R#_Crss_File.dta”
drop _merge
```

*Round 7 only: create a weight for CGs to both living and deceased SPs
gen w7cg_lmlwgt0 = w7cgfinwgt0 if w7cgfinwgt0 > 0
replace w7cg_lmlwgt0 = w7cglmlfinwgt0 if w7cgfinwgt0 == 0

*generate a total income variable that includes only non-imputed income values
gen chi#income2 = chi#dtoincimi1
replace chi#income2 = . if chi#dtoincimreas > 1

*save variables to a new dataset
save newdataset, replace

*use this new dataset for analysis
use newdataset, clear

*use mi import to import data that include reported income and additional imputed income variables

```
mi import wide, imputed (chi#income2 = chi#dtoincimi1 chi#dtoincimi2 chi#dtoincimi3  
chi#dtoincimi4 chi#dtoincimi5 chi#dtoincimi6 chi#dtoincimi7 chi#dtoincimi8 chi#dtoincimi9  
chi#dtoincimi10 chi#dtoincimi11 chi#dtoincimi12 chi#dtoincimi13 chi#dtoincimi14  
chi#dtoincimi15 chi#dtoincimi16 chi#dtoincimi17 chi#dtoincimi18 chi#dtoincimi19  
chi#dtoincimi20 ) drop  
mi set wide
```

*estimate unweighted mean income of all caregivers
mi estimate: mean chi#income2

*Rounds 1 and 5 estimate weighted mean income
mi svyset c#varunit [pweight=w#cgfinwgt0], strata(c#varstrat) singleunit(centered)
mi estimate: svy: mean chi#income2

*Round 7 only: estimate weighted mean income
mi svyset c7varunit [pweight=w7cg_lmlwgt0], strata(c7varstrat) singleunit(centered)
mi estimate: svy, subpop(if chi7toincimif > 0): mean chi7income2

Extent of Missing Data for Total Income

Table 2 shows the percentage of participants that require imputation and the reason for NSOC Rounds 1, 5 and 7. The percentage of respondents who did not require imputation (i.e., they provided an exact value of total income in either NSOC or their spouse provided the amount in NHATS¹) ranged from 63%-68%. An additional 7%-9% had values imputed in NHATS and

¹In NSOC Rounds 1, 5, and 7, spouses/partners of living NHATS SPs skipped income questions because their income was collected in NHATS. We brought over SP's reported (ia#totinc) or imputed (ia#dtoincimi1- ia#dtoincimi20) income from NHATS for these cases.

brought over to NSOC. A categorical response (requiring imputation within a bracket) was reported by 5%-10% and 16%-17% were missing both an exact value and a categorical response (requiring a full imputation). Another 2%-3% provided values that were considered outliers (extremely low or high values) and were also imputed. In Round 7 only, a small percentage skipped the income series by design (those who did not help in the last month or last month of life).

Table 2. Percentage of NSOC I-III Respondents Requiring Income Imputation and Reason, Rounds 1, 5, and 7

	NSOC I (Round 1)	NSOC II (Round 5)	NSOC III (Round 7)	
			Cross-sectional	Longitudinal
Reason for Imputation (%)				
Not imputed: exact value reported	62.8	68.3	64.6	65.0
Reported in NSOC	(50.2)	(55.7)	(53.2)	(52.7)
Reported in NHATS	(10.6)	(12.6)	(11.4)	(12.3)
Imputed in NHATS	8.5	8.7	6.9	6.6
Imputed in NSOC: bracket response only	9.5	5.2	7.5	6.8
Imputed in NSOC: missing exact value and bracket response	16.2	15.7	17.3	15.0
Imputed in NSOC: other	3.0	2.0	2.2	2.3
Did not help last month or in last month of life*			1.5	4.2
Total	2,007	2,204	2,652	1,453
Note: In NSOC III, respondents who did not help last month or in the last month of life skipped the income questions. These cases are not imputed and are set to -4.				

Imputation Approach

For purposes of imputation, we transformed income reports to log income (after adding \$1). For respondents missing an exact value (or reporting an outlier value suspected to be misreporting²), we used multiple imputation by chained equations (MICE) to impute missing covariates and missing income. We used interval regression in Stata, which fills in the missing values of partially observed (censored) income (that is, within a range).

To implement the interval regression approach, we specified a lower and upper limit for all cases. For cases with an exact value, we added \$1 to the reported value, and specified the lower and upper limits to be the log-transform of that amount (essentially assigning the reported value to each imputation). For cases with only categorical information, we used the log-transformed upper and lower amounts of the bracket (plus \$1) as limits. Cases with only an

² Cases with reported income <\$200 and >\$900,000 were considered outliers and therefore imputed.

upper limit received \$1 as their lower limit. Cases with only a lower limit received a missing value for their upper limit, which can take any positive value above the lower limit. For low-end outliers (<\$200), we assumed that the exact value was in the lowest category. For cases missing both sources of income (exact value and categorical information) and for high-end outliers (>\$900,000), the lower limit was specified as \$1 and upper limit set to missing. Before releasing the data, we transformed reported and imputed values back to the original (non-logged) dollar scale.

Imputation Groups

Imputation groups were formed based on available covariates. For NSOC I and II (Rounds 1 and 5), three groups were formed.

1. Non-spouse/partner caregivers to living SPs who completed the entire interview (non-breakoff);
2. Spouse/partner caregivers to living SPs;
3. Non-spouse/partner caregivers to living SPs who did not complete the interview (breakoff);

For NSOC III (Round 7), we also imputed income for a fourth group: caregivers to deceased NHATS Sample Persons.

4. Caregivers to deceased SPs (i.e., Last month of life, LML)

Note that Group 2 has reported or imputed income brought over from NHATS, and therefore does not have any missing income information. That is, imputations were conducted for Groups 1 and 3 for Rounds 1 and 5 and for Groups 1, 3 and 4 for Round 7.

To ensure adequate sample sizes in the imputation models, when imputing Groups 3 and 4, we also included Group 1 and 2 cases (either their reported or the average of their 20 imputed values) in the models.

Variables Used in Imputation

Here we provide an overview of covariates included in the imputation models. See Appendix Table 2 for details by imputation group and round.

- Caregiver characteristics include age, gender, race/ethnicity, number of other people in the household, education and spouse/partner's education, and relationship of caregiver to sample person.
- NHATS Sample Person characteristics include census division and metro/nonmetro residence, income, and whether the caregiver lives with the SP in the current round.
- Caregiver economic resources include insurance coverage (private insurance, Medicaid enrollment), assets (e.g., whether the CG and their spouse/partner own a home, have a checking account, savings account, certificate of deposit, retirement plan, and stocks or

mutual funds); business/farm ownership and whether worked for pay last week. Prior round income is also included in NSOC III (Round 7) longitudinal file cases.

Income Imputation Procedure

For each imputation group, we estimated multiple imputations by chained equations (MICE) first to fill in missing values for covariates and logged income (starting with the least missing). Interval regression was used to impute 20 values of logged income. We chose 20 to be consistent with the NHATS imputation methodology. We also tested various numbers of imputations, ranging from 5 to 50. We found that 20 imputations consistently resulted in stable standard error estimation. The final number of imputations also aligns with recommendations from the literature when up to 30% of respondents have missing values (Graham et al., 2007; Bodener, 2008). For each imputation group, we specified 20 burn-in iterations (i.e., the number of iterations before the first set of imputed values is drawn). Final imputed variables were transformed back to a non-logged dollar scale.

We incorporated the NSOC survey design into the imputation procedure by using NSOC weights, which build upon NHATS weights, in the imputation model (using a pweight statement).³

A summary of imputation model results can be found in the Appendix.

Descriptive Statistics for Total Income

Table 3 presents weighted descriptive statistics for total income using reported and 20 imputed values generated through the interval regression approach. The weighted means are based on averages over the 20 imputations and the standard error calculations take into account the extra variation introduced by the uncertainty of the imputation.

Table 3. Weighted Means and Percentiles for Total Income (in dollars), NSOC I-III (Rounds 1, 5 and 7)

	Reference Year	n	Mean	Standard error	25 th percentile	Median	75 th percentile
Round 7 ^a							
Crosssect.	2016	2,612	66,251	2,246	24,000	50,000	90,000
Longitudinal	2016	1,392	56,070	2,053	21,000	41,000	75,000
Round 5	2014	2,204	59,050	2,246	21,000	43,953	83,436
Round 1	2010	2,007	53,652	2,106	19,039	40,000	75,000

Note: Weighted means and standard errors were computed using “mi estimate: svy” command. Median and percentiles were computed using the “_pctile” command for the average of the 20 imputed variables. Final

³Although we did include publicly available geographic indicators (SP’s Census division and metro/non-metro residence) as controls, we could not directly account for strata and cluster variables in the imputation strategy. We considered other approaches—e.g., adding a variable for each strata by cluster combination or imputing separately by each unique cluster (Heeringa, West & Berglund, 2017; Kalpourtzi et al. 2024; Reiter 2006). However, the study has too many cluster/strata combinations to implement these approaches. For NHATS imputations, we also explored controlling for the full set of replicate weights, which captures the survey’s complex design, but models did not consistently converge.

analytic weights were used in these analyses. Total income is for individuals and their spouse/partners (if applicable).

^aCaregivers who did not help last month or in the last month of life are excluded (n=40 for cross-sectional sample and n=61 for longitudinal sample)

Comparison with Previous Interval Regression Imputation Approach

Previously, NSOC provided on the enhanced caregiver files 5 imputations prepared using an interval regression methodology. Table 4 summarizes the main differences between the current interval regression imputation (yielding 20 imputations) and the previous approach (yielding 5 imputations).

Table 4. Differences between previous vs. current interval regression imputation approaches

Previous imputation (5 imputations)	Current imputation (20 imputations)
<i>Imputation preparation: Editing before multiple imputation</i>	
<ul style="list-style-type: none"> • Reports of zero income were treated as missing and were imputed • Reports of very high incomes were accepted as valid reports • Income brought over from NHATS for caregivers who were spouse/partner of SP was based on 5 imputed hot deck values 	<ul style="list-style-type: none"> • Reports of less than \$200 were treated as missing exact value and imputed into the lowest income category • Reports above \$900,000 were treated as missing an exact value and imputed • Income brought over from NHATS for caregivers who were spouse/partner of SP was based on 20 interval regression imputations
<i>Imputation</i>	
<ul style="list-style-type: none"> • For Groups 1, 3 and 4, SP's income from NHATS was reported or based on 5 imputed hot deck values • After bringing over income from NHATS, Group 2 cases with missing income (values of zero) were imputed • For Groups 3 and 4, Group 1 cases are included in the imputation model 	<ul style="list-style-type: none"> • For Groups 1, 3 and 4, SP's income from NHATS was reported or based on 20 interval regression imputations • After bringing over income from NHATS, Group 2 cases did not have missing values • For Groups 3 and 4, both Groups 1 and cases are included in the imputation model
<i>Delivery</i>	
<ul style="list-style-type: none"> • 5 imputed variables (chi#incim1-chi#incim5) and a flag variable (chi#incimf) indicating which cases were imputed 	<ul style="list-style-type: none"> • 20 imputed variables (chi#dtoincimi1 - chi#dtoincimi20), a flag variable (chi#toincimif) indicating which cases were imputed, and a derived variable indicating reason for imputation (chi#dtoincimreas)
<ul style="list-style-type: none"> • Included in enhanced NSOC I-III data file 	<ul style="list-style-type: none"> • Included in an auxiliary file zipped with the NSOC data files

Table 5 compares imputed income for NSOC I-III (Rounds 1, 5 and 7) based on the two sets of imputations. The percentile distributions for the two sets of estimates are close, but means and standard errors are lower in the 20 interval regression approach, likely because outliers were imputed.

Table 5. NSOC I-III Imputed Income Rounds 1, 5 and 7: 5 vs. 20 Interval Regression Imputations

	n	Mean	Standard error	25 th percentile	Median	75 th percentile
NSOC III (Round 7) ^a						
Cross-sectional file						
5 imputations	2,612	72,168	5,844	23,000	49,673	90,000
20 imputations	2,612	66,251	2,246	24,000	50,000	90,000
Longitudinal file						
5 imputations	1,392	58,461	2,646	19,000	40,000	75,000
20 imputations	1,392	56,070	2,053	21,000	41,000	75,000
NSOC II (Round 5)						
5 imputations	2,204	60,373	2,348	20,600	42,000	80,000
20 imputations	2,204	59,050	2,246	21,000	43,953	83,436
NSOC I (Round 1)						
5 imputations	2,007	55,132	2,449	19,857	40,000	75,000
20 imputations	2,007	53,652	2,106	19,039	40,000	75,000

Note: Weighted means and standard errors were computed using “mi estimate: svy” command. Median and percentiles were computed using the “_pctile” command for the average of the 20 imputed variables. Final analytic weights were used in these analyses.

^aCaregivers who did not help last month or in the last month of life are excluded (n=40 for cross-sectional file, n=61 for longitudinal file)

References

- Bodner, T. E. (2008). What improves with increased missing data imputations?. *Structural equation modeling: a multidisciplinary journal*, 15(4), 651-675.
- Freedman, V. A., Hu, M., and Wolff, J. (2023). National Study of Caregiving IV User Guide: Rounds 11-12 Final Release. Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.NHATS.org.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention science*, 8, 206-213.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). Applied survey data analysis 2nd edition. CRC press.
- Hu, M. & Freedman, V. A. (2024a). National Health and Aging Trends Study Twenty Interval Regression Income Imputations: Rounds 1-12. NHATS Technical Paper #43. Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.nhats.org.
- Hu, M. & Freedman, V. A. (2024b). National Study of Caregiving IV Twenty Interval Regression Income Imputations: Rounds 11-12. NHATS Technical Paper #45. Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.nhats.org.
- Hu, M., Freedman, V. A. & Jiao, R. (2024). National Health and Aging Trends Study Five Hot Deck Income Imputations: Rounds 1-11. NHATS Technical Paper #42. Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.nhats.org.
- Kalpourtzi, N., Carpenter, J. R., & Touloumi, G. (2024). Handling missing values in surveys with complex study design: A simulation study. *Journal of Survey Statistics and Methodology*, 12(1), 105-129.
- Reiter, J. P., Raghunathan, T. E., & Kinney, S. K. (2006). The importance of modeling the sampling design in multiple imputation for missing data. *Survey Methodology*, 32(2), 143.
- Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley
- Rubin, D.B. 1996. Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473-489.

Appendix

Appendix Table 1. Sample Sizes for Income Imputation Groups: NSOC 1-III, Round 1, 5 and 7

	1 Non-Spouse/ Partner	2 Spouse/ Partner	3 Breakoff	4 LML (NSOC III only)
<i>NSOC I (Round 1)</i>				
Number with Reported Income	1,009	422 ^a	0	-
Number with Imputed Income	553	0	23	-
Total Model Sample Size	1,562	422	2,007 ^b	-
<i>NSOC II (Round 5)</i>				
Number with Reported Income	1,227	471 ^a	0	-
Number with Imputed Income	484	0	22	-
Total Model Sample Size	1,711	471	2,204 ^b	-
<i>NSOC III (Round 7) Cross-sectional</i>				
Number with Reported Income	1,220	479 ^a	1	196
Number with Imputed Income	592	0	32	92
Total Model Sample Size	1,812	479	2,324 ^b	2,579 ^b
<i>NSOC III (Round 7) Longitudinal</i>				
Number with Reported Income	553	275 ^a	0	213
Number with Imputed Income	236	0	10	105
Total Model Sample Size	789	275	1,074 ^b	1,382 ^b
^a Includes reported and imputed income from NHATS. ^b Group 1 & 2 cases (reported or average of 20 imputed values) are also included in model estimation.				

Appendix Table 2. Covariates used to impute income in NSOC I-III (Rounds 1, 5 and 7) for each imputation group

Covariates	Imputation Group			
	1 Non-Spouse/ Partner	2 Spouse/ Partner	3 Breakoff	4 LML (NSOC III only)
CG Age	Y	n/a	Y	Y
CG Race / ethnicity	Y	n/a	Y	Y
CG Gender	Y	n/a	Y	Y
Number of other people live in CG household	Y	n/a	Y	Y
CG education	Y	n/a	Y	Y
CG spouse/partner's education	Y ^a	n/a	Y ^b	Y ^b
SP census division	Y	n/a	Y	Y
SP metro / non-metro residence	Y	n/a	Y	Y
CG relationship to SP	Y	n/a	Y	Y
SP's income from NHATS (log-transformed)	Y	n/a	Y	Y ^c
CG in household with SP	Y	n/a	Y	Y ^d
CG has private insurance coverage	Y	n/a		Y
CG has Medicaid coverage	Y	n/a		Y
CG owns home	Y	n/a		Y
CG has checking acct	Y	n/a		Y
CG has savings account	Y	n/a		Y
CG has certificate of deposit	Y	n/a		Y
CG has retirement plan	Y	n/a		Y
CG has stocks or mutual funds	Y	n/a		Y
CG worked for pay last week	Y	n/a		
CG owns business or farm	Y	n/a		
CG income from round 5	Y ^e	n/a	Y ^e	Y ^e
^a Missing values imputed in NSOC I, II, and III cross-sectional samples; treated as missing category in NSOC III longitudinal sample. ^b Missing values treated as missing category in NSOC III. ^c Round 5 SP income was used in NSOC III. ^d Not included in NSOC III longitudinal sample. ^e Only included in NSOC III longitudinal sample.				

Appendix Table 3. Model results for NSOC I (Round 1) by imputation groups

	Grp 1: Non-Spouse/ Partner		Grp 3: Breakoff	
	Coef.	SE	Coef.	SE
CG Age	0.01	0.00	0.01	0.00
CG Race / ethnicity				
1 White, non-Hispanic				
2 Black, non-Hispanic	0.09	0.07	-0.01	0.05
3 Others/Missing	0.01	0.15	-0.04	0.15
4 Hispanic	-0.07	0.10	-0.08	0.09
CG Gender				
Female				
Male	0.04	0.05	-0.01	0.04
Number of other people live in CG household	0.00	0.02	0.00	0.02
CG education				
1 <HS				
2 HS	0.11	0.12	0.27	0.07
3 >HS, <College	0.22	0.11	0.51	0.06
4 College+	0.42	0.12	0.81	0.07
CG spouse/partner's education				
0 No spouse				
1 <HS	0.66	0.12	0.75	0.08
2 HS	0.61	0.08	0.91	0.06
3 >HS, <College	0.70	0.08	0.97	0.06
4 College+	0.80	0.08	1.13	0.06
9 Missing			0.91	0.36
SP census division				
1 Northeast Region: New England Division				
2 Northeast Region: Middle Atlantic Division	-0.08	0.11	-0.17	0.08
3 Midwest Region: East North Central Division	-0.01	0.11	-0.10	0.08
4 Midwest Region: West North Central Division	-0.09	0.10	-0.15	0.08
5 South Region: South Atlantic Division	-0.16	0.10	-0.21	0.08
6 South Region: East South Central Division	-0.29	0.12	-0.36	0.09
7 South Region: West South Central Division	-0.17	0.11	-0.24	0.09
8 West Region: Mountain Division	-0.15	0.14	-0.26	0.11
9 West Region: Pacific Division	-0.13	0.11	-0.20	0.08
SP metro / non-metro residence				
1 Metropolitan				
2 Non-metropolitan	-0.16	0.08	-0.12	0.05
CG relationship to SP				
2 Spouse/partner of SP				
3 Children			0.59	0.08

4 Other relatives	-0.32	0.06	0.14	0.09
5 Non-relatives	-0.18	0.08	0.33	0.10
SP's income from NHATS (log-transformed)	0.02	0.04	0.23	0.03
CG in household with SP				
1 YES				
2 NO	0.11	0.06	0.26	0.06
CG has private insurance coverage				
No				
Yes	0.39	0.07		
CG has Medicaid coverage				
No				
Yes	-0.14	0.12		
CG owns home				
No				
Yes	0.25	0.08		
CG has checking acct				
No				
Yes	0.09	0.12		
CG has savings account				
No				
Yes	0.08	0.06		
CG has certificate of deposit				
No				
Yes	0.08	0.05		
CG has retirement plan				
No				
Yes	0.34	0.07		
CG has stocks or mutual funds				
No				
Yes	0.10	0.06		
CG worked for pay last week				
Yes				
No	-0.34	0.06		
Retired	-0.30	0.11		
CG owns business or farm				
No				
Yes	0.07	0.06		
Constant	8.68	0.42	6.10	0.36
Log of sigma	-0.44	0.05	-0.41	0.03
Sigma	0.64	0.03	0.66	0.02

Appendix Table 4. Model results for NSOC II (Round 5) by imputation groups

Variable Label	Grp 1: Non-Spouse/ Partner		Grp 3: Breakoff	
	Coef.	SE	Coef.	SE
CG Age	0.01	0.00	0.01	0.00
CG Race / ethnicity				
1 White, non-Hispanic				
2 Black, non-Hispanic	0.10	0.07	-0.05	0.05
3 Others/Missing	-0.03	0.09	-0.08	0.08
4 Hispanic	0.09	0.09	0.12	0.08
CG Gender				
Female				
Male	-0.09	0.05	-0.09	0.04
Number of other people live in CG household	0.05	0.02	0.03	0.02
CG education				
1 <HS				
2 HS	-0.03	0.13	0.22	0.08
3 >HS, <College			0.41	0.08
4 College+			0.71	0.08
CG spouse/partner's education				
0 No spouse				
1 <HS	0.50	0.11	0.67	0.08
2 HS	0.78	0.07	1.04	0.07
3 >HS, <College	0.71	0.07	1.09	0.06
4 College+	0.80	0.07	1.22	0.06
9 Missing			0.62	0.21
SP census division				
1 Northeast Region: New England Division				
2 Northeast Region: Middle Atlantic Division	0.08	0.13	0.17	0.10
3 Midwest Region: East North Central Division	0.04	0.12	0.10	0.09
4 Midwest Region: West North Central Division	0.00	0.12	0.11	0.09
5 South Region: South Atlantic Division	-0.11	0.12	-0.03	0.08
6 South Region: East South Central Division	-0.09	0.13	-0.04	0.09
7 South Region: West South Central Division	0.00	0.13	0.09	0.08
8 West Region: Mountain Division	-0.11	0.18	-0.05	0.12
9 West Region: Pacific Division	-0.01	0.12	-0.01	0.09
SP metro / non-metro residence				
1 Metropolitan				
2 Non-metropolitan	-0.12	0.06	-0.09	0.05
CG relationship to SP				
2 Spouse/partner of SP				
3 Children			0.56	0.09
4 Other relatives	-0.09	0.07	0.34	0.11
5 Non-relatives	-0.27	0.08	0.10	0.10
SP's income from NHATS (log-transformed)	0.05	0.03	0.28	0.03
CG in household with SP				

1 YES				
2 NO	0.12	0.06	0.17	0.06
CG has private insurance coverage				
No				
Yes	0.17	0.06		
CG has Medicaid coverage				
No				
Yes	-0.36	0.09		
CG owns home				
No				
Yes	0.19	0.07		
CG has checking acct				
No				
Yes	0.17	0.10		
CG has savings account				
No				
Yes	0.19	0.06		
CG has certificate of deposit				
No				
Yes	0.02	0.05		
CG has retirement plan				
No				
Yes	0.39	0.06		
CG has stocks or mutual funds				
No				
Yes	0.19	0.04		
CG worked for pay last week				
Yes				
No	-0.20	0.06		
Retired	-0.26	0.07		
CG owns business or farm				
No				
Yes	0.09	0.06		
Constant	8.39	0.39	5.64	0.41
Log of sigma	-0.47	0.04	-0.40	0.03
Sigma	0.63	0.03	0.67	0.02

Appendix Table 5. Model results for NSOC III (Round 7) cross-sectional sample by imputation groups

	Grp 1: Non-Spouse/ Partner		Grp 3: Breakoff	
	Coef.	Std. Err.	Coef.	Std. Err.
CG Age	0.01	0.00	0.01	0.00
CG Race / ethnicity				
1 White, non-Hispanic				
2 Black, non-Hispanic	0.00	0.05	-0.07	0.05
3 Others/Missing	-0.23	0.13	-0.24	0.09
4 Hispanic	-0.06	0.08	-0.06	0.08
CG Gender				
1 MALE				
2 FEMALE	-0.07	0.04	0.00	0.04
Number of other people live in CG household	0.00	0.02	-0.01	0.02
CG education				
1 <HS				
2 HS	0.01	0.12	0.25	0.07
3 >HS, <College	0.13	0.12	0.35	0.07
4 College+	0.28	0.13	0.67	0.08
CG spouse/partner's education				
0 No spouse				
1 <HS	0.47	0.11	0.54	0.07
2 HS	0.57	0.07	0.81	0.06
3 >HS, <College	0.61	0.05	0.96	0.05
4 College+	0.78	0.06	1.16	0.05
9 Missing	0.37	0.17	0.64	0.10
SP census division				
1 Northeast Region: New England Division				
2 Northeast Region: Middle Atlantic Division	0.14	0.12	0.19	0.10
3 Midwest Region: East North Central Division	0.03	0.11	0.15	0.09
4 Midwest Region: West North Central Division	0.02	0.12	0.14	0.09
5 South Region: South Atlantic Division	0.00	0.11	0.11	0.09
6 South Region: East South Central Division	-0.16	0.13	-0.01	0.11
7 South Region: West South Central Division	-0.01	0.12	0.10	0.10
8 West Region: Mountain Division	0.09	0.14	0.25	0.15
9 West Region: Pacific Division	0.10	0.11	0.11	0.09
SP metro / non-metro residence				
1 Metropolitan				
2 Non-metropolitan	-0.06	0.05	-0.02	0.05
CG relationship to SP				
2 Spouse/partner of SP				
3 Children			0.44	0.07
4 Other relatives	-0.18	0.05	0.14	0.09
5 Non-relatives	-0.07	0.07	0.16	0.09
SP's income from NHATS (log-transformed)	0.04	0.03	0.27	0.03
CG in household with SP				

1 YES				
2 NO	0.00	0.05	0.15	0.05
CG has private insurance coverage				
No				
Yes	0.24	0.05		
CG has Medicaid coverage				
No				
Yes	-0.37	0.07		
CG owns home				
No				
Yes	0.21	0.05		
CG has checking acct				
No				
Yes	0.27	0.08		
CG has savings account				
No				
Yes	0.12	0.06		
CG has certificate of deposit				
No				
Yes	0.18	0.05		
CG has retirement plan				
No				
Yes	0.35	0.05		
CG has stocks or mutual funds				
No				
Yes	0.15	0.05		
CG worked for pay last week				
Yes				
No	-0.26	0.05		
Retired	-0.23	0.07		
CG owns business or farm				
No				
Yes	0.13	0.05		
Constant	8.72	0.33	6.05	0.38
Log of sigma	-0.58	0.04	-0.49	0.03
Sigma	0.56	0.02	0.61	0.02

Appendix Table 6. Model results for NSOC III (Round 7) longitudinal sample by imputation groups

	Grp 1: Non-Spouse/ Partner		Grp 3: Breakoff		Grp 4: LML	
	Coef.	SE	Coef.	SE	Coef.	SE
CG Age	0.00	0.00	0.00	0.00	0.00	0.00
CG Race / ethnicity						
1 White, non-Hispanic						
2 Black, non-Hispanic	-0.07	0.06	-0.06	0.05	0.07	0.06
3 Others/Missing	-0.01	0.07	-0.06	0.08	0.00	0.06
4 Hispanic	-0.08	0.09	-0.06	0.10	-0.02	0.08
CG Gender						
1 MALE						
2 FEMALE	-0.03	0.05	-0.02	0.05	0.03	0.04
Number of other people live in CG household	-0.02	0.02	-0.04	0.02	-0.03	0.02
CG education						
1 <HS						
2 HS	-0.09	0.09	0.06	0.07	-0.01	0.07
3 >HS, <College	-0.02	0.11	0.15	0.08	0.06	0.07
4 College+	0.00	0.11	0.19	0.08	0.09	0.07
CG spouse/partner's education						
0 No spouse						
1 <HS	0.29	0.15	0.41	0.08	0.36	0.08
2 HS	0.38	0.08	0.44	0.07	0.34	0.05
3 >HS, <College	0.46	0.07	0.52	0.08	0.44	0.06
4 College+	0.54	0.07	0.63	0.08	0.56	0.06
9 Missing	0.13	0.25	0.22	0.25	0.14	0.23
SP census division						
1 Northeast Region: New England Division						
2 Northeast Region: Middle Atlantic Division	0.40	0.13	0.26	0.12	0.16	0.10
3 Midwest Region: East North Central Division	0.32	0.09	0.24	0.08	0.10	0.07
4 Midwest Region: West North Central Division	0.31	0.10	0.25	0.09	0.11	0.08
5 South Region: South Atlantic Division	0.25	0.10	0.23	0.09	0.03	0.08
6 South Region: East South Central Division	0.23	0.13	0.20	0.11	-0.01	0.10
7 South Region: West South Central Division	0.31	0.11	0.26	0.09	0.06	0.09
8 West Region: Mountain Division	0.18	0.12	0.21	0.10	0.20	0.10
9 West Region: Pacific Division	0.18	0.11	0.11	0.10	0.02	0.09

SP metro / non-metro residence						
1 Metropolitan						
2 Non-metropolitan	-0.07	0.05	0.02	0.05	-0.01	0.04
CG relationship to SP						
2 Spouse/partner						
3 Children			0.25	0.10	0.20	0.05
4 Other relatives	-0.01	0.07	0.23	0.13	0.17	0.08
5 Non-relatives	-0.11	0.06	0.04	0.11	0.03	0.06
SP's income from NHATS (log-transformed)	0.08	0.03	0.16	0.04	0.14	0.03
CG income from round 5	0.35	0.06	0.51	0.05	0.37	0.05
CG in household with SP						
1 YES						
2 NO	0.01	0.06	0.02	0.07		
CG has private insurance coverage						
No						
Yes	0.05	0.07			0.09	0.04
CG has Medicaid coverage						
No						
Yes	-0.30	0.08			-0.30	0.07
CG owns home						
No						
Yes	0.21	0.06			0.14	0.05
CG has checking acct						
No						
Yes	0.21	0.14			0.23	0.13
CG has savings account						
No						
Yes	-0.12	0.08			0.02	0.06
CG has certificate of deposit						
No						
Yes	0.07	0.06			0.03	0.04
CG has retirement plan						
No						
Yes	0.19	0.07			0.21	0.05
CG has stocks or mutual funds						
No						
Yes	0.08	0.05			0.03	0.04
CG worked for pay last week						
Yes						
No	-0.32	0.06				
Retired	-0.36	0.08				

CG owns business or farm						
No						
Yes	0.02	0.06				
Constant	5.38	0.58	3.06	0.55	4.50	0.52
Log of sigma	-0.78	0.07	-0.69	0.07	-0.76	0.06
Sigma	0.46	0.03	0.50	0.03	0.47	0.03