NHATS Technical Paper #31

NATIONAL HEALTH AND AGING TRENDS STUDY (NHATS)
Round 11 Income Imputation

November 2022

Suggested Citation: Jiao, Rui, Freedman, Vicki A., Kasper, Judith D., and Schrack, Jennifer. 2022. National Health and Aging Trends Study Round 11 Income Imputation. NHATS Technical Paper #31. Johns Hopkins University School of Public Health. Available at www.NHATS.org. We thank Benjamin Schneider, who played an instrumental role in the development of the Round 11 Income imputation. This technical paper was prepared with funding from the National Institute on Aging (U01AG032947).

Overview

In preparing survey data files for analysis, imputation is often used to address item nonresponse, particularly when complex multi-variate recodes are required that are built up from a collection of more detailed questions (Marker, Judkins, and Winglee, 2001). Rounds 1, 3, 5, 7, 9 and 11 of the National Health and Aging Trends Survey (NHATS) include imputed values for total income. Both a continuous measure and a bracket value are provided, with separate bracket values for single respondents and those who are married or are living with a partner. We used a cyclical n-partition hot deck (see Judkins 1997) to generate five imputations of each measure. This technical paper provides details on the imputation strategy.

Income Sources Collected in NHATS

Rounds 1 and 5 of NHATS collected information on sources of income (yes/no) and amounts for each source. Rounds 3, 7, 9 and 11 also collected information on sources of income (yes/no) but not amounts for each source. Respondents with a spouse/partner were given the option of reporting sources for themselves either together or separately from their spouse/partner. Table 1 shows the income sources included in NHATS Round 11.

Table 1. Summary of Income Sources Collected in NHATS Round 11

Sources of Income	Time frame
Social Security	Last Month
Supplemental Security Income	Last Month
Veteran's Administration	Last Month
Pension plan	--
Earned Income	Last Month
Interest/dividend income from any: mutual funds/stocks, bonds, bank accounts, or CDs	--
Real estate	--
Retirement account withdrawals	--
Total income from all sources	Last Year

--Asked about the existence of relevant asset

Extent of Missing Data for Total Income

In Round 11, 69% of the sample provided a total income amount and an additional 17% reported a bracketed value instead. Thus, a bracketed value could be created from reports for 86% of the sample and required imputation for 14%. An exact value was then imputed for 31% of the sample (17% within a reported bracketed value and 14% within an imputed bracketed value).

Imputation Methodology

Westat's AutoImpute software was used to impute five values of the total income items. AutoImpute uses a cyclical n-partition hot deck (an approach analogous to the Gibbs sampler but using the hot deck to generate the imputations; see Judkins 1997; Judkins et al. 2007; Judkins, Piesse, and Krenzke 2008; Krenzke and Judkins 2008). This software is designed to facilitate preservation of multivariate distributions while also ensuring that imputations maintain skip patterns and adhere to constraints. In this application an example of a constraint is ensuring imputations for specific amounts fall within reported (or imputed) bracket ranges.

The cyclical n-partition hot deck procedure initially imputes all target variables (i.e., items requiring imputation) using a simple hot deck that uses specified auxiliary variables and skip controllers. Using the

initial imputed variables, a model is fit for each target variable using simple forward stepwise regression selection. The predicted values of the target variable from the final model are used to generate imputed values by using predictive mean matching for ordinal (ordered categorical) target variables and clustering for unordered categorical target variables. Predictive mean matching uses a hot deck with the skip controllers as hard boundaries and the predicted values from the final model as soft boundaries. For unordered categorical target variables, a k-means clustering algorithm is used on the vector of predicted values for each level, and then a hot deck is used to impute the target variable with the skip controllers as hard boundaries and the cluster membership indicators as soft boundaries. For more details on the procedure see Judkins et al. (2007).

Variables Used in Imputation

Three classes of variables were used in the imputation (see Appendix tables):

1. Source variables that indicate (yes/no) whether the respondent (and his/her spouse/partner, if applicable) has the particular source of income (referred to below as "fencepost" variables);
2. Auxiliary variables that included respondent characteristics (e.g., age, race/ethnicity, gender, educational attainment, home ownership (in Round 11), veteran's status, labor force status (in Round 11), spouse/partner's labor force status (in Round 11), interviewer observations about the home condition and skip pattern controllers; and
3. Total income (reported or imputed), as well as source variables, from Round 9.

In order to preserve the joint distribution of the full set of income variables, all missing fencepost, auxiliary, and total income variables were imputed.

When imputing total income variables, both the Round 9 and Round 11 source variables and Round 9 total income were used, along with the auxiliary variables. Total income value was also constrained to fall within the reported/imputed bracket amount.

Income Imputation Variables in the SP File

The following imputed variables are included on the SP data file:

Variable name	Label	Description
2020 Total Income Value		
ia11totinc	R11 IA50 TOTAL INCOME	Actual reported $ amounts
ia11toincimf	R11 F IMPUTED TOTAL INC FLG	Flag indicating imputation
ia11toincim1-5	R11 IA50 IMPUTED TOTAL INC1-INC5	Imputed values 1-5 for missing $ amounts and reported bracket amounts
2020 Total Income Range – Respondents who have spouse/partner		
ia11toincesjt	R11 IA51A JOINT EST TOT INCOME	Actual reported $ amounts
ia11eincimjf	R11 F IMPTD JOINT EST TOT INC FLG	Flag indicating imputation
ia11eincimj1-5	R11 IA51A IMP EST JOINT TOT INC1-INC5	Imputed values 1-5 for missing $ amounts and reported bracket amounts
2020 Total Income Range – Respondents who are single		
ia11toincessg	R11 IA51B SNGLE EST TOT INC	Actual reported $ amounts
ia11eincimsf	R11 F IMPUTED SGL EST TOT INC FLG	Flag indicating imputation
ia11eincims1-5	R11 IA51B IMP EST SP SGL TOT INC1-INC5	Imputed values 1-5 for missing $ amounts and reported bracket amounts

Using the Five Versions of the Imputed Variable in Analysis

For each of the three total income variables that was imputed, five sets of imputed variables were generated. For item nonrespondents, the five sets contain five independently generated imputed values. These five sets of imputed variables are provided to enable data users to use multiple imputation variance estimators and analysis techniques (see, for example, Rubin 1996) to account for the effects of item nonresponse and imputation error in variance estimates for analyses that use these income variables. In Round 11, item respondents with actual reported $ amounts were not included in the 5 sets of imputed variables.

Because Round 9 variables were used in the imputation of Round 11 variables, in order to capture the effects of imputation of the Round 9 variables on the precision of estimates involving the Round 11 variables, the five sets of imputed values for the Round 9 variables were used to impute the five sets of imputed values for the Round 11 variables.

References

Judkins, D. (1997). Imputing for Swiss cheese patterns of missing data. *Proceedings of Statistics Canada Symposium '97*, 143-148.

Judkins, D., Krenzke, T., Piesse, A., Fan, Z., and Haung, W.C. (2007). Preservation of skip patterns and covariate structure through semi-parametric whole questionnaire imputation. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 3211-3218.

Judkins, D., Piesse, A., and Krenzke, T. (2008). Multiple semi-parametric imputation. *Proceedings of the Section on Survey Research Methods of the American Statistical Association.*

Krenzke, T. and Judkins, D. (2008). Filling in blanks: Some guesses are better than others – Illustrating the impact of covariate selection when imputing complex survey items. *Chance*, 21(3), 7-13.

Marker, D. A., Judkins, D. R., and Winglee, M. (2001). Large-scale imputation for complex surveys. In *Survey Nonresponse*, Eds. R. M. Groves, D. A. Dillman, E. L. Eltinge, and R. J. A. Little. New York: John Wiley & Sons.

Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association,* 91(434), 473-489.

Appendix. Lists of Variables Used in NHATS Round 11 Income Imputation

Table 1. Round 11 Source ("Fencepost") Variables

#	Variable name	Label	% missing
1	ia11recsspa1	R11 IA1 SP REC SOCIAL SECURITY	1.6
2	ia11recsspa2	R11 IA1 SPOUSE PART REC SOC SEC	1.6
3	ia11recsspa3	R11 IA1 NO SOC SECURTY PAYMNT REC	1.6
4	ia11recssils1	R11 IA4 SP RECEIVD SSI LAST MONTH	2.6
5	ia11recssils2	R11 IA4 SPOUSE PRT REC SSI LST MO	2.6
6	ia11recssils3	R11 IA4 NO SSI RECEIVD LAST MONTH	2.6
7	ia11rvapayls1	R11 IA5 SP REC PAY FRM VA LAST MO	1.2
8	ia11rvapayls2	R11 IA5 SPOUS PA REC VA PAY LSTMO	1.2
9	ia11rvapayls3	R11 IA5 NO VA PAY REC LAST MONTH	1.2
10	ia11penjobou1	R11 IA6 SP HAS PENSION PLAN	2.0
11	ia11penjobou2	R11 IA6 SPOUSE HAS PENSION PLAN	2.0
12	ia11penjobou3	R11 IA6 NO PENSION PLAN	2.0
13	ia11iraothac1	R11 IA7 SP HAS IRA OTH RETIRE ACC	3.7
14	ia11iraothac2	R11 IA7 SPOUSE HAS IRA OTHR ACC	3.7
15	ia11iraothac3	R11 IA7 NO IRA OTHR RETIRE ACCT	3.7
16	ia11mutfdstk1	R11 IA8 SP OWNS MUTUAL FUND STOCK	3.9
17	ia11mutfdstk2	R11 IA8 SPOUSE OWNS FUNDS STOCK	3.9
18	ia11mutfdstk3	R11 IA8 SP SPOUSE OWN FUNDS STOCK	3.9
19	ia11mutfdstk4	R11 IA8 NO FUNDS OR STOCK OWNED	3.9
20	ia11ownbond1	R11 IA11 SP OWNS BONDS	3.7
21	ia11ownbond2	R11 IA11 SPOUSE OWNS BONDS	3.7
22	ia11ownbond3	R11 IA11 SP SPOUSE OWN BONDS	3.7
23	ia11ownbond4	R11 IA11 NO BONDS OWNED	3.7
24	ia11bnkacccd1	R11 IA10 SP OWNS CHECK ACCT	1.7
25	ia11bnkacccd2	R11 IA10 SPOUSE OWNS CHECK ACCT	1.7
26	ia11bnkacccd3	R11 IA10 SP SPOUSE OWN CHECK ACCT	1.7
27	ia11bnkacccd4	R11 IA10 NO CHECK ACCT OWNED	1.7
28	ia11bnkacccd5	R11 IA10 SP OWNS SAVINGS ACCT	3.3
29	ia11bnkacccd6	R11 IA10 SPOUSE OWNS SAVING ACCT	3.3
30	ia11bnkacccd7	R11 IA10 SP SPOUSE OWN SAVNG ACT	3.3
31	ia11bnkacccd8	R11 IA10 NO SAVINGS ACCT OWNED	3.3
32	ia11bnkacccd9	R11 IA10 SP OWNS CDS	4.7
33	ia11bnkaccc10	R11 IA10 SPOUSE OWNS CDS	4.7
34	ia11bnkaccc11	R11 IA10 SP SPOUSE OWN CDS	4.7
35	ia11bnkaccc12	R11 IA10 NO CDS OWNED	4.7
36	ia11realestt1	R11 IA13 SP OWNS REAL ESTATE	1.7
37	ia11realestt2	R11 IA13 SPOUSE OWNS REAL ESTATE	1.7
38	ia11realestt3	R11 IA13 SP SPOUSE OWN REAL ESTTE	1.7
39	ia11realestt4	R11 IA13 NO REAL ESTATE OWNED	1.7
40	lf11workfpay	R11 LF1 WORKED FOR PAY RECENTLY	0.1
41	lf11abstlstwk	R11 LF2 ABSENT FRM JOB LAST WEEK	0.0
42	lf11wrkplstmn	R11 LF3 WORK FOR PAY IN LST MONTH	0.0
43	lf11huswifwrk	R11 LF13 HUSB WIFE PARTN PAY WORK	0.6

Table 2. Auxiliary Variables

	Variable name	Label	% Missing
1	SEX	GENDER	-[1]
2	AGECAT_R5	AGE CATEGORY AS OF 2014	-
3	RTIRACE	RACEETH, 3-CATEGORY	-
4	PER_CAP_INC_5YR	PER CAPITA INCOME [ACS]	-
5	SMPTYPE	SAMPLE TYPE (O=ORIGINAL SMP, R=REPLENISHMENT SMP)	-
6	el5dhigstsch[2]	R5 EL10 D HGHST DGREE SCOOL COMPLD	-
7	rl5dracehisp	R5 D RACE AND HISPANIC ETHNICITY	-
8	va5serarmfor	R5 VA1 SERVED IN ARMED FORCES	-
9	va5memnatgrd	R5 VA3 MEMBER OF NATIONAL GUARD	-
10	fl11facility	R11 F ROUTING FLAG FROM RE4f HT3 5 6 7	-
11	ir11areacond1[3]	R11 IR15 LITTER GLASS ON SDWLK ST	7.5
12	ir11areacond2	R11 IR15 GRAFFITI ON BUILDG WALLS	7.5
13	ir11areacond3	R11 IR15 VACANT HOUSES OR STORES	7.5
14	ir11condhome1	R11 IR16 BROKEN WINDOWS IN HOME	7.5
15	ir11condhome2	R11 IR16 CRUMBLNG FOUNDTN IN HOME	7.5
16	ir11condhome3	R11 IR16 MISSNG BRCKS SIDNG IN HM	7.5
17	ir11condhome4	R11 IR16 ROOF PROBLEM IN HOME	7.5
18	ir11condhome5	R11 IR16 BROKEN STEPS TO HOME	7.5
19	ir11condhome6	R11 IR16 CONTINUOUS SIDEWALKS	7.5
20	hh11dmarstat	R11 D MARITAL STATUS AT R11	-
21	hh11livwthspo	R11 HH11 LIVE WITH SPOUSE PARTNER	-
22	hh11placekind	R11 HH12 KIND OF PLACE LIVE IN	3.0
23	pa11workfrpay	R11 PA17 EVER WORK FOR PAY	0.1
24	lf11mrthnonjb	R11 LF4 MOR THN ONE JOB LAST WEEK	-
25	lf11hrswkwork	R11 LF5 HRS PR WEEK WORK MAIN JOB	0.4
26	lf11hrwrkltwk	R11 LF6 HOURS WORK LAST WEEK	1.6
27	lf11hrwrklstw	R11 LF7 HOW MNY HOURS DID YOU WRK	-
28	lf11oftpaid	R11 LF8 HOW OFTN PAID ON MAIN JOB	0.3
29	hp11ownrentot	R11 HP1 OWN RENT OR OTHER	0.4

[1] A value of "-" is used when there were no missing values.

[2] Equal to EL1HIGSTSCH for original sample cases and EL5HIGSTSCH for replenishment sample cases.

[3] The missingness for IR11AREACOND1-IR11AREACOND3, and IR11CONDHOME1-IR11CONDHOME6 are contributed by either the completed interviews conducted by telephone or the partial completed interviews.

Table 3. Round 9 Source ("Fencepost") Variables[4]

#	Variable name	Label
1	ia9recsspa1	R9 IA1 SP REC SOCIAL SECURITY
2	ia9recsspa2	R9 IA1 SPOUSE PART REC SOC SEC
3	ia9recsspa3	R9 IA1 NO SOC SECURTY PAYMNT REC
4	ia9recssils1	R9 IA4 SP RECEIVD SSI LAST MONTH
5	ia9recssils2	R9 IA4 SPOUSE PRT REC SSI LST MO
6	ia9recssils3	R9 IA4 NO SSI RECEIVD LAST MONTH
7	ia9rvapayls1	R9 IA5 SP REC PAY FRM VA LAST MO
8	ia9rvapayls2	R9 IA5 SPOUS PA REC VA PAY LSTMO
9	ia9rvapayls3	R9 IA5 NO VA PAY REC LAST MONTH
10	ia9penjobou1	R9 IA6 SP HAS PENSION PLAN
11	ia9penjobou2	R9 IA6 SPOUSE HAS PENSION PLAN
12	ia9penjobou3	R9 IA6 NO PENSION PLAN
13	ia9iraothac1	R9 IA7 SP HAS IRA OTH RETIRE ACC
14	ia9iraothac2	R9 IA7 SPOUSE HAS IRA OTHR ACC
15	ia9iraothac3	R9 IA7 NO IRA OTHR RETIRE ACCT
16	ia9mutfdstk1	R9 IA8 SP OWNS MUTUAL FUND STOCK
17	ia9mutfdstk2	R9 IA8 SPOUSE OWNS FUNDS STOCK
18	ia9mutfdstk3	R9 IA8 SP SPOUSE OWN FUNDS STOCK
19	ia9mutfdstk4	R9 IA8 NO FUNDS OR STOCK OWNED
20	ia9ownbond1	R9 IA9 SP OWNS BONDS
21	ia9ownbond2	R9 IA9 SPOUSE OWNS BONDS
22	ia9ownbond3	R9 IA9 SP SPOUSE OWN BONDS
23	ia9ownbond4	R9 IA9 NO BONDS OWNED
24	ia9bnkacccd1	R9 IA10 SP OWNS CHECK ACCT
25	ia9bnkacccd2	R9 IA10 SPOUSE OWNS CHECK ACCT
26	ia9bnkacccd3	R9 IA10 SP SPOUSE OWN CHECK ACCT
27	ia9bnkacccd4	R9 IA10 NO CHECK ACCT OWNED
28	ia9bnkacccd5	R9 IA10 SP OWNS SAVINGS ACCT
29	ia9bnkacccd6	R9 IA10 SPOUSE OWNS SAVING ACCT
30	ia9bnkacccd7	R9 IA10 SP SPOUSE OWN SAVNG ACT
31	ia9bnkacccd8	R9 IA10 NO SAVINGS ACCT OWNED
32	ia9bnkacccd9	R9 IA10 SP OWNS CDS
33	ia9bnkaccc10	R9 IA10 SPOUSE OWNS CDS
34	ia9bnkaccc11	R9 IA10 SP SPOUSE OWN CDS
35	ia9bnkaccc12	R9 IA10 NO CDS OWNED
36	ia9realestt1	R9 IA13 SP OWNS REAL ESTATE
37	ia9realestt2	R9 IA13 SPOUSE OWNS REAL ESTATE
38	ia9realestt3	R9 IA13 SP SPOUSE OWN REAL ESTTE
39	ia9realestt4	R9 IA13 NO REAL ESTATE OWNED
40	lf9workfpay	R9 LF1 WORKED FOR PAY RECENTLY
41	lf9abstlstwk	R9 LF2 ABSENT FRM JOB LAST WEEK
42	lf9wrkplstmn	R9 LF3 WORK FOR PAY IN LST MONTH
43	lf9huswifwrk	R9 LF13 HUSB/WIFE/PARTN PAY WORK
44	ia9totinc	R9 IA50 TOTAL INCOME

[4] Round 9 source variables are not available for 3 R9 respondents who had completed facility questionnaires only, so their imputation process only used Table 1 and Table 2 variables.