NHATS Technical Paper #10

NATIONAL HEALTH AND AGING TRENDS STUDY (NHATS)
Round 3 Income Imputation

April 24, 2015

Suggested Citation: Montaquila Jill, Freedman Vicki A., and Kasper, Judith D. 2015. National Health and Aging Trends Study Round 3 Income Imputation. NHATS Technical Paper #10. 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).

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 and 3 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

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

Table 1. Summary of Income Sources Collected in NHATS Round 3

Sources of Income	Time frame
Social Security	Last Month
Supplemental Security Income	Last Month
Veteran's Administration	Last Month
Pension plan	Last Month
Earned Income	Last Month/Last Paycheck
Interest/dividend income from any: mutual funds/stocks, bonds, bank accounts, or CDs[1]	Last Year
Retirement account withdrawals[1]	Last month/Last year
Total income from all sources	Last Year

[1]Asked after questions about the existence of relevant asset

After reporting about the receipt of income from each individual source, respondents are asked to report total income from all sources. Respondents who report don't know or who refuse are offered a set of five bracketed ranges. These ranges were developed from reports for the 65 and older population in the Survey of Consumer Finance for 2007 adjusted to 2011 dollars. Separate ranges were provided for individually reported (respondent or spouse/partner amounts separately) and for jointly reported (respondent and spouse/partner together) amounts.

Extent of Missing Data for Total Income

Fifty-nine percent of the sample provided a total income amount and an additional 22% reported a bracketed value instead. Thus, a bracketed value could be created from reports for 81% of the sample and required imputation for 19%. An exact value was then imputed for 41% of the sample (22% within a reported bracketed value and 19% 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 3), veteran's status, labor force status (in Round 3), spouse/partner's labor force status (in Round 3), interviewer observations about the home condition) and skip pattern controllers; and
3. Total income (reported or imputed), as well as source variables, from Round 1.

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 1 and Round 3 source variables and Round 1 total income were used, along with the auxiliary variables. Total income value was also constrained to fall within the reported/imputed bracket amount.

Because Round 1 variables were used in the imputation of Round 3 variables, in order to capture the effects of imputation of the Round 1 variables on the precision of estimates involving the Round 3 variables, the five sets of imputed values for the Round 1 variables were used to impute the five sets of imputed values for the Round 3 variables.

Income Imputation Variables in the SP File

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

Variable name	Label	Description
2012 Total Income Value		
Ia3totinc	R3 IA50 TOTAL INCOME	Actual
Ia3toincimf	R3 F IMPUTED TOTAL INC FLG	Flag indicating imputation
Ia3toincim1-5	R3 IA50 IMPUTED TOTAL INC1-INC5	Imputed values 1-5
2012 Total Income Range – Respondents who have spouse/partner		
Ia3toincesjt	R3 IA51A JOINT EST TOT INCOME	Actual
Ia3eincimjf	R3 F IMPTD JOINT EST TOT INC FLG	Flag indicating imputation
Ia3eincimj1-5	R3 IA51A IMP EST JOINT TOT INC1-INC5	Imputed values 1-5
2012 Total Income Range – Respondents who are single		
Ia3toincessg	R3 IA51B SNGLE EST TOT INC	Actual
Ia3eincimsf	R3 F IMPUTED SGL EST TOT INC FLG	Flag indicating imputation
Ia3eincims1-5	R3 IA51B IMP EST SP SGL TOT INC1-INC5	Imputed values 1-5

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 respondents, these five sets contain copies of the reported data. 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.

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 3 Income Imputation

Table 1. Round 3 Source ("Fencepost") Variables

#	Variable name	Label	% missing
1	ia3recsspa1	R3 IA1 SP REC SOCIAL SECURITY	3.5
2	ia3recsspa2	R3 IA1 SPOUSE PART REC SOC SEC	3.5
3	ia3recsspa3	R3 IA1 NO SOC SECURTY PAYMNT REC	3.5
4	ia3recssils1	R3 IA4 SP RECEIVD SSI LAST MONTH	4.3
5	ia3recssils2	R3 IA4 SPOUSE PRT REC SSI LST MO	4.3
6	ia3recssils3	R3 IA4 NO SSI RECEIVD LAST MONTH	4.3
7	ia3rvapayls1	R3 IA5 SP REC PAY FRM VA LAST MO	3.8
8	ia3rvapayls2	R3 IA5 SPOUS PA REC VA PAY LSTMO	3.8
9	ia3rvapayls3	R3 IA5 NO VA PAY REC LAST MONTH	3.8
10	ia3penjobou1	R3 IA6 SP HAS PENSION PLAN	3.9
11	ia3penjobou2	R3 IA6 SPOUSE HAS PENSION PLAN	3.9
12	ia3penjobou3	R3 IA6 NO PENSION PLAN	3.9
13	ia3iraothac1	R3 IA7 SP HAS IRA OTH RETIRE ACC	6.4
14	ia3iraothac2	R3 IA7 SPOUSE HAS IRA OTHR ACC	6.4
15	ia3iraothac3	R3 IA7 NO IRA OTHR RETIRE ACCT	6.4
16	ia3mutfdstk1	R3 IA8 SP OWNS MUTUAL FUND STOCK	5.2
17	ia3mutfdstk2	R3 IA8 SPOUSE OWNS FUNDS STOCK	5.2
18	ia3mutfdstk3	R3 IA8 SP SPOUSE OWN FUNDS STOCK	5.2
19	ia3mutfdstk4	R3 IA8 NO FUNDS OR STOCK OWNED	5.2
20	ia3ownbond1	R3 IA9 SP OWNS BONDS	5.7
21	ia3ownbond2	R3 IA9 SPOUSE OWNS BONDS	5.7
22	ia3ownbond3	R3 IA9 SP SPOUSE OWN BONDS	5.7
23	ia3ownbond4	R3 IA9 NO BONDS OWNED	5.7
24	ia3bnkacccd1	R3 IA10 SP OWNS CHECK ACCT	4.0
25	ia3bnkacccd2	R3 IA10 SPOUSE OWNS CHECK ACCT	4.0
26	ia3bnkacccd3	R3 IA10 SP SPOUSE OWN CHECK ACCT	4.0
27	ia3bnkacccd4	R3 IA10 NO CHECK ACCT OWNED	4.0
28	ia3bnkacccd5	R3 IA10 SP OWNS SAVINGS ACCT	5.0
29	ia3bnkacccd6	R3 IA10 SPOUSE OWNS SAVING ACCT	5.0
30	ia3bnkacccd7	R3 IA10 SP SPOUSE OWN SAVNG ACT	5.0
31	ia3bnkacccd8	R3 IA10 NO SAVINGS ACCT OWNED	5.0
32	ia3bnkacccd9	R3 IA10 SP OWNS CDS	6.1
33	ia3bnkacccd10	R3 IA10 SPOUSE OWNS CDS	6.1
34	ia3bnkacccd11	R3 IA10 SP SPOUSE OWN CDS	6.1
35	ia3bnkacccd12	R3 IA10 NO CDS OWNED	6.1
36	ia3realestt1	R3 IA13 SP OWNS REAL ESTATE	3.7
37	ia3realestt2	R3 IA13 SPOUSE OWNS REAL ESTATE	3.7
38	ia3realestt3	R3 IA13 SP SPOUSE OWN REAL ESTTE	3.7
39	ia3realestt4	R3 IA13 NO REAL ESTATE OWNED	3.7
40	lf3workfpay	R3 LF1 WORKED FOR PAY RECENTLY	1.9
41	lf3abstlstwk	R3 LF2 ABSENT FRM JOB LAST WEEK	5.8
42	lf3wrkplstmn	R3 LF3 WORK FOR PAY IN LST MONTH	0.2
43	lf3huswifwrk	R3 LF13 HUSB WIFE PARTN PAY WORK	2.2

Table 2. Auxilliary Variables

	Variable name	Label	% Missing
1	Sex	HISKEW GENDER	-
2	agecat	HISKEW AGE CATEGORY	-
3	rtirace	HISKEW RACEETH, 3-CATEGORY	-
4	per_cap_inc_5yr	PER CAPITA INCOME [WT1; ACS]	-
5	el1higstschl	R1 EL10 HGHST DGREE SCOOL COMPLD	-
6	rl1hisplatno	R1 RL3 CNSDR YRSF HSPAN OR LATNO	-
7	rl1yourrace1	R1 RL1 RACE OF SP WHITE	-
8	rl1yourrace2	R1 RL1 RACE OF SP AFRICN AMERICN	-
9	rl1yourrace3	R1 RL1 RACE OF SP AMERICN INDIAN	-
10	rl1yourrace4	R1 RL1 RACE OF SP ALASKA NATIVE	-
11	rl1yourrace5	R1 RL1 RACE OF SP ASIAN	-
12	rl1yourrace6	R1 RL1 RACE OF SP NATIVE HWAIIAN	-
13	rl1yourrace7	R1 RL1 RACE OF SP PACIFIC ISLNDR	-
14	rl1yourrace8	R1 RL1 RACE OF SP OTHER SPECIFY	-
15	va1serarmfor	R1 VA1 SERVED IN ARMED FORCES	-
16	va1memnatgrd	R1 VA3 MEMBER OF NATIONAL GUARD	-
17	fl3facility	R3 F ROUTING FLAG FROM RE4f HT3 5 6 7	0.0
18	ir3areacond1	R3 IR15 LITTER GLASS ON SDWLK ST	0.9
19	ir3areacond2	R3 IR15 GRAFFITI ON BUILDG WALLS	0.9
20	ir3areacond3	R3 IR15 VACANT HOUSES OR STORES	0.9
21	ir3condhome1	R3 IR16 BROKEN WINDOWS IN HOME	0.9
22	ir3condhome2	R3 IR16 CRUMBLNG FOUNDTN IN HOME	0.9
23	ir3condhome3	R3 IR16 MISSNG BRCKS SIDNG IN HM	0.9
24	ir3condhome4	R3 IR16 ROOF PROBLEM IN HOME	0.9
25	ir3condhome5	R3 IR16 BROKEN STEPS TO HOME	0.9
26	ir3condhome6	R3 IR16 CONTINUOUS SIDEWALKS	0.9
27	hh3dmarstat	R3 D MARITAL STATUS AT R3	0.0
28	hh3livwthspo	R3 HH11 LIVE WITH SPOUSE PARTNER	0.0
29	hh3placekind	R3 HH12 KIND OF PLACE LIVE IN	1.6
30	pa3workfrpay	R3 PA17 EVER WORK FOR PAY	0.1
31	lf3mrthnonjb	R3 LF4 MOR THN ONE JOB LAST WEEK	0.0
32	lf3hrswkwork	R3 LF5 HRS PR WEEK WORK MAIN JOB	1.5
33	lf3hrwrkltwk	R3 LF6 HOURS WORK LAST WEEK	1.4
34	lf3hrwrklstw	R3 LF7 HOW MNY HOURS DID YOU WRK	0.0
35	lf3oftpaid	R3 LF8 HOW OFTN PAID ON MAIN JOB	0.3
36	hp3ownrentot	R3 HP1 OWN RENT OR OTHER	2.1

Table 3. Round 1 Source ("Fencepost") Variables

#	Variable name	Label
1	ia1recsspa1	R1 IA1 SP REC SOCIAL SECURITY
2	ia1recsspa2	R1 IA1 SPOUSE PART REC SOC SEC
3	ia1recsspa3	R1 IA1 NO SOC SECURTY PAYMNT REC
4	ia1recssils1	R1 IA4 SP RECEIVD SSI LAST MONTH
5	ia1recssils2	R1 IA4 SPOUSE PRT REC SSI LST MO
6	ia1recssils3	R1 IA4 NO SSI RECEIVD LAST MONTH
7	ia1rvapayls1	R1 IA5 SP REC PAY FRM VA LAST MO
8	ia1rvapayls2	R1 IA5 SPOUS PA REC VA PAY LSTMO
9	ia1rvapayls3	R1 IA5 NO VA PAY REC LAST MONTH
10	ia1penjobou1	R1 IA6 SP HAS PENSION PLAN
11	ia1penjobou2	R1 IA6 SPOUSE HAS PENSION PLAN
12	ia1penjobou3	R1 IA6 NO PENSION PLAN
13	ia1iraothac1	R1 IA7 SP HAS IRA OTH RETIRE ACC
14	ia1iraothac2	R1 IA7 SPOUSE HAS IRA OTHR ACC
15	ia1iraothac3	R1 IA7 NO IRA OTHR RETIRE ACCT
16	ia1mutfdstk1	R1 IA8 SP OWNS MUTUAL FUND STOCK
17	ia1mutfdstk2	R1 IA8 SPOUSE OWNS FUNDS STOCK
18	ia1mutfdstk3	R1 IA8 SP SPOUSE OWN FUNDS STOCK
19	ia1mutfdstk4	R1 IA8 NO FUNDS OR STOCK OWNED
20	ia1ownbond1	R1 IA9 SP OWNS BONDS
21	ia1ownbond2	R1 IA9 SPOUSE OWNS BONDS
22	ia1ownbond3	R1 IA9 SP SPOUSE OWN BONDS
23	ia1ownbond4	R1 IA9 NO BONDS OWNED
24	ia1bnkacccd1	R1 IA10 SP OWNS CHECK ACCT
25	ia1bnkacccd2	R1 IA10 SPOUSE OWNS CHECK ACCT
26	ia1bnkacccd3	R1 IA10 SP SPOUSE OWN CHECK ACCT
27	ia1bnkacccd4	R1 IA10 NO CHECK ACCT OWNED
28	ia1bnkacccd5	R1 IA10 SP OWNS SAVINGS ACCT
29	ia1bnkacccd6	R1 IA10 SPOUSE OWNS SAVING ACCT
30	ia1bnkacccd7	R1 IA10 SP SPOUSE OWN SAVNG ACT
31	ia1bnkacccd8	R1 IA10 NO SAVINGS ACCT OWNED
32	ia1bnkacccd9	R1 IA10 SP OWNS CDS
33	ia1bnkaccc10	R1 IA10 SPOUSE OWNS CDS
34	ia1bnkaccc11	R1 IA10 SP SPOUSE OWN CDS
35	ia1bnkaccc12	R1 IA10 NO CDS OWNED
36	ia1realestt1	R1 IA13 SP OWNS REAL ESTATE
37	ia1realestt2	R1 IA13 SPOUSE OWNS REAL ESTATE
38	ia1realestt3	R1 IA13 SP SPOUSE OWN REAL ESTTE
39	ia1realestt4	R1 IA13 NO REAL ESTATE OWNED
40	lf1workfpay	R1 LF1 WORKED FOR PAY RECENTLY
41	lf1abstlstwk	R1 LF2 ABSENT FRM JOB LAST WEEK
42	lf1wrkplstmn	R1 LF3 WORK FOR PAY IN LST MONTH
43	Lf1huswifwrk	R1 LF13 HUSB/WIFE/PARTN PAY WORK
44	ia1totinc	R1 IA50 TOTAL INCOME
45	ia1toincesjt	R1 IA51A JOINT EST TOT INCOME
46	ia1toincessg	R1 IA51B SNGLE EST TOT INC