HRS 1996 Imputations: Documentation

Honggao Cao

Institute for Social Research
University of Michigan

February 2001

This document describes imputation results for HRS Wave 3 (1996) data. For questions related to the results or the documentation, contact the author or hrsquest@isr.umich.edu.

CONTENTS

HRS96 Imputations: Documentation

I. Introduction

The Health and Retirement Study (HRS) is a national longitudinal survey that describes for older Americans the dynamics of economic, health, marital, and family statuses, as well as public and private support systems. A companion study to the Asset and Health Dynamics among the Oldest Old (AHEAD)[1], and administered at the Institute for Social Research (ISR), the University of Michigan, HRS uses an important method to collect data that are usually subject to high rates of non-response (Heeringa, Hill, and Howell, 1995; Hill, 1999; Juster and Smith, 1998). A key feature of the method is to allow a respondent who is unwilling or unable to answer an "amount" question to provide "bracket" information about the "amount." To utilize the "bracket" information collected, one needs to unfold brackets, or impute the amount variables with brackets.

In this document we describe imputation results released in five supplemental, imputation files for HRS Wave 3 (1996) public release data: "h1996i_e," "h1996i_f," and "h1996i_j," "h1996i_n," and "h1996i_r." For simplicity, we shall call these files "HRS96 imputations." An imputation program developed at the ISR, IMPUTE,[2] was used to produce the results, which include variables for imputed values, imputation controls, bracket patterns, and imputation types. Naturally, our focus is on the imputations of *the amount variables with brackets;*[3] examples of

[1] Since 1998, HRS and AHEAD have been integrated into a single study, still named HRS.

[2] A SAS application, the program IMPUTE, along with a documentation file, is available upon request. Send request to: hrsquest@isr.umich.edu

[3] "Amount variable" is a term used in the imputation program, IMPUTE, to indicate a variable for which imputations are done. For the definition of "amount variable," as well as other important terminologies used in the program (and hence in this file), see Section IV or consult Cao (2001).

these variables are given in Section III. But some of the variables without brackets are also covered if they are believed to be integral components of three aggregate variables interested by many researchers: household income, housing equity, and net worth. The rules for selecting the amount variables for imputations are stated formally in Section II, while the strategies for producing and organizing imputation results are discussed in Section IV.

In Section V, we list all of the amount variables included in the imputation files, with comments on the specific treatment to the imputations for each of the variables. The core of the document, this section may be viewed as a supplement to--but not a substitute for--the codebook of "HRS96 imputations," which is released simultaneously with this document.

Finally in Section VI, we list and briefly answer several important questions that we believe users may have regarding the imputations.

Users who are not interested in our imputation strategies may skip the rest of the paper. It is recommended, however, that you consult Section IV, "Producing and Organizing Imputation Results," and Section V, "Questions and Answers," whenever encountering an imputation-related problem.

II. Rules for Variable Selections

Not all of the variables in HRS Wave 3 data have imputations included in the imputation files. For some, imputations may easily be done by researchers themselves, tailoring to their own needs. Most of those variables do not have brackets. For other variables, there are too few valid records to make any imputations meaningful—even though they do have brackets. To be specific, the following rules were used to determine if an HRS96 variable should be imputed and included in the imputation files ---

Box 1. The Rules for Variable Selections

1. The Bracket Rule: In general, all the variables with brackets should be imputed, and included in the imputation file;

2. The Income/Asset Component Rule: A variable without bracket should be imputed if it is determined to be an integral component of household income, housing equity, or net wealth;

3. The Imputability Rule: A variable eligible for imputation based on Rule 1 or 2 will not be imputed if there are no enough valid observations.

III. The Bracketed and Unbracketed Problems: Examples

An imputation problem encountered in HRS96 may be categorized as one of the two types: bracketed and unbracketed. An unbracketed problem, consisting only of "ownership" and "amount" questions, has a familiar structure (Box 2). The ownership question asks a respondent if he or she "owns" the subject of interest (receives welfare income, pays medical expenditure,

Box 2. The Unbracketed Problem: an Example

(Ownership Question)
O1. Did you (or your husband/or your wife/or your partner) receive any income from welfare in (1995/1996)?

	1.	Yes
	5.	No
	7.	Other
	8.	DK (don't know); NA (not ascertain)
	9.	RF (refused)
	Blank	INAP (Inapplicable)

(If "Yes" to the Ownership Question, then)
(Amount Question)
A1. How much did (you/he/she) receive in (1995/1996)?

	1-999996.	Actual value
	999997.	Other
	999998.	DK (don't know); NA (not ascertained)
	999999.	RF (refused)
	Blank.	INAP (Inapplicable)

Etc.) If the answer is "yes," then the amount question asks for the amount that he or she "owns." When a respondent gives a positive answer to the ownership question but an uncertain answer (e.g., "don't know" or "refused") to the amount question, however, an unbracketed problem would not pursue any further the uncertain answer.

By contrast, a bracketed problem not only has "ownership" and "amount" questions, it also has "bracket" questions regarding the "amount" when a respondent giving a positive answer to the "ownership" question fails to provide an exact amount (Boxes 3a and 3b). Varying in the

Box 3a. The Bracketed Problem with a Single Bracket Question: an Example

```
(Ownership Question)
O1. People sometimes receive property or lump sum payments of money
from such things as pension settlements, insurance settlements,
cash in annuities, or inheritances. ... In the past two years, did
you/or your husband/or your wife/or your partner receive a lump sum
of money or property that you have not already told me about?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertained)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)

(If "Yes" to the Ownership Question, then)
(Amount Question)
A1. About how much did you/or your husband/or your wife/or your
partner receive from that?
      0-999996.   Actual value
        999997.   Other
        999998.   DK (don't know); NA (not ascertained)
        999999.   RF (refused)
         Blank.   INAP (Inapplicable)

(If "999998" or "999999" to the Amount Question, then)
(Bracket Question)
B1. Was it more than $50,000?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertained)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)
```

structure, a bracket question generally asks a respondent if the amount of the "thing" in interest that he or she "owns" is more (or less) than a certain break point. Since this question does not ask for an exact amount, it is likely that a respondent unable or unwilling to provide a definite answer to the amount question becomes able or willing to give a definite answer to the bracket question.

A bracketed problem may have one (Box 3a) or more (Box 3b) "bracket" questions; each of them corresponds to a distinct break point. (In the example Box 3a, the only break point is $50,000; while in the example Box 3b, the three break points are $3,000, $5,000, and $10,000.) Since an unbracketed problem is equivalent to a bracketed problem with no bracket question, it

Box 3b. The Bracketed Problem with Multiple Bracket Questions: an Example

```
(Ownership Question)
O1. Are you/or your husband/or your wife/or your partner currently
receive any income from veteran benefits?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertained)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)

(If "Yes" to the Ownership Question, then)
(Ownership Question Continued)
O2. Who received that?
      1.     R only
      2.     Spouse or Partner
      3.     Both
      7.     Other
      8.     DK (don't know); NA (not ascertained)
      9.     RF (refused)
      Blank. INAP (Inapplicable)

(If "1" or "3" to the above question, then)
A1. (Amount Question for R's Veteran Benefits)
How much did you (yourself) receive last month from that?
      0-999996.   Actual value
        999997.   Other
        999998.   DK (don't know); NA (not ascertained)
        999999.   RF (refused)
         Blank.   INAP (Inapplicable)
```

Box 3b. The Bracketed Problem with Multiple Bracket Questions (Continued)

```
(If "999998" or "999999" to the Amount Question, then)
(Bracket Question - 1)
B1. Did you receive more than $3,000?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertained)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)

If "1" to the Bracket Question, then)
(Bracket Question – 2)
B2. Did you receive more than $5,000?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertained)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)

(If "1" to the Bracket Question – 2, then)
(Bracket Question – 3)
B3. Did you receive more than $10,000?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertained)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)
```

may be viewed as a special type of bracketed problem.

IV. Producing and Organizing Imputation Results

a). Translating an Imputation Problem for IMPUTE

In order to use the imputation program, IMPUTE, one first needs to translate an imputation problem into a set of variables required by the program. Three types of variables are essential for IMPUTE, including the amount variable, the control variable, and, for a bracketed problem, the bracket variable(s). The amount variable is one for which imputations are done. In general, this variable corresponds to the amount question of an imputation problem, taking on the value that answers the amount question. Similarly, the control variable corresponds to the

ownership question, dictating if a respondent has a positive value on the amount variable; and a bracket variable corresponds to a bracket question, providing with bracket information about the amount when the exact value on the amount variable is not available. The relationship between an imputation problem and the variables translated from the problem is summarized in Box 4.

Box 4. The Relationship Between an Imputation Problem and the Variables Translated from the Problem

Question in an Imputation Problem		*Translated Variable*
Ownership Question	-------- ->	Control Variable
Amount Question	-------- ->	Amount Variable
Bracket Question(s)	-------- ->	Bracket Variable(s)

The translation of an imputation problem into the variables required appears to be an easy task. But, in reality, that is not always true, especially for creating control variables. To impute "veteran benefits a *respondent* received last month" for the problem in Box 3b, for example, one needs to integrate the two ownership questions, O1 and O2, to create the control variable. That is, whether a respondent received veteran benefits last month may only be determined by combining the information collected from both questions O1 and O2. In fact, as indicated in Section V, about half of the control variables used for HRS96 imputations were created from two or more variables (or ownership questions).

b). The Imputation Strategies

The strategy appropriate for imputing an amount variable depends on the distributional characteristics of the variable as well as the structure of the underlying imputation problem. In producing the HRS96 imputations, we adopted various strategies that we'd believed adequately

reflect the characteristics of various amount variables and their underlying problems (see the last column in Box 10). These strategies, however, have three common features (Box 5).

First, all the imputations were based on a "mixed" method. As detailed in Appendix, Cao (2001), a "mixed" method replaces a missing value with a valid value based on a "hotdeck + regression" technique. Specifically, when it has a closed (e.g., 5,000 – 10,000) or bottom-open (e.g., -- 5,000) bracket, a missing value would be replaced by a value *randomly* selected from an appropriately-formed donor pool, a technique often referred to as "hotdeck"; when it has a top-open bracket (e.g., 5,000 -), it would be replaced by a value from the donor pool based on the predicted score from a *regression* model.

Second, whenever a regression technique was needed, the covariate vector always contained four basic demographic variables – including age, sex, educational attainment, and marital status. When imputing for the income from an asset (e.g., business or farm, stock, and bonds), however, the vector also included the asset variable.

And finally, when imputing a respondent variable and a spouse variable that is otherwise identical (e.g., respondent's veteran benefits and spouse's veteran benefits), a joint donor pool was always formed. The underlying rationale for pooling a respondent variable with a spouse variable is of two-fold: (i) the question regarding the respondent's veteran benefits may be

Box 5. The Common Features of the Imputation Strategies

1. A "mixed" method for all the imputations

2. A covariate vector containing four basic demographic variables (age, sex, educational attainment, and marital status), and, when imputing the income from an asset, the asset variable

3. A joint donor pool for imputing the same variables for a respondent and his/her spouse

treated as the same as the question regarding the spouse's veteran benefits, so a joint donor pool is conceptually reasonable; and (ii) since pooling together the valid values of both respondent and spouse variables always increases the size of the donor pool, the reliability of the imputations for both the respondent variable and the spouse variable tends to be increased.

c). The Conventions Used for Organizing Imputation Results

Anticipating that the HRS96 imputations are most likely to be used together with other HRS96 variables in the public release data, we organized the results according to a set of useful conventions. First, an easy-to-understand connection between an imputation-related variable and the original amount variable is built. An imputation-related variable is always named as an original amount variable plus a single letter extension, a rule that becomes clearer and clearer as we proceed.

Second, for each amount variable imputed, six imputation-related variables are reported. Combined, these variables provide not only information on the structure of the imputation problem, but also information on the way the imputations have been done. A brief description of the variables is given in Box 6.

An imputation-related variable with extension "x"-- or the "x" variable for short -- stands for imputed values. When an original amount variable has a valid value, the "x" variable simply maintains the value; otherwise, it takes on a value out of imputation. By default, a valid value is one that an amount variable may possibly take on. Neither "missing," "don't know," nor "refused" is a valid value, although some missing values may be legitimate. (For example, a missing value on the spouse's social security income for a respondent who is single should be legitimate. As a rule, for a legitimate missing on an amount variable, the "x" variable will be set to zero (0), while the control variable to missing (.).

Box 6. Imputation-Related Variables

Variable Name	*Description*
Q1234	Original amount variable in the public release data
Q1234x	Variable for imputed value (imputed amount variable)
Q1234t	Variable for imputation type or quality
Q1234s	Variable summarizing bracket information
Q1234c	Control variable for the amount variable
Q1234d	Variable for the lower bound value of the bracket
Q1234e	Variable for the upper bound value of the bracket

The "s" variable synthesizes the information on the bracket associated with the amount variable. More often than not, a bracketed problem has two or more questions. To make it easier to use the information collected from all the bracket questions, the "s" variable combines the answers from the questions into a single index. The rule for reading such an index is shown in Box 7.

Box 7. The Rules for Reading an Index of the "s" Variable

1.The General Rule: Let an imputation problem with an amount variable Q1234 have i bracket variables, var1, var2, ..., vari. Each of the variables corresponds to a distinct break point value, with var1 to the smallest value, var2 to the second smallest, etc. Let a missing value on any of the bracket variables be set to zero.

The variable Q1234s is then defined as ---

$$Q1234s = var1 + var2*10 + var3*100 + \ldots + vari*10^{(i-1)}.$$

2. The Special Rule: The following special codes are reserved for an "s" variable ---

'-2': the missing value on the amount variable is legitimate;
'-1': the value on the amount variable is valid;
'0': imputation for the amount variable was based on no bracket information.

The "c" variable is the control variable about which we have already discussed. For "HRS96 imputations," a value "1" on the "c" variable stands for a positive answer (or "yes") to the ownership question, a value "5" for a negative answer (or "no"), while "8" and "9" for "don't know" and "refused," respectively.

The "d" and "e" variables give the lower and upper bound values, respectively, of the bracket. For a closed bracket (e.g., 5000 -10000), neither the "d" or "e" variables should be missing. For an open-end bracket (e.g., 5000 - , - 10000), the variable corresponding to the open end (lower or upper) should be missing. When the amount variable has no bracket information (thus open at both ends), therefore, both the "d" and "e" variables should be missing. *In addition, a special code, "-1," has been reserved for the two variables if the amount variable is about the spouse of an information-providing respondent and the respondent has no spouse.*

Combining the information from "x," "s," "c," "d," and "e" variables, the "t" variable is an index for identifying the type of imputation done to each record of the amount variable. Regardless of the structure of an imputation problem, an imputed value belongs to one of ten categories, which are listed in the last column in Box 8. While a full appreciation of the availability of this imputation quality variable requires a thorough knowledge of the imputation program IMPUTE, users without any knowledge of the program may still find the variable useful if they want to subset, modify our imputation results, or do the imputation work of their own.[4]

d). Post-Imputation Treatment: the "z" Variable

Depending on the structure of an imputation problem, the amount variable for which imputations are actually done may or may not be what we want to know conceptually. Consider

[4] For example, if one does not trust in our work for open brackets, one may keep only those records with the values of the "t" variable less than 5.

Box 8. Imputation Type as Determined by the "x," "c," "d," "e," and "s" Variables [5]

Original Amount Variable (Q1234)	Imputation Variable (Q1234x)	Control Variable (Q1234c)	Summary Bracket Variable (Q1234s)	Break point Value Bottom (Q1234d)	Top (Q1234e)	Imputation Type (Q1234t)
legitimate missing (1)	0	.	-2	.	.	1
legitimate missing(2)	0	.	-2	-1	-1	1
valid value	>=0	(1,5)	-1	>=0	>=0	2
closed bracket	>0	1	>0	>=0	>0	3
bottom-open bracket	>=0	1	>0	.	>0	4
top-open bracket	>0	1	>0	>=0	.	5
bracket with no information	>=0	1	>=0	.	.	6
imputed ownership and Amount	>=0	8,9	0	.	.	7

the example in Box 3b. The amount variable for this problem is naturally "the veteran benefits a respondent received last month." But what about if we want to know another variable, "the veteran benefits a respondent received in the last calendar year"? Can we construct this new variable based on the imputations for the original amount variable?

The work involving the conversion of the imputations for an original amount variable into another related variable is called as "post-imputation treatment," while the new, converted variable is named with the original amount variable plus an extension "z" (e.g., Q1234z). In most

[5] Note that the rule for creating a "t" variable has been changed since July 2000, when we released the AHEAD95 imputations.

of the cases in the "HRS96 imputations," the "x" to "z" conversions was straightforward, involving only simple mathematical computations. For other cases, however, the tasks were quite complicated, as illustrated for the example "the veteran benefits a respondent received in the last calendar year" in Box 9.

Box 9. The "x" to "z" Conversion: an Example

```
Original Amount Variable: E4043 (Veteran Benefits a Respondent
                                 Received Last month (year 1996))
Imputation Variable: E4043x (Veteran Benefits a Respondent
                             Received Last Month (year 1996))
Objective: Construct E4043z (Veteran Benefits a Respondent
                             Received in Year 1995)

The "Starting-time-dependent" Conversion Strategy ---

1. Assumption: monthly veteran benefits a respondent received in
   year 1995 was about the same as he/she received last month

2. Determine when a respondent started to receive the benefits:
   E4051 (starting year), E4052 (starting month)

3. Impute E4051x and E4052x as some of the values of E4051 and E4052
   are missing

4. E4043x to E4043z Conversion:
   If E4051x <= 1994 then E4043z = E4043x*12
   Else if E4051x > 1995 then E4043z = 0
   Else if E4051x = 1995 then E4043z = E4043x*(13-E4052x)
```

In order to know from "the veteran benefits a respondent received last month" (E4043x) "the veteran benefits a respondent received in year 1995 (or last calendar year)" (E4043z), we adopted a so-called "starting-time dependent" x-to-z conversion strategy.[6] We first built the relationship between the monthly benefits in 1996 and in 1995. Without great loss of accuracy, we assumed that "monthly veteran benefits a respondent received--if any--in 1995 were about

[6] As indicated in Box 11, this strategy was used for x-to-z conversion for many other variables as well.

the same as he/she received last month." We then determined if a respondent was receiving any veteran benefits in 1995. To do this, we imputed two "starting-time" variables, E4051 (year starting to receive the benefits) and E4052 (month starting to receive the benefits), as some of the values of the variables were missing. Finally, we made the conversion based on the assumed relationship between the monthly benefits in 1996 and in 1995, and the imputations for the "starting-time" variables.

V. The Structure of "HRS96 Imputations"

"HRS96 imputations" are divided into five different data sets: "h1996i_e," "h1996i_f," "h1996i_j," "h1996i_n," and "h1996i_r." The names of the data files contain both HRS wave (h1996i) and section (e.g., e, f, j, n, and r) identifiers. That is, "h1996i_e" is a file containing imputations for HRS Wave 3 data, section E (health care utilization), "h1996i_f" for section F (housing), "h1996i_j" for section J (income), "h1996i_n" for section N (widowhood and asset changes), and "h1996i_r" for section R (health insurance). Because of the structure of the original data, "h1996i_f," "h1996i_j," and "h1996i_n" are household level data sets, with one record for each HRS Wave 3 household; "h1996i_e" and "h1996i_r" are respondent level data sets, with one record for each HRS Wave 3 respondent.

In this section, we list all the amount variables covered in the five imputation files. For each variable, we list its name, a label defining the variable[7], the variables involved in the creation of the control variable, and, when needed, a brief comment on the imputation strategy used. Users curious about our imputation strategy for any amount variable are encouraged to look up the codebook for the HRS96 public release data as well as the documentation file of the

[7] This is the same label used in the HRS96 public release data, combining the question number with a brief definition of the variable.

program IMPUTE for a better understanding.

We first list the amount variables for which imputations were actually done (Box 10), then the variables for which the "x" to "z" conversions were made (Box 11). For convenience, we also provide the definitions of four aggregate variables included in the imputation files (Box 12). These aggregates are: household income, net wealth (in the file "h1996i_j"), main home equity, and second home equity (in the file "h1996i_f"). While a conscious effort was made to have them constructed as well as possible, users are not required to agree with us on the definitions.

Box 10. The Amount Variables Imputed

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
	SECTION E: HEALTH CARE UTILIZATION	
E10. NURHM OR HOSP R PAY $ E1783	E1770 E1775 E1776 E1781	
E18A.DOCTOR/OUT/ DENTAL R PAY $ E1804	E1790 E1795 E1793 E1798 E1800 E1803	
E21A.PRESCR R PAY $ E1816	E1811 E1815	
E24A.SPECIAL R PAY $ E1834	E1827 E1829 E1831	
E26.TOTAL COST MEDICAL E1841	E1841	
E172.VOL HELP E2169	E2169	
E173.HELP FRIENDS E2172	E2172	
	SECTION F: HOUSING	
F6.HOME VALUE $ E2246	E2226 E2229 E2234	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
F7A.$ PAYMENT ON MORTGAGE E2252	E2251M1 E2251M2 E2251M3 E2253	annualized before imputation
F7D.$ OWN ON MORTGAGE E2256	E2251M1 E2251M2 E2251M3	
F9A.$ PAYMENT 2ND MORTGAGE E2260	E2251M1 E2251M2 E2251M3 E2261	annualized before imputation
F9C.$ OWN ON 2ND MORTGAGE E2263	E2251M1 E2251M2 E2251M3	
F10.$ PAYMENT OTHER LOAN E2267	E2251M1 E2251M2 E2251M3 E2268	annualized before imputation
F10E.$OWE ON LOAN E2270	E2251M1 E2251M2 E2251M3	
F11A.$ COULD BORROW E2275	E2274	
F11C.$ OWED EQUITY LOAN E2277	E2276	
F13.REAL ESTATE TAX E2292	E2226 E2229 E2234	
F14.$ AMOUNT RENT E2301	E2226 E2229 E2234 E2302	annualized before imputation
F20A.$ PARK ASSOC FEES E2317	E2316 E2318	annualized before imputation
E2326X F21D.HOW MUST LUMP_SUM E2326	E2325	
F21D. $ ASSOC FEE E2333	E2332 E2334	annualized before imputation

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
F46E.PURCHASE PRICE 2ND HOME E2434	E2407 E2408 E2409 E2412 E2417	
F46F.$ VALUE 2ND HOME E2435	E2407 E2408 E2409 E2412 E2417	
F47A.$ PAYMENT ON MORTGAGE-2 E2441	E2440M1 E2442	annualized before imputation
F47D.$ OWN ON MORTGAGE-2 E2445	E2440M1	
F50B. $ OWE ON LOAN-2 E2459	E2440M1	
F53.REAL ESTATE TAX-2 E2480	E2407 E2408 E2409 E2412 E2417	
	SECTION J: ASSETS AND INCOME	
J6. R SELF-EMPLOYMENT INCOME AMOUNT E3876	E3875	joint donor pool with E3909
J8.R HOW MUCH WAGES AND SALARY E3883	E3882	joint donor pool with E3916
J10.R HOW MUCH PROF PRACTICE OR TRADE E3890	E3889	joint donor pool with E3923
J11A.R HOW MUCH TIPS, BONUSES, COMMISSIO E3897	E3896	joint donor pool with E3930
J12A.HOW MUCH OTHER INCOME LCY E3903	E3902	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J14.SP HOW MUCH SELF-EMPLOYMENT E3909	E3908	joint donor pool with E3876
J16.SP HOW MUCH WAGES AND SALARY E3916	E3915	joint donor pool with E3883
J18.SP HOW MUCH PROF PRACTICE OR TRADE E3923	E3922	joint donor pool with E3890
J19A.SP HOW MUCH TIPS, BONUSES, COMMISSI E3930	E3929	joint donor pool with E3897
J20A.SP HOW MUCH OTHER INCOME E3936	E3935	
J23.AMT UNEMPL COMP LCY E3944	E3942 E3943	joint donor pool with E3952
J28. AMT S/P UNEMPL COMP LCY E3952	E3842 E3943	joint donor pool with E3944
J35.AMT WORKERS COMP LCY E3963	E3961 E3962	joint donor pool with E3974
J40.AMT S/P WORKERS COMP LCY E3974	E3961 E3962	joint donor pool with E3963
J47.R AMT RECVD SS LAST MO E3985	E3983 E3984	joint donor pool with E3996
J50A1.SP HOW MUCH SS BENEFITS E3996	E3983 E3984	joint donor pool with E3985
J53. R SOC SEC EXPECTED AMOUNT E4010	E4008 E4011	annualized before imputation joint donor pool with E4016
J57. SP SOC SEC EXPECTED AMOUNT E4016	E4014 E4017	annualized before imputation joint donor pool with E4010
J61.R/S SSI $ E4023	E4021	
J66.AMT WELFARE LCY E4038	E4036	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J69.R VET BENEFIT LAST MONTH E4043	E4041 E4042	joint donor pool with E4056
J72.AMT SP VET BEN LAST MO E4056	E4041 E4042	joint donor pool with E4043
J77.REAL ESTATE $ E4071	E4070	
J82.RENTAL RECEIVE $ LCYEAR E4089	E4085 E4086 E4087 E4088	integration with annualized "per-period income" (E4087) to increase the size of the donor pool
J84.BUSINESS OR FARM $ E4097	E4096 E4100 E4101	
J89.BUS/FARM INCOME $ LCYEAR E4117	E4112 E4113 E4114 E4115	integration with annualized "per-period income" (E4114) to increase the size of the donor pool
J93.TOTAL $ IRA ACCOUNT E4127	E4124	
J95.IRA NO REG INC WITHDRAW $-1 E4138	E4137 E4139	IRA-1 total withdrawal in last 2 years annualized before imputation
J100.TOTAL $ IRA ACCOUNT-2 E4149	E4125	
J102.IRA NO REG INC WITHDRAW $-2 E4159	E4158 E4160	IRA-2 total withdrawal in last 2 year annualized before imputation
J107.TOTAL $ IRA ACCOUNT-3 E4171	E4125	
J109.IRA NO REG INC WITHDRAW $-3 E4178	E4177 E4179	IRA-3 total withdrawal in last 2 years annualized before imputation
J113.AMT CONVERT E4188X	E4187M1 E4187M2	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J114.ANNUITY AMOUNT E4189	E4187M1 E4187M2 E4190	annuity income in last 2 years annualized before imputation
J124.AMT WITHDRAW E4204	E4187M1 E4187M2	
J132.AMT PENSION LAST MONTH E4214_1	E4209 E4210	joint donor pool with E4237_1
J132.(J30D)AMT PENSION LAST MONTH E4214_2	E4209 E4210 E4212	joint donor pool with E4237_2
J136.AMT IN PENSION E4222_1	E4221_1	joint donor pool with E4245_1
J136.AMT IN PENSION E4222_2	E4221_2	joint donor pool with E4245_2
J143.OTHER PENSIONS E4232	E4209 E4211 E4212	joint donor pool with E4258
J148.AMT PENSION LAST MONTH E4237_1	E4209 E4210 E4234	joint donor pool with E4214_1
J148.(J31D)AMT PENSION LAST MONTH E4237_2	E4235 E4209 E4210 E4234	joint donor pool with E4214_2
J152.AMT IN PENSION E4245_1	E4244_1	joint donor pool with E4222_1
J152.(J31H)AMT IN PENSION E4245_2	E4244_2	joint donor pool with E4222_2
J159.OTHER PENSIONS SP E4258	E4234 E4235	joint donor pool with E4232
J168.AMT ANNUITY INC E4269_1	E4263 E4264	joint donor pool with E4298_1
J168.(J33D)AMT ANNUITY INC E4269_2	E4263 E4264 E4265 E4266	joint donor pool with E4298_2

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J176.OTHER ANNUITIES E4291	E4266	joint donor pool with E4320
J181.AMT ANNUITY INC SP E4298_1	E4263 E4264	joint donor pool with E4269_1
J181.(J35C)AMT ANNUITY INC SP E4298_2	E4263 E4264 E4295 E4296	joint donor pool with E4269_2
J188A.OTHER ANNUITY INC SP E4320	E4295 E4296	joint donor pool with E4291
J208.TOTAL $ STOCKS E4340	E4339	
J213.STOCK INCOME $ LCYEAR E4362	E4359 E4360 E4361	integration with annualized "per-period income" (E4360) to increase the size of the donor pool
J216.TOTAL $ BONDS E4383	E4382	
J221.BOND INCOME $ LCYEAR E4405	E4401 E4402 E4403	integration with annualized "per-period income" (4403) to increase the size of the donor pool
J223.TOTAL $ CHECKING, SAVINGS E4426	E4425	
J228.CHECKING INCOME $ LCYEAR E4445	E4441 E4442 E4443 E4444	integration with annualized "per-period income" (E4443) to increase the size of the donor pool
J230.TOTAL $ CD E4465	E4464	
J236.CD INCOME $ E4487	E4483 E4486 E4484 E4485	integration with annualized "per-period income" (E4485) to increase the size of the donor pool
J238.TRANSPORTATION E4501	E4500	
J240.OTHER ASSETS $ E4514	E4513	
J242.OTHER ASSETS INCOME E4519	E4518	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J245.TRUST FUNDS $LCY E4564	E4563	
J248.AMT ALIMONY $ LCY E4568	E4567	
J250.OTHER SOURCES $ 1994 E4575	E4574	
J274.AMT HU MEMBERS INCOME E4625	E4625	
J277.AMT ASSETS OTH HHMEM E4633	E4632	
J284.TOTAL FAMILY INCOME E4643	E4643	
J290.USE ASSETS $ E4666	E4652 E4667	use assets in 2 years annualized before imputation
J293.ASSET ADD $ E4689	E4687	
J299.$ TRUSTS E4713	E4709	
J303.DEBTS $ E4727	E4726	
J309.LUMP $-1 E4752	E4748	
J315.LUMP $-2 E4759	E4755	
J321.LUMP $-3 E4766	E4762	
SECTION N: WIDOWHOOD AND ASSET CHANGES		
N2. SELLING PRICE E4833	E4829 E4830	
N4. PURCHASE PRICE E4839	E4839	
N6. SELLING PRICE E4846	E4843 E4846	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
N8. PURCHASE PRICE E4852	E4852	
N10. IMPROVEMENT COST E4857	E4856	
N12. PURCHASE PRICE E4863	E4862	
N14. REAL ESTATE SOLD E4868	E4867	
N16. BUY MORE/LESS THAN SOLD E4873	E4872	
N18. IMPROVEMENT COST E4878	E4877	
N20. IMPROVEMENT COST E4883	E4882	
N22. SELLING PRICE E4888	E4887	
N25. HOW MUCH E4897	E4895 E4896	
N26. INCLUDING AUTOMATIC REINVESTMENT E4901	E4895	
N28. HOW MUCH E4906	E4905	
	SECTION R: Health Insurance	
R96.POLICY FACE VALUE -ONLY ONE POLICY E5286	E5284 E5285	
R97.POLICIES FACE VALUE -2+ POLICIES E5287	E5284 E5285	
R100A. LIFE INS POLICY FACE VALUE E5295	E5294	
R102.LAPSED POLICY FACE VALUE	E5301	

Box 11. The "z" Variables

Label Variable Name Strategy	Variables Used for Creating the Control Variable	Comments on the Imputation
J47.R AMT RECVD SS LAST MO E3985Z	E3990 E3991	starting-time dependent
J50A1.SP HOW MUCH SS BENEFITS E3996Z	E4001 E4002	starting-time dependent
J61.R/S SSI $ E4023Z	E4027 E4028	starting-time dependent
J69.R VET BENEFIT LAST MONTH E4043Z	E4051 E4052	starting-time dependent
J72.AMT SP VET BEN LAST MO E4056Z	E4064 E4065	starting-time dependent
J89.BUS/FARM INCOME $ LCYEAR E4117Z	E4122	E4117Z=E4117X if E4122=5, 0 otherwise
J95.IRA NO REG INC WITHDRAW $-1 E4138Z		E4138Z=E4138X/2
J102.IRA NO REG INC WITHDRAW $-2 E4159Z		E4159Z=E4159X/2
J109.IRA NO REG INC WITHDRAW $-3 E4178Z		E4178Z=E4178X/2
J114.ANNUITY AMOUNT E4189Z		E4189Z=E4189Z/2
J124.AMT WITHDRAW E4204Z		E4204Z=E4204X/2
J132.AMT PENSION LAST MONTH E4214_1Z	E4223_1 E4224_1	starting-time dependent
J132.(J30D)AMT PENSION LAST MONTH E4214_2Z	E4223_2 E4223_2	starting-time dependent

Box 11. The "z" Variables (Continued)

Label Variable Name Strategy	Variables Used for Creating the Control Variable	Comments on the Imputation
J148.AMT PENSION LAST MONTH E4237_1Z	E4246_1 E4247_1	starting-time dependent
J148.(J31D)AMT PENSION LAST MONTH E4237_2Z	E4246_2 E4247_2	starting-time dependent
J168.AMT ANNUITY INC E4269_1Z	E4275_1 E4276_1	starting-time dependent
J168.(J33D)AMT ANNUITY INC E4269_2Z	E4275_2 E4276_2	starting-time dependent
J181.AMT ANNUITY INC SP E4298_1Z	E4304_1 E4305_1	starting-time dependent
J181.(J35C)AMT ANNUITY INC SP E4298_2Z	E4304_2 E4305_2	starting-time dependent
J290.USE ASSETS $ E4666Z		E4666Z=E4666X/2
J309.LUMP $-1 E4752Z	E4750 E4751	starting-time dependent
J315.LUMP $-2 E4759Z	E4757 E4758	starting-time dependent
J321.LUMP $-3 E4766Z	E4764 E4765	starting-time dependent

VI. Questions and Answers

Q1. What variables were imputed for the HRS 96 data?

A1. All the variables imputed for the HRS96 data are listed in Box 10. These variables were selected based on the rules detailed in Box 1.

Box 12. Household Income, Housing Equity, and Net Wealth

1. Household income in the last calendar year:

ehhinc = E3876x+E3909x+E3883x+E3916x+E3890x+E3923x+
E3897x+E3930x+E3903x+E3936x+E3944x+E4519x+
E3952x+E3963x+E3974x+E4038x+E4564x+E4568x+
E4575x+E3985z+E3996z+E4023z+E4043z+E4056z+
E4089x+E4117z+E4138z+E4159z+E4178z+E4214_1z+
E4214_2z+E4237_1z+E4237_2z+E4269_1z+E4269_2z+
E4298_1z+E4298_2z+E4362x+E4405x+E4445x+E4487x+E4189z

2. Net wealth at the time of interview:

eassets = E4071x+E4097x+E4127x+E4149x+E4171x+E4340x+
E4383x+E4426x+E4465x+E4501x+E4514x+E4713x-E4727x

3. Main home equity:

ehome1 = E2246x-E2256x-E2263x-E2270x-E2277x

4. Second home equity:

ehome2 = E2435x-E2445x-E2459x-E2466x

Q2. How did you impute a variable?

A2. For all of the variables selected for imputation, we used a "mixed" imputation method, which combines a "hotdeck" imputation procedure and a "linear regression" procedure. Please see Cao (2001), Appendix, for detail.

Q3. Can I impute a variable myself?

A3. Yes, you certainly can. If the variable you want to impute was not selected for imputation by us--most likely, the variable would have no bracket--you may impute it based on any imputation procedure that you believe reasonable ("hotdeck," for example). If you want to do imputations yourself systematically, using a different method, you may do so by taking advantage of the availability of the original bracket variables in the data sets. And you may use a summary bracket variable (the "s" variable) to simplify your work.

Q4. What is the main difference between an “x” variable and a “z” variable?

A4. An “x” variable is an original imputation variable, while a “z” variable is one converted from a corresponding “x” variable. The conversion is needed usually because the “x” variable is not what we want conceptually. A “z” variable must have a corresponding “x” variable, but not vice versa.

Q5. Can I define “hhinc” myself?

A5. Yes, you can, and sometimes you should. For example, if you don’t want to have household income contain anything from assets, you should drop the asset income components from the equation in Box 12.

Q6. Why are there sixty (60) households with missing values on all the variables in the household level files “h1996i_f,” “h1996i_j,” and “h1996i_n”?

A6. There are no financial respondents for those households. However, since the households were included in the public release data, we include them in our imputation files to make the files consistent with the public release data.

Q7. Why are there multiple imputation-related variables for each amount variable?

A7. For each amount variable, there are six imputation-related variables: the “x,” “c,” “s,” “d,” “e,” and “t” variables. (For definitions of the variables, see Box 6.) We include these variables in the data sets to accommodate various needs of perspective users. Those who just want to have imputed values may keep only “x” and, if applicable, “z” variables when working with the data.

Q8. How do you create a “c” variable?

A8. The creation of a "c" variable may be complicated or very simple, depending on the structure of the imputation problem. Sometimes we need only a single variable to create the control variable for the problem. In other cases, we need to incorporate information from several variables to determine if a respondent or household receives welfare income, pays medical expenditures, etc.

Q9. What should I do if I find any problems with the imputations?

A9. Problems with the imputations may exist for various reasons. If you believe that the imputations for a variable do not look correct, please contact hrsquest@isr.umich.edu.

References

Cao, H. "IMPUTE: A SAS Application System from Missing Value Imputations --- With Special Reference to HRS Income/Assets Imputations, Version 2.0," Institute for Social Research, University of Michigan, February 2001.

Herringa, S. G., Hill, D. H., and Howell, D. A. "Unfolding Brackets for Reducing Item Non-Response in Economic Surveys," Health and Retirement Study Working Paper, No. 94-029, June 1995.

Hill, D. H. "Unfolding Bracket Method in the Measurement of Expenditures and Wealth," in J. P. Smith, and R. J. Willis (eds.), Wealth, Work, and Health, University of Michigan, 1999.

Juster, F. T., and Smith, J. P. "Improving the Quality of Economic Data: Lessons from the HRS and AHEAD," *Journal of the American Statistical Association*, Vol. 92, No. 440, 1998.