# Health and Retirement Study

# HRS 2001 HUMS
# College Tuition Imputations
## Restricted Version

# Data Description and Usage

# Final Release, Version 1.0
# June 2005

# 1. Overview

The Health and Retirement Study (HRS) is a national longitudinal study of the economic, health, marital, and family status, as well as public and private support systems, of older Americans. The HRS is a rich source of longitudinal, cross-sectional data for researchers and policymakers who study aging. Funding for the Health and Retirement Study is provided by the National Institute on Aging at NIH (U01 AGO9740), with supplemental support from the Social Security Administration. The study is conducted by the Institute for Social Research (ISR) at the University of Michigan.

The original HRS sample consists of individuals born between 1931 and 1941, inclusive. This sample came from a screening of 69,336 households that was conducted in 1992.That sample of households was generated using a multi-stage, clustered area probability frame. The second sample was generated for what began as a separate study: Asset and Health Dynamics among the Oldest Old (AHEAD). This sample consists of individuals born in 1923 or before. Those born between 1914 and 1923, and about half of those born in 1913 or before, were identified through the same household screening used to identify the original HRS sample. The other half, those born in 1913 or before, was identified using the Medicare enrollment files maintained by the Health Care Financing Administration (HCFA, since renamed the Centers for Medicare, Medicaid Services, or CMS). The HRS and AHEAD studies were merged, with a single interview schedule, in 1998, and at the same time the third and fourth samples were added. The War Baby (WB) sample consists of those born between 1942 and 1947, inclusive, and was obtained from the same 1992 household screening. The Children of the Depression Age (CODA) sample consists of those born between 1924 and 1930 (the 'missing' birth cohorts between the HRS and AHEAD samples), and were identified from the Medicare enrollment file.

The HRS is funded under a cooperative agreement between the National Institute on Aging (NIA) and the Survey Research Center at the University of Michigan.  The HRS is designed to study labor force, health, and family transitions of the U.S. population aged 51 and older, and the impact of those transitions on economic resources, claims on structured programs such as Social Security, Medicare, and Medicaid, and informal assistance and transfers to and from family members.

In the fall of 2001, two different questionnaires were mailed to two random sub-samples of the Health and Retirement Study (HRS) households interviewed in 2000.  One of the questionnaires, the 2001 Human Capital Mail Survey (HUMS), asked about parental economic investment in the education of children.  The HUMS included questions about children's educational attainment and the costs associated with attending college.

Funding for the 2001 HRS HUMS was provided by the National Institute on Aging at NIH (U01 AGO9740).  The Institute for Social Research (ISR) Survey Research Center at the University of Michigan conducted the survey.

## 1A. Sample Interviewed in the 2001 HRS HUMS

HUMS questionnaires were mailed to 3,862 households interviewed in the HRS 2000 core survey that were likely to have had at least one child (ever) 18 years of age or older. If a selected household had two panel members in it, one panel member was randomly selected to receive the questionnaire.

Data from 3,031 questionnaires are included in the 2001 HUMS, for a simple response rate of 78.5%. The simple response rate does not adjust for mortality or undeliverable questionnaires. The HRS 2002 Tracker file notes that 3,040 questionnaires were returned (see the variable MCSTATUS), which is nine more than contained in the 2001 HRS HUMS data. The nine cases were deleted from the 2001 HUMS data because the returned questionnaires did not contain responses to the questions.

## 1B. 2001 HRS HUMS Sections

The 2001 HRS HUMS contained one section, referred to as Section H: Child Educational Attainment and Expenses. The questions in Section H were designed to facilitate research in a number of areas including parental investment in the human capital of their children, the allocation of parental resources across their offspring, the effect of educational expenditures on mid-life savings of parents, etc.

Section H contains questions about high school attendance, college attendance (number, duration, type), and the proportion of college expenses (tuition, food, housing) paid for by the parent of the child. Respondents answered the questions for each child (up to and including 10 children). Section H concludes with a question about who filled out the questionnaire and how much time it took the respondent to fill out the questionnaire.

## 1C. File Structure

The HRS 2001 HUMS College Tuition Imputations (Restricted) data are constructed at the child level, or one record per child in a given household for which responses were obtained (n=10,437). The data file, HUMS2001TR, contains 10,437 cases and 29 variables.

The primary Identification variables (IDs) are:
```
HHID          HOUSEHOLD IDENTIFIER
PN            PERSON NUMBER
OPN           CHILD PERSON NUMBER
```

Secondary Identification variables are:
```
GSUBHH        2000 SUB-HOUSEHOLD IDENTIFIER
```

Records in the data files are sorted in order by HHID, GSUBHH and OPN. The variable PN refers to the respondent to whom the questionnaire was mailed. The variable OPN is the person number of a child and links to the child in the 2000 core data. There are 11 children (across seven different households) in the HUMS data that do not have OPNs in the 2000 core data. They are as follows:

```
          HHID            GSUBHH      OPN
          047596               0      112
```

```
                        047596               0      113
                        047596               0      114
                        047596               0      116
                        084119               2      160
                        084119               2      163
                        140925               0      124
                        201774               0      101
                        202238               0      114
                        207009               0      101
                        207281               0      116
```

Identification variables in the HRS 2001 HUMS College Tuition Imputations data are stored in character format. For further information about HRS identification variables and merging data, see the Data Description for the HRS 2000 Core.

## 1D. Data Quality

### Recoded or constructed variables:

The variables in the HRS 2001 HUMS College Tuition Imputations (Restricted) data mirror the questions in the HUMS questionnaire. However, occasionally an extension indicator was applied to a stem question/variable that could generate two answers. For example:

```
VARNAME
H1       Corresponds to the yes/no part of the H1 question asking "Did
         this child attend a private high school?" whereas,
H1a      Corresponds to the "how many years?" dimension of H1 if the
         answer to H1 was yes.
```

### Consistency:

The data have some inconsistencies. Respondents did not always follow the correct skip patterns and in most instances their "incorrect" answers were preserved. Likewise, there were times when respondents interpreted questions in different ways.

### Extreme values:

While cleaning the data, when extreme or unlikely values were encountered, the hard copies of the questionnaires were checked to determine if errors were introduced during the data entry process. If the respondent reported the extreme value, it appears unchanged in the data. The constructed variable for the year of college attendance (COLLYR) is a case in point. There are several respondents who gave values that indicate that the last year that a child attended college was 2002, 2003, or 2005 (two or more years after the HUMS survey was conducted). After careful review of these cases, it appears that respondents were optimistically looking ahead to when their children would graduate from college.

### Missing data:

When respondents did not provide an answer to a question, their answers were treated as missing. There are no codes for "don't know" or "refused". Instead, codes of "9" (or "99", etc. depending on the variable width) have been used to designate "Answer not given" when respondents answered a lead-in question, but did not provide answers to the follow-up

question(s).  For example, 2127 respondents said "Yes" at H1. PRIVATE HIGH SCHOOL, but only 1798 provided a response at H1a. NUM YRS PRIVATE HS.  Therefore, 329 respondents (2127-1798 = 329) were coded as "99. Answer not given".  Respondents for whom the question was inapplicable, are coded as blank or dot.

# 2. Obtaining the Data

## 2a. Restricted Data Agreement

This restricted data set is intended for exclusive use by you and the persons specified in the *Agreement for Use of Restricted Data from the Health and Retirement Study* and/or the *Supplemental Agreement with Research Staff for Use of Restricted Data from the Health and Retirement Study*. If you have technical questions about the content and use of this data set, refer to the HRS Restricted Data Web site or contact Michael A. Nolte at 734.936.1903 (voice) or manolte@isr.umich.edu  (email). All procedural questions should be directed to Cathy Liebowitz at 734.763.4180 (voice) or catlieb@isr.umich.edu (email).

## 2b. Access to Restricted Data

Although some HRS data sets are unrestricted, other HRS data sets are restricted, and are available only under specific contractual conditions. If you wish to obtain the HUMS 2001 College Tuition Imputations (Restricted) data set, you should begin the application process by reviewing the application materials and procedures at the HRS Restricted Data Web site. If you decide to go forward, you should submit a preliminary application package containing a research proposal and data protection plan to HRS. If problems are found with your application, the HRS staff members will contact you to provide assistance in resolving them. Otherwise, once you have been notified by HRS that your application is complete, you may submit your application to your local Institutional Review Board and/or Contracting Authority. Once IRB approval is obtained, you should submit the signed originals of the institutional approval documents to the HRS. At this point the application is complete and it is forwarded to the HRS Data Confidentiality Committee (DCC) for review and approval. If approval is granted by the DCC, the Principal Investigator of the HRS will sign the restricted data agreement, and you will be sent the HUMS 2001 Tuition Imputation File by the HRS. Note: the access code for this restricted data product is sent separately.

If your institutional or physical circumstances change, (e.g., new persons added to the project, hardware/software changes, or office moves) you should contact the HRS in order to modify the underlying agreement. A yearly re-certification of your restricted data agreement is required; if you wish to continue with analysis of the data after the expiration of the initial agreement, you will need to submit a renewal request. Finally, an HRS agent will periodically audit you in order to ensure that all conditions of the Restricted Data Agreement are being met.

## 2c. Publications Based on Restricted Data

Under terms of the Restricted Data Agreement, you have agreed to inform HRS of any papers, publications, or presentations based on this HRS restricted data set. Please send a copy of such publications, with a bibliographical reference if appropriate, to:

```
        Health and Retirement Study
```

```
        Attn: Papers and Publications
        The Institute for Social Research, Room 3050
        P.O. Box 1248
        Ann Arbor, Michigan  48106-1248
```

Alternately, you may contact us by e-mail at *hrsquest@isr.umich.edu* with "Attn: Papers and Publications" in the subject line.

# 3. Setup

## *3a. Distribution Set*

The files are packaged for distribution in a .ZIP file, *hums2001tr.zip*, which contains five sub-directories:

```
        hums2001tr\Data
        hums2001tr\Doc
        hums2001tr\SAS
        hums2001tr\SPSS
        hums2001tr\Stata
```

In order to keep the contents secure the .ZIP file has been encrypted using PGP (Pretty Good Privacy). For users working in a PC environment (Windows XP/2000), a PGP self-decrypting archive file is distributed; UNIX users receive a standard PGP file. Users who require special file formats should contact the HRS Help Desk at hrsquest@isr.umich.edu.

You'll typically want to decompress the data files, at least one set of program statement files, and the documentation files.

### 3a-1. PC environment (Windows XP/2000)

Run the self-decrypting archive file, *Hums2001trSda.exe*, from the command line.  When you are prompted for a pass-phrase, respond with the pass-phrase you received via e-mail.  The output from the self-extracting archive file will be the zipped file, *hums2001tr.zip*, containing the five-subdirectories listed above.  Use your file decompression software (i.e. Winzip)  to extract the file. If you have problems please contact the HRS Help Desk at hrsquest@isr.umich.edu.

### 3a-2. UNIX environment

Access the PGP software installed on your system to decrypt *Hums2001tr.pgp*. When you are prompted for a pass-phrase, respond with the pass-phrase you received via e-mail. Use appropriate file decompression software such as unzip to extract files from *Hums2001tr.zip*.  If you have problems please contact the HRS Help Desk at hrsquest@isr.umich.edu.

## *3b. Filename Suffixes*

For all datasets filename suffixes mean the following:

```
    File            Description
    *.DA            Data
    *.DCT           STATA dictionary
    *.DO            STATA "Do" file
```

```
    *.SAS          SAS program statements
    *.SPS          SPSS program statements
    *.TXT          ASCII text files, (data descriptions and codebooks)
```

## 3c. Directory Structure

While a particular setup is not required for using HRS files, we have traditionally suggested a
directory structure.  By using this directory structure, you will not have to change the path name
in your data descriptor files. If you use a different structure, just change the directory references
in the program files.

```
    Directory               Contents

    c:\hums2001tr\          Distribution files

    c:\hums2001tr\codebook\ hums2001tr.txt
                            hums2001tr_dd.pdf (this document)

    c:\hums2001tr\data\     hums2001tr.da

    c:\hums2001tr\sas\      hums2001tr.sas

    c:\hums2001tr\spss\     hums2001tr.sps

    c:\hums2001tr\stata\    hums2001tr.do
                            hums2001tr.dct
```

# 4. Program Statements

The HUMS 2001 College Tuition Imputations (Restricted) data file is provided in ASCII format,
with fixed-length records.   Use the associated SAS, SPSS or Stata program statements to read
the data file into the analysis package of your choice.

## 4A. Using the Files with SAS

To create a SAS system file for a particular dataset, two file types must be present for that
dataset -- .SAS program statement files and .DA data files. To create a SAS system file, load the
*.SAS file into the SAS Program Editor.

If the *.SAS file is located in "c:\hums2001tr\sas" and the data file is located in
"c:\hums2001tr\data", you can run the file as is.  A SAS system file (*.SD2 or *.SAS7BDAT)
will be saved to directory "c:\hums2001tr\sas". If the files are not located in the specified
directories, you will need to edit the *.SAS file to reflect the proper path names prior to running
the file.

## 4B. Using the Files with SPSS

To create an SPSS system file for a particular dataset, two file types must be present for that
dataset -- .SPS program statement files and .DA data files. To create an SPSS system file, open
the *.SPS file in SPSS as an SPSS Syntax File.

If the *.SPS file is located in "c:\hums2001tr\spss" and the data file is located in "c:\hums2001tr\data", you can run the file as is.  An SPSS system file (*.SAV) will be saved to directory "c:\hums2001tr\spss". If the files are not located in the specified directories, you will need to edit the *.SPS file to reflect the proper path names prior to running the file.

### 4C. Using the Files with Stata

To use Stata with a particular dataset, the following three file types must be present for that dataset -- .DCT files, .DO files, and .DA data files.

Files with the suffix .DA contain the raw data for Stata to read.  Files with the suffix .DCT are Stata dictionaries used by Stata to describe the data. Files with the suffix .DO are short Stata programs ("do files") which you may use to read in the data.  Load the .DO file into Stata and then submit it.

If the *.DO and *.DCT files are located in "c:\hums2001tr\stata" and the data file is located in "c:\hums2001tr\data", you can run the .DO file as is. If the files are not located in these directories, you must edit the *.DO and *.DCT files to reflect the proper path names before you run the files.

Note that the variable names provided in the .DCT files are uppercase.  If you  prefer lower case variable names, you may wish to convert the .DCT files to lower case prior to use.  You may do this by reading the .DCT file into a text or word processing program and changing the case.  For instance in Microsoft Word the following sequence will work: *Edit > Select All > Format > Change Case > lowercase*.

# 5. Linking Children to Household- and Child-level Files

Child records in the HUMS 2001 College Tuition Imputations (Restricted) data set can be linked to household records from 2000 and prior waves by HHID and GSUBHH. The sub-household identifiers, ASUBHH for 1992, BSUBHH for 1993, C2SUBHH for 1994, DSUBHH for 1995, ESUBHH for 1996, FSUBHH for 1998 and GSUBHH for 2000 can be used to link household data with the cross-sectional respondent data.  When linking the child records in HUMS to the child records in 2000 use HHID, GSUBHH and OPN.

### 5A. Merge Example 1 (SAS)

HRS 2000 core household level data with HUMS 2001 College Tuition Imputations (Restricted) data

```
    libname hums 'c:\hums2001tr\data\';
    libname hrs00 'c:\hrs2000\data\';

    proc sort data= hums.hums01_c out= hums01_c;
      by hhid gsubhh;
    run;
    proc sort data= hrs00.h00d_h out= h00d_h(keep=hhid gsubhh G1980 G1983);
      by hhid gsubhh;
    run;

    data new;
```

```
      merge hums01_c(in=a) h00d_h(in=b);
      by hhid gsubhh;
      if a;
    run;
```

## 5B. Merge Example 2 (SAS)

HRS 2000 core child level data with HUMS 2001 College Tuition Imputations (Restricted)

```
    libname hums 'c:\hums2001tr\data';
    libname hrs00 'c:\hrs2000\data';

    proc sort data=hums.hums01_c out=hums01_c;
      by hhid gsubhh opn;
    run;
    proc sort data=hrs00.h00pr_mc(keep=hhid gsubhh opn G10) out=h00pr_mc;
      by hhid gsubhh opn;
    run;

    data new;
      merge hums01_c(in=a) h00pr_mc(in=b);
      by hhid gsubhh opn;
      if a;
    run;
```

# 6. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines for using the HUMS 2001 College Tuition Imputations (Restricted) data release. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

## 6a. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access public data, metadata, and other relevant information, point your Web browser to the HRS Web site: http://hrsonline.isr.umich.edu/. For restricted data information, visit the HRS Restricted Data Web site.

## 6b. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

Internet: Help Desk at our Web site

E-mail: hrsquest@isr.umich.edu

Postal service:
Health and Retirement Study
The Institute for Social Research, Room 3050
The University of Michigan

P.O. Box 1248
Ann Arbor, MI 48106-1248

FAX: (734) 647-1186