Data Description
2001 Human Capital Mail Survey (HUMS)

1.   Overview
In the fall of 2001, two different questionnaires were mailed to two random subsamples 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.

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.

Funding for the 2001 HRS HUMS (Early Release, Version 1.0) 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.

By receiving the dataset, you agree to use it for research and statistical purposes only and make no effort to identify respondents.   In addition, you agree to send us a copy of any publications you produce based on the data. See Obtaining the Data (at the end of this document for additional details).

1A.   The Sample Interviewed in the 2001 HRS HUMS

In the fall of 2001, HUMS questionnaires were mailed to 3,862 households interviewed in the HRS 2000 core survey, and who 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 Early Release data (Version 1.0), 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 file. The nine cases were deleted from the 2001 HUMS Early Release 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.

It is important to note that while the 2001 HRS HUMS questionnaire asked respondents to provide the name, city, and state for colleges attended by each child, these data are NOT included in the 2001 HUMS data. Also, the expenses associated with attending a given college in a given year are NOT included in the 2001 HUMS data. The expenses associated with attending a given college will be calculated and released in the future. The name, city and state variables will not be released in public files.

1C.   File Structure

The 2001 HRS HUMS Early Release (Version 1.0) data are at the child level, or one record per child in a given household for which responses were obtained (n=10,437).

The data file is called HUMS01_C, loosely translated as: Human Capital Mail Survey 2001, with the "_C" designating the file is at the child level.

The HUMS01_C data file contains 10,437 cases and 22 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 Early Release 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 2001 HUMS 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 will have some inconsistency. 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.   Distribution Files

The following extensions are used for the six different types of distribution files:

.DA for data files,
.SAS for SAS program statements,
.SPS for SPSS program statements,
.DO for Stata DO statements,
.DCT for Stata dictionary statements, and
.TXT for codebook files.

For example,

HUMS01_C.DA contains ASCII data from Sections H. HUMS01_C.SAS contains corresponding SAS program statements,
HUMS01_C.SPS contains corresponding SPSS program statements,
HUMS01_C.DO contains corresponding Stata DO statements,
HUMS01_C.DCT contains corresponding Stata dictionary statements, and
HUMS01_C.TXT contains the ASCII codebook.

The 2001 HRS HUMS Early Release data are provided in ASCII format, with fixed-length records. Use the associated SAS, SPSS or Stata program statements to read the data into the analysis package of your choice. In addition, you will probably want to download the codebook file (HUMS01_ C.TXT) and the data description (this document).

3. Program Statements

3A.   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:\HUMS2001\sas" and the data file is located in "c:\HUMS2001\data", you can run the file as is. A SAS system file (*.SD2 or *.SAS7BDAT) will be saved to directory "c:\HUMS2001\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.

3B.   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:\HUMS2001\spss" and the data file is located in "c:\HUMS2001\DATA", you can run the file as is. An SPSS system file (*.SAV) will be saved to directory "c:\HUMS2001\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.

3C.   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:\HUMS2001\stata" and the data file is located in "c:\HUMS2001\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, Edit, Select All, Format, Change Case, lowercase.

4.   Linking Children to Household-level and Child-level Files

4A. Child records in the HUMS Early Release (Version 1.0) 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.

4B. Merge Example (SAS): HRS 2000 core household level data with 2001 HRS HUMS Early Release data (Version 1.0)

```
libname hums 'c:\HUMS2001\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;
```

4C. Merge Example (SAS): HRS 2000 core child level data with 2001 HRS HUMS
Early Release data (Version 1.0)

```
libname hums 'c:\HUMS2001\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;
```

5.   Registration and Downloading the Data

5A. HRS data are available for free to researchers and analysts at the HRS Web site.
In order to obtain public release data, you must first register at our Web site.
Once you have completed the registration process, your username and password
will be sent to you via e-mail. Your username and password are required to
download any data files.

By registering all users, we are able to document for our sponsors the size and
diversity of our user community allowing us to continue to collect these
important data. Registered users receive user support, information related to
errors in the data, future releases, workshops, and publication lists. The
information you provide will not be used for any commercial use, and will not be
redistributed to third parties.

5B.   Conditions of Use
By registering, you agree to the Conditions of Use governing access to Health
and Retirement public release data. You must agree to

- not attempt to identify respondents
- not transfer data to third parties except as specified
- not share your username and password
- include specified citations in work based on HRS data
- provide information to us about publications based on HRS data
- report apparent errors in the HRS data or documentation files
- notify us of changes in your contact information

For more information concerning privacy issues and conditions of use, please read "Conditions of Use for Public Data Files" and "Privacy and Security Notice" at the Public File Download Area of the HRS Web site.

5C.   Publications Based on Data
As part of the data registration process, you agree to include specified citations and to inform HRS of any papers, publications, or presentations based on HRS data. Please send a copy of any publications you produce based on HRS data, with a bibliographical reference, if appropriate, to the address below.

Health and Retirement Study
Attn: Papers and Publications
The Institute for Social Research, Room 3050
P.O. Box 1248
Ann Arbor, MI (USA) 48106-1248

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

6.   If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines to using the HUMS 2001 Early Release (Version 1.0) data. We recommend that you also read the data description for the HRS 2000 Final Release (Version 1.0). If you have questions or concerns that are not adequately covered here or on our Web site, 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 the data and other relevant information, point your Web browser to the HRS Web site.

   http://hrsonline.isr.umich.edu/

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