

HEALTH AND RETIREMENT STUDY

**Respondent Cross-Year Detail Earnings**

Data Description and Usage

Version 4.2, May 2015

To the Restricted Data Investigator: 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 there are any questions about this data set and its use, refer to the restricted data pages on the HRS Web site (<http://hrsonline.isr.umich.edu/rda>). You can also contact the HRS Help Desk via electronic mail (<mailto:hqsquestions@umich.edu>) or on the Web (<http://hrsonline.isr.umich.edu/help>).

This document may not be reproduced without the written consent of the staff of the Health and Retirement Study, the Institute for Social Research, the University of Michigan.

## Table of Contents

<b>1. Overview and Acknowledgments.....</b>	<b>2</b>
<b>2. Obtaining the Data.....</b>	<b>2</b>
2a. Restricted Data Agreement.....	2
2b. Access to Restricted Data .....	2
2c. Publications Based on Restricted Data .....	2
<b>3. Setup .....</b>	<b>3</b>
3a. Distribution Set.....	3
3a-1. Windows Environment .....	3
3a-2. UNIX/Linux Environment .....	3
3a-3. Macintosh OS X Environment (10.4.x and above).....	3
3d. Program Statements .....	4
3d-1. Using the Files with SAS.....	4
3d-2. Using the Files with SPSS .....	4
3d-3. Using the Files with Stata.....	4
<b>4. Data Set Content .....</b>	<b>5</b>
<b>5. If You Need to Know More.....</b>	<b>5</b>
5a. HRS Internet Site .....	5
5b. Contact Information.....	6
<b>Appendix A: Data Management Notes.....</b>	<b>7</b>
<b>Appendix B: Cohort Permission Status.....</b>	<b>8</b>

## 1. Overview and Acknowledgments

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 the National Institutes of Health (U01AG009740), 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 HRS gratefully acknowledges the special assistance of the Social Security Administration's Office of Research and Statistics for their assistance in retrieving the administrative records of HRS respondents who gave consent for those records to be used for research purposes.

## 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 questions about the content and use of this data set, refer to the [HRS Restricted Data Web](#) site or contact the [HRS Help Desk](#) by email.

### 2b. Access to Restricted Data

Although most HRS data sets are available to the public without restriction, certain HRS data sets contain sensitive respondent information and are only available under terms of a formal agreement negotiated between the researcher and HRS. If you are a researcher who has found a restricted data file that is of interest, you should begin the application process by reviewing the application materials and procedures. 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. After 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 restricted data product(s) by the HRS. Note: access codes for your restricted data product(s) are 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. Note: 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

Your restricted data agreement specifies that you will inform HRS of any papers, publications, or presentations based on this restricted data set. Please send a copy of such publications in PDF format via

e-mail to [hqsquestions@umich.edu](mailto:hqsquestions@umich.edu) with "Attn: Papers and Publications" in the subject line. If you wish, you may include a bibliographical reference.

As an alternative, you may transmit publications in paper format by postal mail:

Health and Retirement Study  
Attn: Papers and Publications  
The Institute for Social Research, Room 3410  
P.O. Box 1248  
Ann Arbor, Michigan 48106-1248

## 3. Setup

### 3a. Distribution Set

This data set is packaged for distribution in a ZIP file, `XDetErn.zip`. In order to keep the contents secure, the ZIP file has been encrypted using **WinZIP** 256 bit AES encryption. Extract the data file(s), the program statement file(s) matching your analysis environment, the data description (this file), and the codebook file. If you require a special file format or experience system problems, please contact the [HRS Help Desk](#). If all files are decompressed, they will require approximately 130MB of free space on your storage device.

#### *3a-1. Windows Environment*

Copy the ZIP file to the Windows folder where you plan to do your work. Do not use the built-in Windows decompression utility; instead use a third-party tool such as **7-Zip** or **WinZIP** to extract the ZIP folder contents. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output will be the files listed in Table 1.

#### *3a-2. UNIX/Linux Environment*

Copy the ZIP file to the folder where you plan to do your work. Use the ZIP file decompression software installed on your system, (e.g. **7-Zip**, **gunzip**) to decrypt and decompress the ZIP folder contents. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output will be the files listed in Table 1.

#### *3a-3. Macintosh OS X Environment (10.4.x and above)*

Copy the ZIP file to the folder where you plan to do your work and use **stuffit** to decrypt and decompress the ZIP folder contents. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output will be the files listed in Table 1.

<b>Directory<sup>1</sup></b>	<b>File</b>	<b>Type</b>
c:\xyrdetern	XDetErn.zip	zip file
c:\xyrdetern\docs\	XyrDetErn2012.txt	Codebook file (ASCII text)
c:\xyrdetern\data\	XyrDetErn2012.da (n=552291)	Data file (ASCII text)
c:\xyrdetern\sas\	XyrDetErn2012.sas	SAS program statements
c:\xyrdetern\spss\	XyrDetErn2012.sps	SPSS program statements
c:\xyrdetern\stata\	XyrDetErn2012.dct/.do	Stata dictionary and "do" files

### 3d. Program Statements

Each data file comes with associated SPSS, SAS or Stata program statements to read its contents. Files containing SPSS statements are named with an .sps extension, those with SAS statements with an .sas extension, and those with Stata statements with .do and .dct extensions.

#### 3d-1. Using the Files with SAS

To build SAS system files for a restricted data set, two file types must be present for that data set -- .sas program statement files and .DA data files. Create SAS system files by loading the \*.sas file into the SAS Program Editor and submitting the commands. If the \*.sas file is located in "c:\xyrdetern\sas" and the data file is located in "c:\xyrdetern\data", you can run the file as is. A SAS system file will be saved to directory "c:\xyrdetern\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.

#### 3d-2. Using the Files with SPSS

To build SPSS system files for a restricted data set, two file types must be present for that data set -- .sps program statement files and .da data files. Create SPSS system file by opening the \*.sps file in SPSS as an SPSS Syntax File and submitting the commands. If the \*.sps file is located in "c:\xyrdetern\spss" and the data file is located in "c:\xyrdetern\data", you can run the file as is. An SPSS system file (\*.sav) will be saved to directory "c:\xyrdetern\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.

#### 3d-3. Using the Files with Stata

To use Stata with a particular data set, the following three file types must be present for that data set -- .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:\xyrdetern\stata" and the data file is located in "c:\xyrdetern\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.

<sup>1</sup> While a specific folder arrangement is not required for using HRS files, we have traditionally suggested a directory structure for the Windows environment. By using this directory structure, you will not have to change the path name references in your data descriptor files. If you do not use Windows, or if you use a different directory structure, just change the directory references in the program files.

## 4. Data Set Content

Version 4.1 of the *Respondent Cross-Year Detail Earnings* data set contains Master Earnings File (MEF) earnings data for 19,626 respondents derived from eight permission sets as summarized below:

Primary Cohort	Permission Set								Total
	1992	1998	2000	2004	2006	2008	2010	2012 <sup>2</sup>	
<i>HRS</i>	3471	14	5	2195	43	2338	1837	424	10327
<i>AHEAD</i> <sup>3</sup>	11	12	7	99	11	428	253	0	821
<i>CODA</i>	0	495	63	169	439	265	120	2	1553
<i>War Baby</i>	0	377	41	200	421	499	289	196	2023
<i>EBB</i>	0	0	0	600	146	509	1066	549	2870
<i>MBB</i>	0	0	0	0	0	0	1657	347	2004
<i>N.A.</i>	0	0	0	0	0	0	1	1	2
<b>Total</b>	3482	898	116	3263	1060	4039	5223	1519	19600

This version of the *Respondent Cross-Year Detail Earnings* data set contains additional variables that supplement the set contained in the previous version (3.0). In addition the duplicate records that existed in v4.0 have been removed. For more information, see the latest codebook. The data file contains one record per employer<sup>4</sup> per year for each consenting respondent who received wages during years 1978-2012,<sup>5</sup> and whose records were available in the MEF. See Appendix B for more information on cohort permission timing.

To help preserve respondent confidentiality, earnings data have been rounded to the nearest \$100, except for fields that were originally between \$0.01 and \$49.99, which have been recoded to **.N**. Entries that were originally coded as 0 remain as 0. Observations that are missing due to absence of information for a given year are coded **.M**. Amounts greater or equal to \$250,000 or less than or equal to -\$250,000 were top/bottom coded. See Appendix A for further details.

## 5. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines for using this data set. 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.

### 5a. HRS Internet Site

The Health and Retirement Study maintains a Web site (<http://hrsonline.isr.umich.edu>) that provides access to public release data, questionnaires, codebooks, sample and response rate documentation, a bibliography, user guides, and extensive background information. If you want to find out more about

<sup>2</sup> Prospective permissions began in 2006

<sup>3</sup> Note: This data set treats HRS Wave 1 "overlap" cases as members of the AHEAD cohort. There were a number of original HRS Wave 1 households eligible to be either an HRS or AHEAD household. An interview was attempted for each in HRS Wave 1. Afterwards, a random sub-sampling was performed, with 60% of the cases staying in HRS and the remaining going to AHEAD. We refer to the 134 cases given over to AHEAD as the HRS/AHEAD overlap. Detail earnings information was not obtained for AHEAD respondents whose permission was sought in 1993/1995.

<sup>4</sup> Employer identifiers are scrambled and are unique to each respondent.

<sup>5</sup> **Detail earnings records for 1978 and 1979 are incomplete and should be used with caution, if at all.**

restricted data products and procedures, visit the restricted data pages (<http://hrsonline.isr.umich.edu/rda>).

## **5b. 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 (<http://hrsonline.isr.umich.edu/help>)

E-mail: [hqsquestions@umich.edu](mailto:hqsquestions@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

## Appendix A: Data Management Notes

Table 3: Missing Data Codes			
Bottom/Top Code Category	SAS/Stata Special Missing	Numeric Missing	Numeric Recode
<= -\$500,000	.a	-999998	
-\$300,000 to 499,999	.b	-999997	
-\$250,000 to 299,999	.c	-999996	
-1 to -49	.m	-2	-25
1 to 49	.n	-1	25
\$250,000 to 299,999	.x	999996	272000
\$300,000 to 499,999	.y	999997	370000
>= \$500,000	.z	999998	1000000

To preserve respondent confidentiality, all data elements are rounded and top/bottom-coded. A code of zero (0) represents a true zero dollar amount. A code of .N represents an amount between 1 and 49 dollars, inclusive. A code of .M represents an amount between -1 and -49 dollars, inclusive. Amounts with an absolute value of 50 dollars or more were rounded to the nearest 100 dollars. An absolute value ending in 01 to 49 was rounded down, and a value ending in 50 to 99 was rounded up.

Amounts greater or equal to \$250,000 were recoded as follows:

\$250,000 to 299,999 = .X;  
 \$300,000 to 499,999 = .Y;  
 >= \$500,000 = .Z.

Amounts less than or equal to -\$250,000 were recoded as follows:

-\$250,000 to -299,999 = .C;  
 -\$300,000 to -499,999 = .B;  
 <= -\$500,000 = .A.

SPSS users may wish to convert the SAS special missing codes to SPSS equivalents. The following code fragment serves as an example:

```
* input WTOTCOMP as a string variable named xWTOTCOMP .
compute WTOTCOMP=0.
compute WTOTCOMP=number(xWTOTCOMP, F10).
recode xWTOTCOMP (convert) ('.n'=25) ('.x'=999996)
('y'=999997) ('.z'=999998) into WTOTCOMP.
missing values WTOTCOMP (999996, 999997, 999998).
desc WTOTCOMP.
execute.
```

## Appendix B: Cohort Permission Status

Cohort	1992/94/96	1993/95	1998/2000	2004	2006: Asked If Face-to-face and...	2008: Asked if Face-to-face and...	2010: Asked if Face-to-face and...	2012: Asked if Face-to-face and...
New spouses or never-interviewed R in any cohort	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all
<b>AHEAD</b> (Born <=1923)	NO	Asked all	NO	NO	NO	NO	NO	NO
<b>CODA</b> (1924-1930)	NA	NA	Asked all	NO	R was not asked in 2004	R refused, did not provide valid permission information, was not asked, or was non-response in 2004 or 2006	R refused, did not provide valid permission information, was not asked, or was non-response in 2006 or 2008	R refused, did not provide valid permission information, was not asked, or was non-response in 2006, 2008 or 2010
<b>HRS</b> (1931-1941)	Asked all	NO	NO	Asked all	R was not asked in 2004			
<b>WAR BABY</b> (1942-1947)	NA	NA	Asked all	Only asked refusers	R was not asked in 2004			
<b>EARLY BOOMERS</b> (1948-1953)	NA	NA	NA	Asked all	R refused or was non-response in 2004	R refused, did not provide a valid permission form or was non-response in 2004 or 2006	R refused, did not provide a valid permission form or was non-response in 2006 or 2008	R refused, did not provide a valid permission form or was non-response in 2006, 2008 or 2010
<b>MID BOOMERS</b> (1954-????)	NA	NA	NA	NA	NA	NA	NA	NA

The *Respondent Cross-Year Summary Earnings* dataset is derived from Master Beneficiary Record (MBR) data for primary beneficiary and other (secondary) insured. The inputs to this dataset were received at different points in time over the past decade from the Social Security Administration:

- In 1992, 1994 or 1996 for HRS cohort members
- In 1993 or 1995 for AHEAD cohort members
- In 1998 or 2000 for CODA and War Baby cohort members as well as new spouses
- In 2004 for EBB and HRS cohort members, new spouses, and War Baby 1998/2000 refusers
- In 2006, 2008, 2010 or 2012 for respondents interviewed face-to-face who were not asked previously or refused or were non-response. These permissions are prospective and provide for biennial updates through 2030.

Prior to 2006, respondent permissions were based on separate agreements between HRS and SSA. This meant that information was obtained only for the time period covered by the agreement. In 2006 and thereafter, the terms of the permission agreement allow HRS to obtain biennial updates of earnings and benefits data through the current year. Since each new record set obtained from SSA reflects the most recent information available, the latest version of the *Respondent Cross-Year Detail Earnings Earnings* data set will always be the most accurate. If current information for a given respondent is missing, the most recent data elements from past datasets are used.