

# HEALTH AND RETIREMENT STUDY

## **Restricted Data**

### **Cancer Site**

All HRS Cohorts, 1995-2018

Data Description and Usage

November 2021 Version

**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 HRS Restricted Data Web Site (<http://hrsonline.isr.umich.edu/rda>) or contact the HRS Help Desk ([hrsquestions@umich.edu](mailto:hrsquestions@umich.edu)).

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</b> .....	<b>2</b>
<b>2. Obtaining the Data</b> .....	<b>2</b>
2a. Access to Restricted Data.....	2
2b. Restricted Data Agreement.....	2
2c. Publications Based on Restricted Data.....	2
<b>3. Data Set Content</b> .....	<b>2</b>
3a. Errata .....	4
<b>4. If You Need to Know More</b> .....	<b>4</b>
4a. HRS Internet Site .....	4
4b. Contact Information.....	4
<b>Appendix: Installation (Traditional License Only)</b> .....	<b>5</b>
1. Distribution Set .....	5
1-1. Windows Environment.....	5
1-2. UNIX/Linux Environment .....	5
1-3. Macintosh OS X Environment (10.4.x and above) .....	5
Table A-1: Contents of Distribution Package.....	6
2. Program Statements.....	6
2-1. Using the Files with SAS .....	6
2-2. Using the Files with SPSS.....	6
2-3. Using the Files with Stata .....	6
3. Non-Windows Environments.....	7
3-1. SPSS in an OSX environment.....	7
3-2. STATA in an OS X Environment .....	7

## 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 AG009740), with supplemental support from the Social Security Administration. The study is conducted by the Institute for Social Research (ISR) at the University of Michigan.

## 2. Obtaining the Data

### 2a. 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. Prospective users of HRS restricted geocode data have two access options:

- [MiCDA Enclave](#) Virtual Desktop Infrastructure (recommended)
- Traditional Licensing Agreement (deprecated)

For instructions on how to proceed, visit the [HRS Restricted Data Web](#) site or contact the HRS Restricted Data Applications Processing Team ([hrrsdaapplication@umich.edu](mailto:hrrsdaapplication@umich.edu)) by email.

### 2b. Restricted Data Agreement

This restricted data set is intended for exclusive use by you and the persons specified in the *Confidentiality Agreement for Use of Restricted Data from the Health and Retirement Study*. Traditional licensing agreements that include multiple users are also bound by the *Supplemental Agreement with Research Staff for Use of Restricted Data from the Health and Retirement Study*.

### 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. You may send a bibliographical reference (including a URL link whenever possible) for each item to [hrrsdaapplication@umich.edu](mailto:hrrsdaapplication@umich.edu) with "Attn: Papers and Publications" in the subject line. If possible, you should also include a PDF-formatted copy of the publication.

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 3450  
P.O. Box 1248  
Ann Arbor, Michigan 48106-1248

## 3. Data Set Content

This restricted data set contains cancer site information collected in biennial interviews (through 2018) with the exception of interview years 1992, 1993, 1994 (Core/Exit), and 1996 (Exit). Information collected at these time points is included in the publicly released data available (with registration) at the HRS Web site. The data set matches the 2018 Tracker File and is keyed on HHID and PN.

The dataset was first released as a cumulative file after the 2010 wave. Since Core and Exit interviews are mutually exclusive, they were combined into a single wave specific variable. For 1998 and 2000, where there are two possible mentions, the Exit interview was combined under the first mention.

<b>Data Collection</b>	<b>Original Cancer Site Variables</b>	<b>Released Variable Name(s)</b>	<b>Release Type</b>	<b>Question</b>
1995	Core D810 Exit N810	D810	Restricted	<i>B5g. In which organ or part of your body did your cancer(s) start?</i>
1996	Core E810	E810	Restricted	<i>B5g. In which organ or part of your body did your cancer(s) start?</i>
1998	Core F1138M1, F1138M2 Exit Q1138	F1138M1, F1138M2	Restricted	<i>B5g. In which organ or part of your body did your cancer(s) start?</i>
2000	Core G1271M1, G1271M2 Exit - R1183	G1271M1, G1271M2	Restricted	<i>B5g. In which organ or part of your body did your cancer(s) start?</i>
2002	Core HC026 Exit SC026	HC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2004	Core JC026 Exit TC026	JC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2006	Core KC026 Exit UC026	KC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2008	Core LC026 Exit VC026	LC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2010	Core MC026 Exit WC026	MC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2012	Core NC026 Exit XC026	NC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2014	Core OC026 Exit YC026	OC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2016	Core PC026 Exit ZC026	PC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>
2018	Core QC026 Exit XQC026	QC026	Restricted	<i>C026. In which organ or part of your body did your cancer(s) start?</i>

### 3a. Errata

As part of this update, corrections were made to the cancer site variables that were previously missing data. A list of the affected waves and the number of records updated is listed below:

Wave	Number of Records
2006	1
2010	454
2012	1
2014	2

## 4. If You Need to Know More

This brief overview is intended to provide guidelines for using the *Cancer Site* 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.

### 4a. 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 or to find out more about restricted data products and procedures, visit the [HRS Web site](https://hrs.isr.umich.edu).

### 4b. Contact Information

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

Internet: Help Desk at the HRS Web site (<https://hrs.isr.umich.edu/help>)

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

Postal Service:

Health and Retirement Study  
The Institute for Social Research  
426 Thompson Street, 3450 ISR  
Ann Arbor, Michigan 48104

## Appendix: Installation (Traditional License Only)

### 1. Distribution Set

The *Cancer Site* data set is packaged for distribution in a ZIP file, **CSxwave18.zip**. In order to keep the contents secure, the .ZIP file has been encrypted using WinZIP 256 bit AES encryption. SAS, Stata, and SPSS system files are provided as part of this distribution set and can be used as is. Extract these files plus the data description (this file), and the codebook file.

If you wish to build your own versions of the system files, you will need to extract the ASCII text 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 10MB of free space on your storage device.

#### 1-1. Windows Environment

Copy the ZIP file to the Windows folder where you plan to do your work Use a third-party<sup>1</sup> file compression/decompression tool such as **WinZIP** or **7-zip** 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 A-1.

#### 1-2. 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. **7z**, **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 A-1.

#### 1-3. 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 A-1.

**Note: MiCDA Enclave Virtual Desktop Environment users are given access to pre-built SAS, Stata and SPSS versions of this dataset; therefore the information in this appendix does not apply to such users.**

---

<sup>1</sup> The built-in Windows decompression utility will not process AES-256bit encrypted zip files; it halts with "an unexpected error is keeping you from copying the file".

*Table A-1: Contents of Distribution Package*

<b>Directory</b>	<b>File</b>	<b>Type</b>
.\\cs\\docs\\	CSXWAVE18.txt	Codebook files (ASCII text)
.\\cs\\ascii\\data\\	CSXWAVE18.da	Data files (ASCII text)
.\\cs\\ascii\\sas\\	CSXWAVE18.sas	SAS program statements
.\\cs\\ascii\\spss\\	CSXWAVE18.sps	SPSS program statements
.\\cs\\ascii\\stata\\	CSXWAVE18.dct/CSXWAVE18.do	Stata dictionary and "do" files
.\\cs\\built\\sas\\	CSXWAVE18.sas7bdat	SAS system file
.\\cs\\built\\spss\\	CSXWAVE18.sav	SPSS system file
.\\cs\\built\\stata\\	CSXWAVE18.dta	Stata system file

## **2. Program Statements**

This restricted data product is distributed with pre-built versions of SAS, Stata, and SPSS system files. Once these have been extracted from the distribution medium no further action is necessary. Users who wish to build system files using the included ASCII text file(s) and SPSS, SAS or Stata program statements should read on. <sup>2</sup>

### *2-1. Using the Files with SAS*

To create a SAS system file for a particular data set, two file types must be present for that data set -- .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:\\cs\\ascii\\sas" and the data file is located in "c:\\cs\\ascii\\data", you can run the file as is. A SAS system file will be saved to directory "c:\\cs\\ascii\\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.

### *2-2. Using the Files with SPSS*

To create an SPSS system file for a particular data set, two file types must be present for that data set -- .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:\\cs\\ascii\\spss" and the data file is located in "c:\\cs\\ascii\\data", you can run the file as is. An SPSS system file will be saved to directory "c:\\cs\\ascii\\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.

### *2-3. 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:\\cs\\ascii\\stata" and the data file is located in "c:\\cs\\ascii\\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>2</sup> While a specific setup is not required for using HRS files, we have traditionally suggested a directory structure for the Windows environment to be used in building system files. By using this directory structure (or a Unix equivalent), you will not have to change the path name references in your data descriptor files. If you use a different structure, just change the directory references in the program files.

### 3. Non-Windows Environments

Non-Microsoft users should modify the default Windows file structure syntax to match that of their own operating system. The following examples should work for both Macintosh OS X and any Unix/Linux distribution. Open the SAS program file(s), SPSS syntax file(s) or the Stata do/dct files in an ASCII editor and make the changes indicated below.

#### 3-1. SPSS in an OSX environment

In this example, we assume that the user has extracted the files from the *Respondent Date of Birth* data set and placed them in a **Desktop** folder called **cs18** with the ASCII data file stored in subfolder **data** and the syntax file in subfolder **spss**. Then the commands in the syntax file would be modified to look like this:

```
FILE HANDLE cs18 /name='Desktop/cs18/data/csxwave18.da' RECL=257.
DATA LIST FILE= cs18/
HHID 1-6(A)
[rest of syntax file goes here]
.
execute.
SAVE /outfile 'Desktop/cs18/spss/csxwave18.sav'.
Execute.
```

#### 3-2. STATA in an OS X Environment

In the following example we assume that:

- The username is “user1”
- The encrypted zip file containing *Respondent Date of Birth* data has been copied to the user’s desktop from the CDROM sent by HRS.
- The user has decrypted /decompressed the zip file (use Stuffit for OS X) into a desktop folder named **cs18**
- The statistical package is stata

File CSXWAVE18.do should be modified as follows:

*Change...*

```
infile using c:\cs\ascii\stata\csxwave18.dct
```

*To...*

```
infile using /Users/user1/Desktop/cs18/stata/csxwave18.dct
```

*Change...*

```
save c:\cs\ascii\stata\csxwave18.dta
```

*To...*

```
save /Users/user1/Desktop/cs18/stata/csxwave18.dta
```

File CSXWAVE18.dct should be modified as follows:

*Change...*

```
dictionary using c:\cs\ascii\data\csxwave18.da {...}
```

*To...*

```
dictionary using /Users/user1/Desktop/cs18/data/csxwave18.da {...}
```