

HEALTH AND RETIREMENT STUDY

2003 Diabetes Study
Version 2.0, April 2007
(Sensitive Health Data)

Data Description and Usage

To the researcher: This data set is intended for exclusive use by you under the terms specified in the *Sensitive Health Data Use Agreement*. If there are any questions about its use, please contact the HRS Help Desk (hqsquest@isr.umich.edu).

This document may be reproduced only with the written consent of the staff of the Health and Retirement Study, Institute for Social Research, University of Michigan.

1. Introduction

1a. Background

The HRS 2003 Diabetes Study was funded by the National Institute on Aging through a competing supplement to the Health and Retirement Study. Its aims were to collect self-reported questionnaire data on aspects of treatment and self-management of diabetes, and to collect a clinical biomarker of glucose control: glycosylated hemoglobin, or HbA1c. The survey questions were drawn from several sources, including validated instruments from the Michigan Diabetes Research and Training Center (<http://www.med.umich.edu/mdrtc/survey/index.html>). The blood spot assays for HbA1c were performed by Flexsite Diagnostics, Inc. (<http://www.flexsite.com>). Both data collections were done through mail surveys.

There were 3,194 interviewed respondents in the 2002 HRS who reported a diagnosis of diabetes (including respondents whose interviews were given by proxies). Of these, 680 were excluded from the 2003 Diabetes Study because of their participation in the Consumption and Activities Mail Survey (CAMS). This exclusion is random, with the slight exception that proxy cases from 2000 were ineligible for CAMS but eligible for the Diabetes Study, so they are represented at slightly higher rates prior to weighting adjustments.

Of the 2,514 eligible 2002 participants, 129 were subsequently determined to have died prior to the October 2003 start of the Diabetes Study, and so are ineligible for inclusion in the sample. Of the 2,385 remaining eligible cases, 1,901 returned questionnaires, for a response rate of 79.7%.

A total of 1,233 valid blood spots were returned for HbA1c assays. That is 64.9% of those who returned questionnaires, and 51.8% of all eligible cases.

1b. File Structure

The 2003 Diabetes Study (Version 1.0) data consist of two respondent-level files. Each person is uniquely identified in the files by a diabetes study identifier (QXDIABID) or by the combination of a household ID (HHID) and person number (PN). The primary identification variables used in the HRS (HHID, PN) are included in both files to facilitate merging with other HRS data sets.

The sample data file (DIAB2003S) contains the 3,194 respondents who reported diabetes in 2002 and includes a variable indicating whether each was sampled for the participation in the 2003 Diabetes Study (DIABSAMP), sample section weights, and relevant identifiers. This file is provided so that analysts can identify the non-respondents for analysis if desired.

The questionnaire data file (DIAB2003A) contains one record for each respondent or proxy informant who completed and returned a questionnaire. This file also includes the results of the HbA1c blood kit test for respondents who choose to participate in this portion of the study.

Codebooks for each data file, the questionnaire, and this document may be viewed on the HRS public Web site.

1c. Data Quality

1c-1. Recoded and constructed variables

The majority of variables in the questionnaire data file match the code frame structure of the questions in the Diabetes Study questionnaire. However, questions that allowed for multiple mentions have been recoded such that each possible response is represented as a dichotomous variable. In addition, new variables have been created to categorize responses to open ended questions.

An imputed variable representing the Total Illness Burden Index (TIBI) score of diabetes comorbidities is also included in the file. The TIBI is a composite measure which was developed to assess comorbid illness in older adults using self-reported symptoms and medical event history (for more info see: Greenfield S, Sullivan L, Duker KA, *et al*: Development and testing of a new measure case mix for use in office practice. *Med Care* 33 (suppl): AS47-AS55, 1995).

1c-2. Extreme values and missing data

Extreme or unlikely values in the data were checked against the hard copies of the questionnaires to determine whether errors occurred during data entry. Confirmed cases in which the respondent reported an extreme value remain unchanged in the data.

When respondents did not provide an answer to a question, their answers were treated as missing. However, variables may contain missing values for other reasons. In an attempt to distinguish the various types of possible missing responses, meaningful missing codes are used in the SAS and Stata versions of the data. Generally, these codes distinguish data that are missing because a respondent chose not to respond to a question from those questions which the respondent was instructed to skip. The meaningful missing codes used in the data adhere to the classification in the table below.

Code	Reason for Missing Data
.D	Uncertain, can't say
.M	Missing
.S	Skip (information not available due to skip patterns, typically from the preceding question)
.X	Answer not applicable

1d. Sample Weights

Separate sampling weights were created for the questionnaire respondents and respondents with valid HbA1c scores. The appropriate weight must be used to provide unbiased estimates. The questionnaire weights should be used for analyses that do not include HbA1c scores, and will represent the full U.S. population of persons with diabetes born in 1948 or earlier. The HbA1c weights should be used for analyses that include HbA1c scores. The HbA1c weights will also represent the full U.S. population of persons from the smaller sample of persons with valid HbA1c results. Weights were created in a multi-stage process.

In the first stage, weights were created for the 2,381 eligible cases to adjust for the selection of the sample from the full sample of persons with diabetes in the HRS. This was done in two steps. The first step adjusted for the slightly different selection probabilities of proxy cases and cases with poor

cognitive status. The second step adjusted the weights so that the sum of weights of selected eligible cases by categories of diabetes severity and cardiovascular co-morbidities exactly matched the category sums of 2002 respondent weights for all persons reporting diabetes in 2002 and surviving to October 2003.

The second stage created questionnaire weights, again in two steps. In the first step, a propensity model was estimated by logistic regression, predicting the probability of returning a questionnaire among all selected eligible cases. Predictor variables included race, sex, age, education, measures of chronic disease and physical function, and type of diabetes treatment reported in 2002. The inverse of the fitted probabilities of participation formed the preliminary weight. These were then trimmed at the fifth and ninety-fifth percentiles. The second step adjusted these trimmed preliminary weights so that the sum of weights of questionnaire participants by categories of diabetes severity and cardiovascular co-morbidities exactly matched the category sums of 2002 respondent weights for all persons reporting diabetes in 2002 and surviving to October 2003.

The third stage created weights for the HbA1c scores using the same procedure as was used for questionnaire weights.

2. Obtaining the Data

2a. How to Apply

1. Visit the [HRS User Registration/File Download Web site](#). If you do not already have a username and password, you must register in order to enter the site.
2. Identify the desired data set (Diabetes 2003, in this case) under the *HRS Special Access File Request Information* heading.
3. Download and complete the *Data Use Agreement* in hardcopy format (See Appendix).
4. Send a signed copy of the agreement via surface mail to:

Health and Retirement Study
DUA Review Committee
426 Thompson Street, Room 3050 ISR
Ann Arbor, Michigan 48104-2321

2b. Approval Process

The Health and Retirement Study will review your request, and based on the information that you provided when you registered at the [HRS User Registration/File Download Web site](#), verify your identity and institutional affiliation. Once this authentication process has been completed to our satisfaction, we will authorize access to the Diabetes 2003 data. We will communicate with you at the email address that you provided when you registered at the download site.

2c. How to Download

Once you receive your approval notification, login to the [HRS User Registration/File Download Web site](#). In the HRS Special Access Files box on the right hand side of the page you will now see links to the Diabetes 2003 data set. Click on the appropriate link and follow instructions to download the Diabetes 2003 files.

2d. Publications Based on Sensitive Health Data

You must inform HRS of any papers, publications, or presentations based on this HRS sensitive health data set. Please send a copy of any such publications, with a bibliographical reference if appropriate, to this address:

Health and Retirement Study
Attn: Papers and Publications
The Institute for Social Research, Room 3050
426 Thompson Street
Ann Arbor, MI (USA) 48104

Papers and publications may also be submitted via electronic mail; send message and attachment(s) to hrequest@isr.umich.edu. Please include "Attn: Papers and Publications" on the subject line.

The following citation should be included in any research reports, papers, or publications based on these data:

In text: "The HRS 2003 Diabetes Study is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and was conducted by the University of Michigan."

In references: "Health and Retirement Study, 2003 Diabetes Study. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (2006)."

3. Setup

3a. Distribution Set

The HRS 2003 Diabetes Study (V1.0) files are packaged for distribution in a .zip file, diab2003.zip. In order to keep the contents secure, this file has been password-protected. For users working in a Microsoft Windows environment, a self-decompressing .exe file is distributed. UNIX, Linux, and Macintosh users will receive a standard .zip file. Users who require special file formats should contact the HRS Help Desk (hrequest@isr.umich.edu).

Extract the data file, the program statement file that matches your analysis environment, the data description (this file), and the codebook file.

3a-1. Windows Environment

Run the self-decompressing .zip file, diab2003.exe, from the command line or from Windows Explorer. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output from the self-extracting process will be the files listed below. If you have problems, please contact the [HRS help desk](#).

3a-2. UNIX Environment

Use the file decompression software installed on your system (i.e. `gunzip`) to expand `diab2003.zip`. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail to output the files listed below. If you have problems, please contact the [HRS help desk](#).

3b. Filename Suffixes

For all data sets, 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 (codebook)
*.pdf	PDF format

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:\diab2003\	Distribution file
c:\diab2003\docs\	diab2003A.txt diab2003S.txt diab2003dd.pdf (this file) diab2003qnaire.pdf
c:\diab2003\data\	diab2003A.da diab2003S.da
c:\diab2003\sas\	diab2003A.sas diab2003S.sas
c:\diab2003\spss\	diab2003A.sps diab2003S.sps
c:\diab2003\stata\	diab2003A.do diab2003A.dct diab2003S.do diab2003S.dct

Decompress the selected .zip files into the appropriate subdirectories. You will need approximately 5 MB of free space on your storage device to store the uncompressed files.

3d. Program Statements

The distribution package provides SPSS, SAS, or Stata program statements describing the 7 variables in the sample data set and the 321 variables in the questionnaire data set in this data set. These are designed to assist the user in creating system files from the ASCII data file. Files containing SPSS statements are named with a .sps extension, those containing SAS statements with a .sas extension, and those containing Stata statements with .do and .dct extensions. The ASCII file for the sample file contains 3,194 records, with a logical record length of 51. The ASCII file for the questionnaire file contains 1,901 records, with a logical record length of 707. System files will differ in size, but should contain the same number of records and variables.

3d-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:\diab2003\sas" and the data file is located in "c:\diab2003\data", you can run the file as is. A SAS system file (*.sas7bdat, assuming that you are using SAS V7 or later) will be saved to directory "c:\diab2003\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 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:\diab2003\spss" and the data file is located in "c:\diab2003\data", you can run the file as is. An SPSS system file (*.sav) will be saved to directory "c:\diab2003\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:\diab2003\stata" and the data file is located in "c:\diab2003\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.

4. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines for using the HRS 2003 Diabetes Study data. If you have questions or concerns that are not adequately covered here or on our restricted data Web site, or if you have any comments, please contact us. We will do our best to provide answers.

4a. HRS Internet Sites

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 at <http://hrsonline.isr.umich.edu/>.

Visit the HRS Restricted Data Web site at <http://hrsonline.isr.umich.edu> if you want to find out more about restricted data products.

4b. 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/intro/hrsquest.php>)

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

Appendix: Data Use Agreement

Health and Retirement Study Sensitive Health Data Use Agreement

Data Set Requested: _____

By signing this form and obtaining the requested data set from HRS, the User agrees:

1. To use the data set solely for statistical reporting and analysis.
2. Not to share these data with, or provide copies of these data to, any other person or organization. Note: Each research assistant/associate, graduate student or undergraduate student, or other individual working on the research project must sign and submit his/her own form.
3. To return or destroy the data set, and any derivative data files, upon request from HRS.
4. To make no attempt to link this data set with individually identifiable records from any source, or in any other way attempt to identify the persons in this or other HRS data sets.
5. That if the identity of any person or establishment in this data set is inadvertently discovered, then (a) no use will be made of this knowledge, (b) the Director of HRS will be advised of this incident immediately, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by HRS, and (d) no one else will be informed of the discovered identity.
6. To employ the following guidelines when producing tabulations for distribution:
 - Magnitude Data: Ensure that no cells/strata with $n < 5$ are produced.
 - Frequency Data: Apply a marginal threshold of $n \geq 5$ and cell threshold of $n \geq 5$ to all tabulations.
 - Protecting against complementary disclosure: Additional cells may be suppressed, i.e., complementary disclosure, to make sure the primary suppressions cannot be derived by subtraction from published marginal totals.
7. Aggregate statistical summaries of the data and analyses (frequency tabulations, magnitude tabulations, means, variances, regression coefficients, and correlation coefficients): These are approved under this agreement and may be freely published by the User, subject to the provisions above.
8. To cite HRS as the data source in any publications or research based upon these data, and to provide a copy of any publications to the HRS. Please refer to the data description for the correct citation that should be included in any research reports, papers, or publications based on these data.
9. To maintain a secure computing environment for storage and use of this data set and any data sets derived from it.
10. To hold harmless and indemnify HRS and the University of Michigan, its agents and employees, for any claims of breaches of confidentiality arising out of his/her research, defined as failure to abide by any section of this agreement or any accidental or intentional violation of privacy of any contributor to any HRS data resource.

UserID (from HRS File Download site):

Signature

Title

Name (printed or typed)

Employer/Institutional Affiliation

For Office Use Only	Date	Initials
Received		
Verified		
Approved		
Access Granted		