

README file for Marshall, McGarry, Skinner (2011) replication distribution set.

Version: 2.0

Version date: September 21, 2018

1 Overview

The files in this distribution set can be used to create approximate Out of Pocket (OOP) expenditure files from the Health and Retirement Study (HRS) exit interviews between 1998-2006. For unknown reasons, these files do not create an exact replication of the files used in Marshall, McGarry, Skinner (2011). If any of these issues can be resolved, this distribution set will be updated. There are several known issues with these files; see Section 4 for details.

2 Using the do files to create the datasets

2.1 Code folder structure

- build – these files combine datasets
- dependencies – these folder contains files that do background operations for the main programs
- impute – these files ‘impute’ the OOP expenditures for each category in each wave
- tables – create some of the tables in the paper
- updated – contains some corrected files

2.2 Order of Operations

The files must be run in this order to work:

1. globals.do – initialize the paths for where datasets are stored and where output is created. This is the only file that you should need to edit. The raw data is in the data/raw folder; you may however modify this file to point towards your own data folder.

2. `exit_expenditures.do` – create the cross wave set of reported expenditure values that are used for the imputation.
3. `yyyybuild.do` – merge the HRS raw files, create the helper file and append the `exit_expenditures` file.
4. `exityr.do` – do the imputations for each type of medical expenditure.
5. `exit_combine.do` – Combine the XOOpyr files and modifyv the variables to be consistent across waves.

3 Imputation notes

When a respondent doesn't know the amount of expenditure on a category, say doctor visits for example, a sequence of questions is then asked to gauge and upper and lower bounds of expenditure. In the 1998 and 2000 waves, the response to each of these question was recorded separately in its own variable. Starting with the 2002 wave, the HRS consolidated these responses into a two variables containing the final upper and lower bound. The `xYYcategory` do files in the `dependencies` folder create these 'consolidated' variables for each type of expense in 1998 and 2000.

The `exit_expenditures` file is appended, not merged. This file combines reported OOP expenditures from all waves on each type of expenditure. In the imputation function 'repute', this data is used to calculate the mean expenditure between the reported upper and lower bound. So, if the respondent reports that they spent more than ℓ , but less than v , using real expenditures from all waves, we then calculate $E[x \mid \ell < x < v]$.

4 Errors

- The helper file needs to be recoded for the DK variable and 98 is wrong
- imputation function only assigns means to the first 100 instances of upper and lower bounds (sorted low to high)
- The existing data uses a 2006 non-final version as well as the 2006 traker file

- In Section 2.3 of the 2000 imputation file, 00exit, for the second insurance variable "ins_2", there is a coding mistake when the conditional mean is applied. The mean is applied for cases when R2620 = 999998 or 999999, when the correct variable is R2636. As a result, 63 individuals are not assigned a value here.
- The imputation for private medigap expenditures uses the total reported expenditure on private medigap plans. That is, the 'all expenditures' value is the sum of all private medigap plans rather than individual plans.
- *Inconsistency*: The conditional mean assignments in 2002-2006 for the private insurance plans, after summing all of them together, is inconsistent.
- *Inconsistency*: The first mean assignment for long term care is inconsistent between 2002 - 2006.
- *Inconsistency*: The hospital/NH mean assignment in 2000 is both inconsistent and miscoded.