Economic models -

We follow the recommendation by the official report carefully written on the data that for harmonization purpose, general diagnosis of chronic medical conditions (e.g. heart diseases), rather than a more specific condition (e.g. congestive heart failure), may be used to maximize the number of studies that may be included. We thus use the diagnosed chronic conditions surveyed for all HRS-sister studies, which were hypertension, diabetes, cancers, chronic lung diseases, stroke, and arthritis or rheumatism.

We focus on diagnosis and treatment of chronic medical conditions, as they were available for all surveys and are considered the most suitable for comparative purposes.

Limitations:

An important aspect is who is allowed to provide such diagnosis. HRS specifically excludes diagnosis made by nurses/nurse practitioners, chiropractors, and dentists. However, both CHARLS and LASI allow diagnosis by nurses, practitioners of traditional medicine, and other health care professionals.

**HRS diseases:** RwDIABE, RwCANCRE, RwLUNGE, RwHEARTE, RwSTROKE, and RwARTHRE are indicator

variables denoting whether or not the Respondent reports a doctor has ever told her/him that s/he had the specified condition. The conditions are 1) diabetes or high blood sugar; 2) cancer or a malignant tumor of any kind except skin cancer; 3) chronic lung disease except asthma such as chronic bronchitis or emphysema; 4) heart problems, which include heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems; 5) stroke or transient ischemic attack (TIA); and 6) arthritis or rheumatism. The Rand HRS 1992\_2018v1 files we use for these doctor-diagnosed conditions, with the exception of cases that dispute a report from a prior wave, each of these variables is set to "yes" if the Respondent answered yes to the pertinent question in the current or any prior wave, and to "no" if the Respondent responded no at the current and all prior waves.

We have constructed three new variables that capture prevalence of chronic conditions. “chronic” refers to having at least one of the conditions cited above. “chronic\_sum” refers to the total number of conditions ever diagnosed. “chronic\_severe” is a dummy variable that measures comorbidity, where an individual is diagnosed with having three or more of those conditions.

To also evaluate further differences in onset and such, we also have the variable RADIAGDIAB, which indicates the age at which the respondent was first diagnosed with diabetes. RwRECCANCR indicates the most recent age at which the respondent was diagnosed with cancer. Respondents are asked the year in which they were most recently diagnosed with cancer, and these responses are converted to their age at diagnosis. Previous responses are carried forward if the respondent does not report a new cancer diagnosis. RwRECHRTATT indicates the most recent age at which the respondent had a heart attack. RAFRHRTATT indicates the age at which the respondent had their first heart attack. However, a first analysis showed that in the sample of HRS more than 75% was missing for age variables. So, in order to really investigate the onset age and time one should look at the samples longitudinally, and not really on these variables.

We will actually call it “disease-free life expectancy”, like the paper by Elisenda.