Read Me File for

„The Value of Health – Empirical Issues when Estimating the Monetary Value of a QALY using Well-being Data“

(Himmler, Stöckel, van Exel, and Brouwer; Health Economics, 2021)

Contact: Sebastian Himmler (himmler@eshpm.eur.nl) & Jannis Stöckel (stöckel@eshpm.eur.nl)

Link to Publication: https://onlinelibrary.wiley.com/doi/10.1002/hec.4279

Data Access and Replication File Instructions:

We are not able to make the underlying SOEP Data fully available to the public. However, access to the SOEP data releases is free of charge for researchers after signing a contract with the German Institute for Economic Research (DIW) guaranteeing compliance with confidentiality requirements.

Access can be requested via this link (link created March 2021):
https://www.diw.de/en/diw_02.c.222829.en/access_and_ordering.html

The results presented in the paper are based on the **SOEP-Core Data Distribution 1984-2018 (soep.v35)** (link created March 2021). **We highly recommend to use this specific release** when replicating our results. This prevents any problems arising from future SOEP data releases retro-actively change the file types or coding conventions of certain variables used in our analysis or intermediate cleaning exercises going from raw data to analysis data.

As we cannot provide our analysis dataset the replication package also includes the dofiles to replicate the intermediate steps creating our analysis dataset from the raw SOEP data. To ensure that the files in the replication package and the relative file-paths within dofiles can work properly you will need to unpack the entire content of the replication zip-folder. The resulting folder should include:

- A set of 22 (sub) dofiles enumerated in groups from 00 to 06:

Dofile	Main Purpose
00	Master dofile running all subsequent dofiles covering data-formatting and analyses.
01	Extracting relevant variables from various SOEP source files.
02	Merging the datasets together into a long-form raw dataset of waves 1984-2018.
03	Cleaning for missing information on the variables used in the analysis.
04	Preparing the data for analysis; creating relevant datasets and defining variables.
05	Creating various descriptive figures presented throughout the paper.
06	Running analysis dofiles and creating tables and figures presenting the results.

- A set of folders titled:
  - **Data_source** → Location for the source-files from SOEP data release v35 (see below).
  - **Data_merge** → Location for intermediate files created by dofiles **01**
  - **Data_panel** → Location for intermediate/final panel datasets created by dofile **02**
  - **Output** → Location for log-files and two subfolders with Graphs/Tables in .tex format.
- A folder containing all log-files, tables and figures as in the paper as obtained before uploading this replication package: **Output_HE_18032021**
- An excel file titled “CIV_overview” which contains the estimated CIV_{QALY} estimates for all specifications depicted in Figure 3.
- A pdf document titled “CPI_DESTATIS_1948-2019_annual_Feb2020”, containing an official release by the German Statistical Office. First column on page 3 contains the indices used to calculate income in 2015 prices.

- An excel file titled "cpi_germanyDESTATIS_Feb2020" containing the consumer price indices as listed in the pdf file mentioned above and the same file in stata .dta format ("cpi_Germany_Feb2020.dta")

To ensure that the master-dofile (00_SOEP_QALY_master.do) and subsequent dofiles can run smoothly using the encoded file-paths **you will need to place the SOEP source files in different folders:**

SOEP v35 Filename	Content	Designated folder
pgen.dta	Individual-level generated variables (e.g. marital status), all years	Data_source
pl.dta	Individual-level survey responses (e.g. life satisfaction, work history), all years	Data_source
hgen.dta	Household-level generated variables (e.g. net-income, region of residence), all years	Data_source
bhkind.dta bikind.dta	Datasets on children for years 2017 (bh) and 2018 (bi)	Data_source/Children/unharmonised
skind.dta to bgkind.dta	Datasets on children for years 2002 (s) to 2016 (bg)	Data_source/Children/source
sp.dta to bip.dta	Datasets on changing individual module health survey for 2002 (s) to 2018 (bi)	Data_source/sf12health/source
health.dta	Re-occurring health survey module, all years	Data_source

The zipped folder already has the structure necessary to run our analyses so be aware that **moving dofiles will require you to adjust the pathways to reflect a change in your working directory.**

Running the dofile "00_SOEP_QALY_master" will create figures and tables presented in the paper in the following order in square brackets the specific dofile is enumerated that produces the set of results. Please note that color-schemes might be different to the published paper[1]:

Descriptive Figures (in order of creation)

[04_01_SOEP_QALY_prep_instrument.do]

Figure A3.3: Impact of industry on labour income

Figure A3.4: Impact of occupation on labour income

[05_SOEP_descriptivefigures.do]

Figure 2: SF12 index values using UK and Dutch tariffs

Figure A3.1: Average Monthly Net Labour Income by Industry

Figure A3.2: Average Monthly Net Labour Income by Occupation

Figure A4.1: Health change sample overview

Figure A4.2: Health change sample – mental/physical component score differences

[1] Executing individuals analysis dofiles (**06_01-06_03**) requires that the globals in **06_00**_ are generated beforehand)

Results Graphs and Tables (in order of creation)

[06_00_SOEP_QALY_analysis_manuscript.do]

Table 1: Descriptive Statistics

Table A1.2: Descriptives for sample no industry or occupation information

Table A4.1: Descriptives for health state dependence sample

Table 2: Baseline Results

Table A3.1: First Stage results of IV regression

FigureA3.5: QQ-plot of observed and predicted labour income

FigureA3.6: Impact of industry on life satisfaction

FigureA3.7: Impact of occupation on life satisfaction

Table 3: Results by region and time period

Table A2.1: Additional subgroup results

Figure 1: Relationship between life satisfaction and income across income splines

Table 4: Income specifications

Table 5: Choice of SF6D tariffs

Table 6: Health state dependence

Table 4.2: Health state dependence – excluding large income losses

Table 4.3: Health state dependence – excluding large income losses & unemployed/retired

Table 4.4: Health state dependence – excluding large income losses & unemployed/retired & only severe health changes

Table 7: Robustness Checks

Table A2.2: Results for unweighted and separate SF-6D levels

Figure 3: Results Overview Graph

Further tables

Table A1.1: Dataset conditioning → Numbers extracted after respective steps in the data creation process. Respective steps and numbers are marked red in the output when running the do-files (**03** and **04_02**).