**What data exist about causes of death for different countries by age group over time?**

After some digging these were the most promising data sources that I found: Institute for Health Metrics and Evaluation’s (IHME) Global Burden of Disease (GBD) project, the CDC’s Wide-ranging Online Data for Epidemiologic Research (WONDER), and the WHO’s Global Health Observatory (GHO).

The GBD project has very extensive data on all cause deaths, DALYs, probability of death, and other measurements of mortality by age from 1990 to 2017 for all of the countries we are interested in. They have a results query tool that can be found here: (<http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2017-public/aacc56b34b3cfbb6b5bfe92dad365783>) Below are a more detailed description of what data are included in the project.

* All GBD causes, risks, etiologies, impairments, injuries by nature, and sequela aggregates
* Measures: deaths, years of life lost (YLLs), years lived with disability (YLDs), disability-adjusted life years (DALYs), prevalence, incidence, life expectancy, probability of death, healthy life expectancy (HALE), maternal mortality ratio (MMR), and summary exposure value (SEV)
* Metrics (units): number, rate, percent, years, probability of death
* Years: 1990-2017; annual results for all measures
* All GBD age groups
* Sexes: males, females, both sexes
* Locations: GBD super regions, regions, countries, select subnational units, and custom regions (WHO regions, World Bank Income Levels, and more)

IHME has more data available for years going back before 1990 but it is less clean and standardized. This page (<http://ghdx.healthdata.org/record/ihme-data/gbd-2017-all-cause-mortality-and-life-expectancy-1950-2017>) has all-cause mortality and life expectancy until 1950 but is not disaggregated by age before 1990.

The WHO’s GHO, collects information about Morality, DALYs, YLL, and YLD but unfortunately goes only from 2000-2016. The information on Mortality is quite detailed and includes causes of death, adult and child, life tables, maternal mortality, and others, broken down by country and age group.

CDC’s (WONDER) contains quite detailed information on US mortality and causes of mortality. Specifically, two datasets seem most interesting: Detailed Mortality and Compressed Mortality.

The detailed mortality dataset goes from 1999-2017 and contains mortality and population counts for all U.S. counties. The data are derived from death certificates that identify a single underlying cause of death and demographic data. We can obtain the data by age group (as specific as one year), race, gender, year, and cause of death (4 Digit ICD).

The compressed mortality dataset goes from 1968-2016 and we can obtain counts and rates of death by underlying cause, state, county, age, race, sex, and year. 1968-78 use ICD 8 codes, 79-98 ICD 9, and 1999-2016 ICD10.

**What data exist about unemployment rates and whatever else we can think of across countries over time?**

The most promising data sources I have been able to find are the OECD, UN University World Institute for Development Economics Research (UNUWIDER), and Tradingeconomics. Additionally, the national statistical bureaus of the countries we are interested in can also be an option. The Luxembourg income/wealth study appears to be very individual/household focused. Given that most of the data we have previously been working with is at the national level, I’m not sure how we could integrate/aggregate it to the national level. Further complicating this process is that the Luxembourg study is quite protective of their data and the only way we are allowed access is to send them program files such as do files and allow them to run the program, after which they will send us the results back.

The UNUWIDER dataset focuses on inequality. Not only does it contain the GINI coefficients for the countries of interest, it also contains the share of income held by each quintile and decile for the countries we are interested in as far back as 1867, though the data become quite sparse pre WWII.

The OECD has the most comprehensive data not related to inequality. It contains quite extensive data on employment rate, as well as employment rate broken down by certain groups such as education, age, and gender. It also contains information on inequality such as the size of transfers that is absent from the UNUWIDER data set. Most if not all of the indicators that we might consider are probably tracked by the OECD, though data before the 1980s seems to be increasingly incomplete compared to more recent periods.

Tradingeconomics is a website that maintains a database of quite a few macroeconomic indicators such as labour participation, capital flows, government debt to GDP, consumer credit, and much more by country over time. Similar to the OECD, most of the indicators we are interested in are probably tracked here. The completeness of the data varies quite a bit by country and much of the data are pulled from the national statistical bureaus of the respective countries. The final option would be to pull the data from the individual statistical bureaus of each of the countries we decide to include in the study.