

The data appendix is organized into three parts. We provide the information needed to reproduce the results presented in the paper.

### **I. Data Sources**

All data used in this paper are publicly available online from their original sources, with the exception of county code identifiers on death certificates, for which application is necessary. Links to all data sources, and to the application information page for obtaining county code information for death certificates, are provided below.

### **II. Dataset Creation**

Data pulled from the web are cleaned and organized into datasets that can be merged and used to create the figures and tables presented. Information on the Stata dofiles used to create these datasets is provided below. Dofiles are provided in folders by type of data created. Folder labels are self-explanatory. Small crosswalk datasets are provided with the dofiles.

### **III. Figures and Tables**

Stata dofiles used to create the paper's figures and tables are noted below, together with details on variable creation, sample sizes and age ranges. Dofiles used to create tables appear in the Tables folder and dofiles used to create figures appear in the Figures folder.

<b>I. DATA SOURCES</b>	<b>page</b>
A. US Mortality Data	2
B. US Population Estimates	2
C. International Mortality Data	2
D. US Economic Data	2-3
E. US Morbidity Data	3
<b>II. DATASET CREATION</b>	
A. US Mortality Data	4-6
B. US Population Estimates	7
C. International Mortality Data	8-10
D. US Economic Data	10-11
E. US Morbidity Data	11

### III. FIGURES AND TABLES

12-16

#### I. DATA SOURCES

##### A. US Mortality Data

Age-specific all-cause and cause-specific mortality data from the CDC-Wonder Underlying Cause of Death Files (1999-2015) and National Vital Statistics System individual death records (1989-2015):

<https://wonder.cdc.gov/ucd-icd10.html>

[https://www.cdc.gov/nchs/data\\_access/vitalstatsonline.htm#Mortality\\_Multiple](https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple),

US Mortality Data disaggregated by education, and by ethnicity (prior to 1999)

Individual death records from the National Vital Statistics System:

[https://www.cdc.gov/nchs/data\\_access/vitalstatsonline.htm#Mortality\\_Multiple](https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple)

US Couma mortality estimates

We obtained permission to use county identifiers on death records from 1989-2015, and created couma-level mortality statistics by cause. Information on the application process is available here:

[https://www.cdc.gov/nchs/nvss/dvs\\_data\\_release.htm](https://www.cdc.gov/nchs/nvss/dvs_data_release.htm)

county-couma Stata data crosswalks are provided with the dofiles.

##### B. US Population Estimates for use with individual death records

Population estimates from the CDC Bridged-Race Population series

<https://wonder.cdc.gov/Bridged-Race-v2015.HTML>

Population estimates by level of education were obtained from annual March Current Population Survey files by age, sex, race, educational attainment, available from the CEPR:

<http://ceprdata.org/cps-uniform-data-extracts/march-cps-supplement/march-cps-data/>

##### C. International Mortality Data

All-cause mortality data by year and country, sex and age, are available through the Human Mortality Database [www.mortality.org](http://www.mortality.org)

Annual data on all-cause and by-cause mortality by 5-year age group and sex are available by combining mortality data pulled from the World Health Organization Mortality database:

[http://www.who.int/healthinfo/mortality\\_data/en/](http://www.who.int/healthinfo/mortality_data/en/) and population data from the UN:

<https://esa.un.org/unpd/wpp/Download/Standard/Population/>

##### D. US Economic Data

Data on real wages by age, education, race, and ethnicity from the Current Population Survey Outgoing Rotation Groups, available from the CEPR:

<http://ceprdata.org/cps-uniform-data-extracts/cps-outgoing-rotation-group/cps-org-data/>

Data on labor force participation, educational attainment, marriage, household income from the March Current Population Surveys, available from the CEPR:

<http://ceprdata.org/cps-uniform-data-extracts/march-cps-supplement/march-cps-data/>

#### **E. US Morbidity Data**

Annual data from the National Health Interview Survey 1997-2015

<https://www.cdc.gov/nchs/nhis/>

for chronic joint pain, sciatic pain, serious mental distress, difficulty socializing.

Annual data from the Behavioral Risk Factor Surveillance System 1993-2015

<https://www.cdc.gov/brfss/>

for body-mass index, self-assessed health status, and an indicator for being at risk for heavy drinking.

## II. DATASET CREATION

We have included all crosswalk datasets that are called in the dofiles that follow. They are stored with the dofiles to which they are relevant. Dofiles are provided in folders by type of data created. Folder labels are self-explanatory. The initial infiles for each set of dofiles need to be pulled from the web using the links provided. For all data work, you will need to change the directory information in the dofiles to point to the folders where your data are stored.

### A. US Mortality Data:

Data were cleaned and coded by cause of death.

Dataset creation dofiles must be run in the following order:

- ProcessMortality03012017.do
- ProcessMortality03012017\_2015.do
- Read\_mort\_8915.do
- Make\_WNHrates\_sex\_5yr\_state\_ed\_9215.do

NOTE: The final dofile in this list is provided as a template – this one creates mortality rates by cause for white non-Hispanics by sex and education group (1992-2015) by five-year age group. Prior to running this dofile, data from annual March CPS files must be processed (see below), to provide estimates of the population at risk.

#### 1. Reading raw data

Exact file names for raw data will depend on whether they were pulled from the NVSS website, or were sent on a CD. (Ours is the latter case, because we have access to county identifiers, but we have checked that data pulled from vitalstatsonline are read successfully using the code we provide.)

**infiles: raw data annual files from NVSS.**

[https://www.cdc.gov/nchs/data\\_access/vitalstatsonline.htm#Mortality\\_Multiple](https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple),

**dofiles: for 1989-2014: ProcessMortality03012017.do;  
for 2015: ProcessMortality03012017\_2015.do**

**outfiles: mort1989.dta – mort2015.dta**

#### 2. Selecting observations

Remove foreign residents, keep deaths for ages of interest, merge with a file to attach state fips codes to death records (crosswalk data provided), create an education identifier (edclass) =1 high school or less, 2=some college, 3=BA or more education. Create race/ethnicity indicators.

**infiles: mort1989.dta – mort2015.dta, state\_lab\_fips.dta (provided)**

**dofiles: read\_mort\_8915.do**

**outfiles: mort\_8915.dta**

#### 3. Create mortality rates by cause, race, education, age

These files will need to be modified and run depending on whether you want mortality rates by single years of age or, alternatively, by age groups; by race; and whether you are planning to disaggregate by level of education. Having made these choices, you must use the appropriate population at risk to calculate rates. We use CDC bridged-race population data when not disaggregating by education, and data from the March CPS files when disaggregating by education.

Causes of death are identified using ICD codes.

ICD codes:

Poisoning: *accidental and intent undetermined drug overdose and alcohol poisoning*

ICD9 codes: 850-860, 980; ICD10 codes: X40-45, Y10-15, Y45, Y47, Y49

Suicide:

ICD9 codes: 950-959; ICD10 codes: X60-84, Y87.0

Alcoholic Liver Disease and Cirrhosis:

ICD9 codes: 571; ICD10 codes: K70, K73, K74

Heart Disease:

ICD9 codes: 390-429; ICD10 codes: I00-I09, I11, I13, I20-I51

Cancer: *malignant neoplasms*

ICD9 codes: 140-208; ICD10 codes: C00-C97

Educational attainment appears uniformly on the death certificates after the revision of 1989, but it took some time for all states to report education on death records. For this reason, we use education-related data only beginning in 1992, and remove from **all** education-related analyses four states that had not made a regular practice of recording education by 1992 (these are GA, OK, RI, SD). Mortality records with missing values for education in the remainder of the states (e.g., 4 percent of cases for white non-Hispanics ages 25-64) are assigned to an education category based on the fraction in each education group by year, race, sex, age group, and cause of death.

The following dofiles provide a template. The first creates mortality rates for five-year age groups, by education class, by sex (0=all, 1=male, 2=female), for white non-Hispanics. The second creates the same for black non-Hispanics.

**infile:** mort\_8915.dta and stcross.dta (state identifier crosswalk from CPS to NVSS)

**dofile:** make\_WNHrates\_sex\_5yr\_state\_ed\_9215.do

make\_BNHrates\_sex\_5yr\_state\_ed\_9215.do

**outfile:** WNHrates\_sex\_5yr\_state\_ed\_9215.dta

BNHrates\_sex\_5yr\_state\_ed\_9215.dta

## **CDC WONDER DATA**

We also use many data series drawn from CDC Wonder

<https://wonder.cdc.gov/ucd-icd10.html>

Selection on cause of death, years, age categories, sex, race, and level of urbanization are made online. Data were exported as txt files. We used Stat/Transfer to read the exported files into Stata, and then processed them using dofiles. As an example, the dofiles we used to read and combine all-cause mortality rates (Trate) for white non-Hispanics, black non-Hispanics, and Hispanics by 5-year age group and sex are provided in the dofile folder “CDC\_Wonder”:

**infile:**            **WNH\_Trate\_5yr\_sex\_9915.dta, BNH\_Trate\_5yr\_sex\_9915.dta, and  
                          Hisp\_Trate\_5yr\_sex\_9915.dta**  
**dofiles:**           **read\_WNH\_BNH\_Trate\_5yr\_sex.do and prep\_vars\_1.do**  
**outfile:**           **WNH\_BNH\_Trate\_5yr\_sex.dta**

## **COUMA MORTALITY ESTIMATES**

Geographic identifiers are death certificates are US counties. We combined counties into units that can be matched with Public Use Microdata Areas (PUMAs) used by the US Census Bureau.

<https://www.census.gov/geo/reference/puma.html>

We call these new units coumas. They are at least the size of a county, and at least the size of a PUMA (and so contain at least 100,000 people). Because PUMAs change after a decennial census, we match data from 2000-2011 into one set of coumas, and 2012-14 into a second set. Some counties (e.g. Philadelphia) contain many PUMAs. Some counties (e.g. in southwestern Montana) are sparsely populated, and need to be combined to match one PUMA. In some cases, to obtain clean boundaries, we need to combine multiple PUMAs and multiple counties into one couma. All observations from a county belong to the same couma. The crosswalks between counties and coumas are provided in couma00\_county.dta (for 2000-2011), and couma10\_county.dta (for 2012-14). These crosswalk datasets are included in the USMortalityData folder. Note that identification is by state and couma.

To calculate mortality rates for 45-54 year old white non-Hispanics by year-state-couma:

**Infiles:**            **WNH\_4554\_pop\_cty\_0014.dta** (bridged-race population for white non-Hispanics by  
                          county, 2000-2014)  
                          **Mort\_8915.dta** (created as described above)  
                          **Couma00\_county.dta** (crosswalk for 2000-2011)  
                          **Couma10\_county.dta** (crosswalk for 2012-2014)  
**Dofile:**            **make\_couma00\_couma10\_WNHrates\_4554.do**  
**Outfile:**           **couma00\_couma10\_WNHrates\_4554.dta**

## **B. US Population Estimates**

Population estimates from the CDC Bridged-Race Population series

<https://wonder.cdc.gov/Bridged-Race-v2015.HTML>

Estimates for population by education group, labor force participation, educational attainment, marriage, and household income were obtained from annual March Current Population Survey, available from the CEPR:

<http://ceprdata.org/cps-uniform-data-extracts/march-cps-supplement/march-cps-data/>

### **March CPS Data:**

**infile:** cepr\_march\_1980.dta – cepr\_march\_2016.dta

**dofile:** -make\_cps\_march\_incdata\_8089.do

-make\_cps\_march\_incdata\_9099.do

-make\_cps\_march\_incdata\_0016.do

-read\_cps\_march\_inc\_8016.do

**outfile:** cps\_march\_inc\_8016.dta

## C. International Mortality Data

### 1. HMD

Annual data on all-cause mortality by **single year of age by year** (1x1) by country are available through the Human Mortality Database [www.mortality.org](http://www.mortality.org)

After you register with the HMD, you can pull down death rates as txt files and import them into Stata.

We pulled deaths and populations separately, because we need these components to make age-adjusted rates for Figure 1.3.

Infiles: 1x1 mortality and population files from HMD

Dofile: make\_Trate\_1x1\_update.do.

Outfile: Trate\_1x1\_update.dta

### 2. WHO/UN

Annual data on all-cause and by-cause mortality by **5-year age group** and sex are available by combining mortality data pulled from the World Health Organization Mortality database:

[http://www.who.int/healthinfo/statistics/mortality\\_rawdata/en/](http://www.who.int/healthinfo/statistics/mortality_rawdata/en/)

and population data from the UN:

<https://esa.un.org/unpd/wpp/Download/Standard/Population/>

[On April 29, 2017] For the WHO mortality data, you need to pull 3 files:

Mortality ICD9; Mortality ICD10 (part 1/2), and Mortality ICD10 (part 2/2).

For the UN population data, you need to pull 3 files:

WPP2015\_Pop\_F15\_2\_ANNUAL\_POPULATION\_BY\_AGE\_MALE.xls;

WPP2015\_Pop\_F15\_3\_ANNUAL\_POPULATION\_BY\_AGE\_FEMALE.xls; and

WPP2015\_Pop\_F15\_1\_ANNUAL\_POPULATION\_BY\_AGE\_BOTH\_SEXES.xls

These are labeled on the website. For example:

Annual Population by Five-Year Age Groups - Both Sexes. De facto population as of 1 July of the year indicated classified by five-year age groups (0-4, 5-9, 10-14, ..., 95-99, 100+). Data are presented in thousands.

US mortality for all races and ethnicities in the US combined are included in the data files from the WHO. If you want to merge the international data by year, age group, sex and cause with US data where US mortality counts are given by race and ethnicity (US white non-Hispanics, US black non-Hispanics and US Hispanics), you will also need US population counts by race, age, year, sex. We drew those from

<https://wonder.cdc.gov/bridged-race-population.html> and read them into Stata to create pop\_race\_5yr\_1990\_2014.dta, which is called below. Note that the international mortality data are *not* broken down by race or ethnicity.

To process the data for use in creating tables and figures, the following dofiles must be run.

- **Process\_who\_icd10\_update.do**, which calls **process\_deaths.do**
  - Infiles:



- Morticd10\_part1.dta
    - Morticd10\_part2.dta
    - Country\_codes.dta (these are provided)
  - Outfiles:
    - alldeathsicd10\_update.dta      for all-cause mortality
    - causeXicd10\_update.dta      for each causeX
- **Process\_who\_icd9\_AC.do**
  - Infiles:
    - morticd9.dta
    - country\_codes.dta
  - Outfiles:
    - Alldeathsicd9.dta      for all-cause mortality
    - causeXicd9.dta      for each causeX

Note: East and West Germany (country codes 4090, 4100) each appear from 1980-1990, and Germany (country code 4085) appears from 1990-present. This dofile is written to drop East and West Germany, but could be changed to join them prior to 1990 rather than delete them.

- **Combineicd9\_10\_update.do**
  - Infiles:
    - Alldeathsicd9.dta
    - Alldeathsicd10\_update.dta
  - Outfiles:
    - Alldeathsicd9\_10\_update.dta      for all-cause mortality
    - causeXicd9\_10\_update.dta      for each causeX
- **UNPop\_male\_female\_age.do**
  - Infiles:
    - WPP2015\_Pop\_F15\_2\_ANNUAL\_POPULATION\_BY\_AGE\_MALE.xls
    - WPP2015\_Pop\_F15\_3\_ANNUAL\_POPULATION\_BY\_AGE\_FEMALE.xls
    - WPP2015\_Pop\_F15\_1\_ANNUAL\_POPULATION\_BY\_AGE\_BOTH\_SEXES.xls
  - Outfiles:
    - Age\_cats\_males\_females\_both.dta

- **Process\_popsicd9\_10\_update.do**, which calls **create\_rates.do**
  - Infiles:
    - Alldeathsicd9\_10\_update.dta
    - causeXicd9\_10\_update.dta
    - Age\_cats\_male\_female\_both.dta
    - UN\_WHO\_crosswalk.dta (these are provided)
  - Outfiles:
    - Alldeathsicd9\_10\_pop.dta
    - CauseXicd9\_10\_pop.dta      for each causeX

Note: to run process\_popsicd9\_10\_update.do, you must first run UNpop\_male\_female\_age.do (above)

- **Make\_rates\_icd9\_10\_update.do**
  - Infiles:

- Alldeathsicd9\_10\_pop.dta
- CauseXicd9\_10\_pop.dta for each causeX
- Outfiles:
  - Alldeathsicd9\_10\_pop.dta\_iso.dta
  - CauseXicd9\_10\_pop.dta\_iso.dta

Note: this dofile will scroll through all data files in a given folder (alldDeaths, poison, suicide, etc) and assign country isocodes. If you want to use this, make sure the data are all stored in the same folder.

- **Make\_mortcounts\_5yr\_race.do \*US Data\***

- Infiles:
  - Mort\_8915.dta (see US Mortality Data, above, for creation)
- Outfiles:
  - Mortcounts\_5yr\_race.dta

Note: cause of death assignment is slightly different when we compare mortality rates internationally. There are slight differences in the ICD9 codes assigned in cause of death relative to what we use when we look only at US data. This dofile, **make\_mortcounts\_5yr\_race.do**, assigns cause of death in the US to match the categories in the international data. If you are interested only in the period 1999-2015, you can pull mortality by cause from CDC Wonder: the ICD10 codes used are the same in the international data and in the US death records.

- **Make\_rates\_5yr\_racenew.do \*US Data\***

- Infiles:
  - Mortcounts\_5yr\_race.dta
  - Pop\_race\_5yr\_1990\_2014.dta (from CDC bridged-race population data)
- Outfiles:
  - Rates\_5yr\_race.dta

- **Make\_intl\_rates\_5yr\_update.do**

- Infiles:
  - Alldeathsicd9\_10\_pop.dta\_iso.dta
  - CauseXicd9\_10\_pop.dta\_iso.dta
  - Rates\_5yr\_race.dta
- Outfiles:
  - Intl\_rates\_5yr.dta

#### D. US Economic Data

Data on real wages by age, education, race, and ethnicity from the Current Population Survey Outgoing Rotation Groups are available from the CEPR:

<http://ceprdata.org/cps-uniform-data-extracts/cps-outgoing-rotation-group/cps-org-data/>

**CPS-ORG Data:**

**infile:** cepr\_org\_1979.dta – cepr\_org\_2015.dta  
**dofile:** make\_ORG\_rw\_sex\_ed\_7915.do

**outfiles:**           **ORG\_rw\_sex\_ed\_7915.dta**

Notes on CPS-ORG Data:

Annual real wage data from the Current Population Series Outgoing Rotation Group obtained from The Center for Economic and Policy Research, Washington DC: CEPR\_data.org

Real (\$2015) wages (variable **rw**) have been converted to real using CPI-U-RS. For details on wage construction of **rw** (usual hourly earnings excluding overtime/bonuses for hourly workers; usual hourly earnings including overtime/bonuses for non-hourly workers; assignment of mean wages above the top-code for top coded non-hourly workers), see

<http://ceprdata.org/cps-uniform-data-extracts/cps-outgoing-rotation-group/cps-org-faq/> .

## **E. US Morbidity Data**

Annual data from the National Health Interview Survey 1997-2015

<https://www.cdc.gov/nchs/nhis/>

for chronic joint pain, sciatic pain, serious mental distress, difficulty socializing

Annual data from the Behavioral Risk Factor Surveillance System 1993-2015

<https://www.cdc.gov/brfss/>

for body mass index, self-assessed health status, and an indicator for being at risk for heavy drinking.

### **NHIS Data:**

**infiles:**           **samadult1997.dta – samadult2015.dta and personsx1997.dta – personsx2015.dta**

**dofiles:**           **-NHIS\_sample\_adult.do, which calls:**

**create\_race.do, processvars.do, label\_var.do, converttonull.do and final\_process.do**

**-make\_WNHsa\_1997\_2015.do**

**-make\_WNHsa\_1997\_2015\_addlvars.do**

**outfiles:**           **wnhsa\_1997\_2015\_addlvars.dta**

Notes on NHIS data

Sample restricted to white non-Hispanic sample adults. Sample adult weights [wtfa\_sa] applied for all analyses. Responses missing either because the respondent does not know an answer, or refuses to say, are excluded from variable construction. Years: 1997-2015.

### **BRFSS Data:**

Annual data from the Behavioral Risk Factor Surveillance System

<https://www.cdc.gov/brfss/>

for body mass index, self-assessed health status, and an indicator for being at risk for heavy drinking.

**infiles**           **cdbrfs93.dta – cdbrfs10.dta and llcp2011.dta-llcp2015.dta**

**dofiles:**           **combineBRFS.do**

**outfiles:**       **combinebrfs.dta**

Notes on BRFSS data

Sample restricted to white non-Hispanic adults. Sample weight [A\_FINAL] applied for all analyses. Responses missing either because the respondent does not know an answer, or refuses to say, are excluded from variable construction. Years: 1993-2015.

### III. FIGURES AND TABLES

#### Table 1.

**Infiles:** wnh\_bnh\_5yr\_sex.dta (CDC Wonder 5-year mortality rates by race for 50-54 year olds)  
WNHrates\_sex\_5yr\_state\_ed\_9215.dta

**Dofiles:** **Table1.do**

#### Table 2.

**Infiles:** wnh\_bnh\_5yr\_sex.dta  
intl\_rates\_5yr.dta

**Dofiles:** **Table2.do**

Note on Table 2: program calls updated data on US white non-Hispanics, US black non-Hispanics and US Hispanics (to include mortality rates for 2015) downloaded from CDC Wonder, and merges them with the international data.

#### Figure 1.1

**[Program: graph\_Fig1\_1.do]**

This file takes all-cause mortality rates by year, race and 5-year age groups, pulled from the CDC Wonder website <https://wonder.cdc.gov/ucd-icd10.html> (discussed above) and merges them by year and 5-year age groups with mortality rates for white non-Hispanics with a high school degree or less. It then plots all-cause mortality for men and women (combined) ages 50-54, for all white non-Hispanics (WNH), all black non-Hispanics (BNH), all Hispanics, and WNH with a high school degree or less education.

#### Figure 1.2

**[Program: graph\_Fig1\_2.do]**

This dofile calls data created using make\_WNHrates\_sex\_5yr\_state\_ed\_9215.do (above) and a parallel dofile for BNH (make\_BNHrates\_sex\_5yr\_state\_ed\_9215.do, above), and graphs WNH and BNH all-cause mortality rates, by 5-year age group, for WNH and BNH with a high school degree or less education.

#### Figures 1.3A and 1.3B

**[Programs: graph\_Fig1\_3A.do and graph\_Fig1\_3B.do]**

US mortality estimates are created using data mort\_8915.dta (see notes above) and two population files from the CDC bridged-race population series. Figure 1.3A uses annual data on total white non-Hispanic population aggregated over ages 45-54, for each year 1990-2015. Figure 1.3B uses annual data on white non-Hispanic population by single year of age to create mortality rates by single year of age, which are then weighted by the share of each age in the population 45-54 in year 2010, to create age-adjusted mortality rates for age group 45-54.

Comparison country estimates are created using dataset Trate\_1x1\_update.dta (see program notes above).

#### Figure 1.4

**Dofile: graph\_Fig1\_4.do**

**Figure 1.5**

**Dofile:** graph\_Fig1\_5.do

**Figure 1.6 and Figure 1.9**

**Infiles:** couma00\_couma10\_WNHrates\_4554.dta

These data can be used to map couma-level mortality rates. We used ArcGIS to map these. Counties, which ArcGIS readily processes, need to be combined in ArcGIS (using the couma00\_county.dta and couma10\_county.dta crosswalks provided) into coumas for mapping.

**Figure 1.7**

**Infiles:** mort\_8915.dta

**Dofiles:** make\_WNHrates\_SY\_MF\_2564\_ed12\_again\_9215.do  
graph\_Fig1\_7.do

Note: Panels of Figure 1.7 are drawn for white non-Hispanics ages 25-64 in birth cohorts 1935-1980, with less than a BA, drawn from mortality records 1992-2015.

**Figure 1.8**

**Dofile:** graph\_Fig1\_8.do

**Figure 1.10**

**Dofile:** graph\_Fig1\_10.do

**Figure 1.11**

**Dofile:** graph\_Fig1\_11.do

**Figure 1.12:**

**Dofile:** graph\_fig1\_12.do

BRFSS: Indicator for excellent/very good health =1 if respondent, in answer to “Would you say that in general your health is ...” reports either ‘excellent’ or ‘very good’ on a five-point scale that also includes ‘good’ ‘fair’ and ‘poor’ as options. Figure 1.12 presents results from local polynomial regressions in which data are pooled (1993-1996) (1999-2002) (2005-2008) and (2012-2015). Sample restricted to white non-Hispanics ages 30-74.

Sample sizes for Figure 1.12:

Years:	1993-96	1999-02	2005-08	2012-15
Education:				
High school or less	119,326	185,230	334,058	340,341
Some college	71,559	126,345	256,701	299,738
BA or more	77,249	156,647	357,812	444,931

**Figure 1.13:**

**Dofile:** graph\_fig1\_13.do

NHIS: Indicator for sciatic pain is defined as =1 if respondent reports lower back pain in the past three months (painlb) that spread down either leg to areas below the knee (painleg); =0 otherwise.

Sample sizes for Figure 1.13:

Birth cohort interval:	Less than BA	BA or more
1940-42	10,286	3,622
1950-52	13,196	6,816
1960-62	15,002	6,840
1970-72	11,185	6,898
1980-82	4,893	3,412

### Figure 2.1

Dofile: graph\_fig2\_1.do

### Figure 2.2

Dofiles: graph\_fig2\_2A.do  
graph\_fig2\_2B.do

### Figure 2.3

Dofile: graph\_fig2\_3.do

### Figure 2.4

Data were drawn directly from the US Census Bureau, Historical Income Tables: Households <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html>

### Figure 3.1:

Dofile: graph\_fig3\_1.do

Sample restricted to white non-Hispanic men ages 25-59. Sample weights [wgt] applied for all analyses. Observations in Figure 3.1 are mean percent not in the labor force by (year of birth interval x individual year of age) cells. (SOURCE: March CPS 1980-2016)

Number of observations for percent not in the labor force, used in Figure 3.1:

Birth cohort interval:	Percent not in the labor force Observations		Ages at which cohort is observed
	Less than BA	BA or more	
1930-31	9,918	2,915	49-59
1940-51	16,718	6,441	39-59
1950-51	32,641	18,391	29-59
1960-61	41,342	17,969	25-56
1970-71	22,206	12,832	25-46
1980-81	10,864	5,515	25-36

**Figure 3.2:**  
**Dofile: graph\_fig3\_2.do**

Sample restricted to white non-Hispanic men and women ages 25-64. Sample weights [wgt] applied for all analyses. Observations in Figure 3.2 are mean percent never married by (year of birth x individual year of age) cells. (SOURCE: March CPS 1980-2016)

Number of observations for percent never married, used in Figure 3.2:

Birth cohort:	Percent never married Number of observations		Ages at which cohort is observed
	Less than BA	BA or more	
1940	21,048	6,223	40-64
1950	38,366	18,637	30-64
1960	43,542	18,232	25-56
1970	22,961	14,407	25-46
1980	11,480	6,826	25-36

**Figures 3.3 and 3.4:**  
**[Program: graph\_fig3\_3\_4.do]**

Figures 3.3 and 3.4: Sample restricted to WNH ages 25-64 with less than a four-year college degree. Regressions of conditions are run on a complete set of age indicators, and a complete set of year of birth indicators (1940-1988).

In Figure 3.3, year of birth indicators are then plotted in panels:

- Panel A contains coefficients from suicide mortality rates, chronic joint pain, sciatic pain, difficulty socializing, at risk for heavy drinking, and serious mental distress.
- Panel B contains coefficients from drug mortality, never married, not married, and not in the labor force (men).

In Figure 3.4, year of birth indicators are plotted for not in the labor force (men only) and (minus) log real wages (men only).

Data for Figures 3.3 and 3.4:

Drug mortality (SOURCE: NVSS 1992-2015) is the mean mortality rate from drug and alcohol poisoning, accidental or intent undetermined, by (year of birth x age) cell.

Suicide mortality (SOURCE: NVSS 1992-2015) is the mean suicide mortality rate by (year of birth x age) cell.

Log wages (SOURCE: CPS ORG 1979-2015) are mean log real wages (men only) by (year of birth x age) cells. Sample weights [orgwgt] applied. Figure 3.4 reports the negative of the year of birth effects.

Not in labor force (SOURCE: CPS March 1980-2016) is the fraction not in the labor force (men only) by (year of birth x age) cell. Sample weights [wgt] applied.

Never married (SOURCE: CPS March 1980-2016) is the fraction of men and women never married by (year of birth x age) cell. Sample weights [wgt] applied.

Not married (SOURCE: CPS March 1980-2016) is the fraction of men and women not currently married by (year of birth x age) cell. Sample weights [wgt] applied.

Chronic joint pain (SOURCE: NHIS 2002-2015) is the fraction of respondents reporting joint pain in past 30 days (jntsymp) that began more than 3 months ago (jntchr) by (year of birth x age) cell. Sample weights [wtfa\_sa] applied.

Sciatic pain (SOURCE: NHIS 1997-2015) is the fraction of respondents by (year of birth x age) reporting sciatic pain in the past three months, where sciatic pain is defined as =1 if respondent reports lower back pain in the past three months (painlb) that spread down either leg to areas below the knee (painleg); =0 otherwise. Sample weights [wtfa\_sa] applied.

Difficulty socializing (SOURCE: NHIS 1997-2015) is the fraction of respondents reporting the activity is *more than* "a little bit difficult" by (year of birth x age) cell. Sample weights [wtfa\_sa] applied.

Mental distress (SOURCE: NHIS 1997-2015) is the fraction of respondents by (year of birth x age) cell for whom the sum of responses to feeling sad, restless, worthless, hopeless, nervous, and everything was an effort was  $\geq 13$ . Each response is coded, during the past 30 days, the respondent felt this feeling: 0=none of the time, 1=a little of the time, 2=some of the time, 3=most of the time, 4=all of the time. Sample weights [wtfa\_sa] applied to calculate the fraction in mental distress.

At risk for heavy drinking (SOURCE: BRFSS 2001-2015) is the fraction of respondents by (year of birth x age) cell assessed by the BRFSS to be at risk for heaving drinking, which they define as more than 2 drinks daily (men) or more than 1 drink daily (women). Sample weights [A\_FINAL] applied.