**RDC Research Proposal**

Proposal Format Updated 07/2020

|  |  |
| --- | --- |
| **General Information** | |
| **Date:** | October 27, 2020 |
| **Title of Project:** | Understanding the Education Trend in Heart Disease Mortality |
| **NCHS Data System and Years:** | NHANES III and continuous NHANES 1999-2016NHANES III Adult Demographic, Laboratory, and Examination FilesNHIS 1999-2013 **2015 Restricted Use Linked Mortality Data** |
| **Non-NCHS Data Files:** | American Community Study (ACS) 5-year estimates, 2005-2009 **Census Data** |
| **Mode of Access:** | [\_] NCHS RDC, Hyattsville, MD [\_] NCHS RDC, Rockville, MD  [\_] NCHS RDC, Washington, DC (Government Only) [\_] NCHS RDC, Atlanta, GA  [X] Federal Statistical RDC, specify: National Bureau of Economic Research |
| **Statistical Software:**  (Check all that apply) | [\_] SAS/SUDAAN [X] Stata [\_] Other, specify: |
| **Proposed Start Date:** |  |
| **Funding Source:** |  |
| **Billing Address:**  (include contact person) |  |

List the name, institution, contact information, and role for anyone who will contribute to publications resulting from this project. Everyone listed must submit a C.V or resume. Add sections as needed.

|  |  |  |
| --- | --- | --- |
| **Research Team** | | |
| **Role:** | **Principal Investigator** | **Co-Investigator** |
| **Name** | David Cutler |  |
| **Email** | [dcutler@harvard.edu](mailto:dcutler@harvard.edu) |  |
| **Phone** | 617-496-5216 |  |
| **Institution** | Harvard University |  |
| **Mailing Address** | Department of Economics |  |
| **SSS Status? Y or N** | N |  |
| **US Citizen? Y or N** | Y |  |
|  | **Programmer** | **Programmer** |
| **Name** | Bryan Kim |  |
| **Email** | [kimcwa@bc.edu](mailto:kimcwa@bc.edu) |  |
| **Phone** | 949-274-2363 |  |
| **Institution** | National Bureau of Economic Research |  |
| **Mailing Address** | 78 Glenville Ave, Allston, MA, 02134 |  |
| **SSS Status? Y or N** | N |  |
| **US Citizen? Y or N** | Y |  |
|  | **Advisor (For Students and Post-Docs)**  [\_] [RDC-Student-Advisor Form](http://www.cdc.gov/rdc/Data/B3/Student_Agreement.pdf) | **Other, specify:** |
| **Name** |  |  |
| **Email** |  |  |
| **Phone** |  |  |
| **Institution** |  |  |
| **Mailing Address** |  |  |
| **SSS Status? Y or N** |  |  |
| **US Citizen? Y or N** |  |  |

Complete as applicable for your project. Address any “Yes” responses in the body of the proposal.

|  |  |  |
| --- | --- | --- |
| **RDC Proposal Summary Information** |  |  |
|  | **YES** | **NO** |
| **Geographic variables\*\*** |  |  |
| Level of geography to be shown in **output** (check all that apply) |  |  |
| National | X |  |
| Regional |  | X |
| State |  | X |
| MSA |  | X |
| County |  | X |
| Urban/rural classification |  | X |
| Census tract |  | X |
| Latitude/Longitude |  | X |
| Other |  | X |
| Will geographic identifier(s) be removed after merge |  | X |
| If yes, can true geographic identifiers be replaced with masked versions of these variables |  |  |
| Is GIS or mapping proposed |  | X |
| **Dates and Temporal information\*\*** |  |  |
| Are exact dates requested other than to calculate time of follow-up |  | X |
| If more than 1 year/cycle, will years/cycles be presented separately |  | NA |
| **Merging of data with NCHS restricted data\*\*** |  |  |
| Are external data being merged with NCHS data | X |  |
|  |  |  |
| **Linked Data Products\*\*** |  |  |
| Are you requesting linked Medicare/Medicaid files |  | X |
| If yes, are you using multiple years |  |  |
| Are you using public-use mortality data |  | X |

1. ***Abstract:*** *Please limit the project description abstract to 300 words.*

In a series of papers and a book, Case and Deaton (2015, 2017, 2020) document that the midlife mortality declines experienced in other wealthy countries over the last 20 years have not been as large in the US. This is especially true for non-Hispanic whites with less education. A good part of this difference is a result of less rapid declines in cardiovascular disease mortality in the US than in other countries. In this project, we examine why cardiovascular disease outcomes have not declined as rapidly for low education groups as for higher education groups. We will first examine trends in cardiovascular risk factors by education group and trends in survival conditional on risk. This will indicate whether differential mortality trends by education are due to differential risk factor management or to differential trends in outcomes conditional on risk. Second, we will estimate regression models for risk factor control and survival as a function of education and local area characteristics. The question we will examine is whether the effects of education are explainable with information on local area characteristics. This will help determine causally why trends have been adverse for low education groups.

1. ***Research Question:*** *Describe study purpose, hypotheses, goals, or research questions.*

This research has two specific aims.

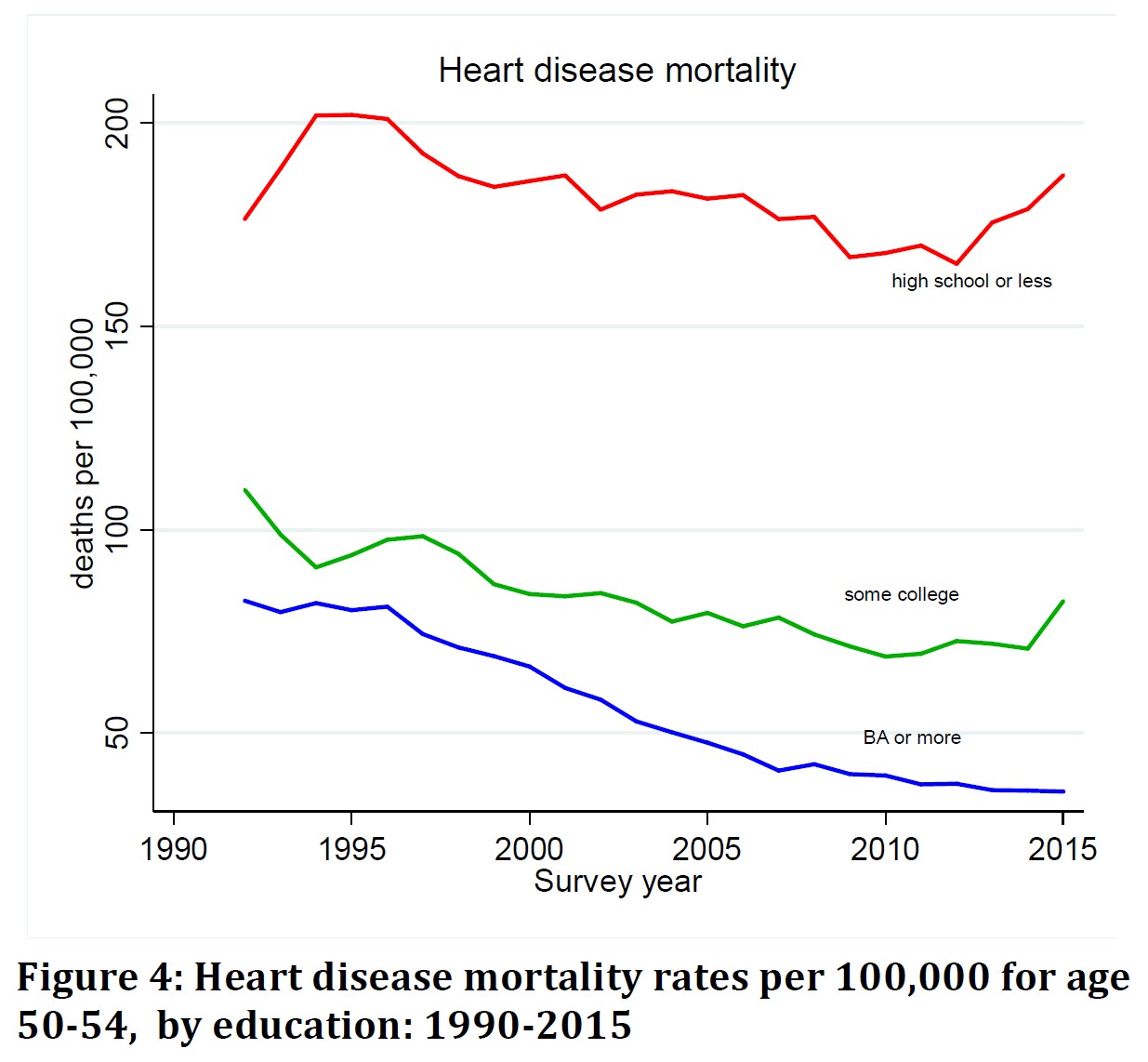
* First, we seek to differentiate the slowdown in cardiovascular disease mortality reductions for lower education groups into changes in risk factors for cardiovascular disease death and changes in death for people with the same level of risk.
* Second, we will relate adverse trends in risk and survival given risk to three measures of economic and social change: differential access to medical care for high and low education groups; economic changes that have been adverse to low education groups; and differential rates of despair.

1. ***Background:*** *Include a short literature review, no more than 2 pages, focusing on papers that discuss your topic and address the methodology that you plan to use. Please limit your reference list to 10 items or less.*

The slowdown in mortality reductions the United States has witnessed in the past two decades has been unprecedented. Life expectancy has declined for several years in a row, which is exceedingly rare. In a series of papers and a book, Case and Deaton (2015, 2017, and 2020) document these trends. The decline in the rate of mortality improvement has been particularly large for non-Hispanic whites, especially those without a college degree and those in late middle ages. By cause of death, the major contributing factors are “deaths of despair” (suicide, alcohol-related deaths, and drug overdose) and heart disease. Heart disease mortality, the subject of this proposal, has been flat for nearly two decades and has increased among lower socioeconomic status groups. This is very different from trends in the US prior to 2000 as well as from trends in European countries. The trend in heart disease deaths by education for people aged 50-54 is shown in the chart labeled figure 4 on the next page.

Several explanations have been put forward for the poor mortality performance in the US. The bulk of the research has focused on drug overdose deaths, the largest contributor to mortality increases in this time. Case and Deaton (2020) attribute those deaths, along with cardiovascular disease deaths, to despair from adverse economic and social change. Pierce and Schott (2020) show that trade liberalization was associated with higher numbers of drug overdoses.

However, other explanations have also been posited for the slowdown in heart disease mortality. The 1990s and 2000s were a period of widespread diffusion of new technologies for heart disease: statin drugs, new anti-hypertensives, and drug- eluding stents to name a few. It is possible that innovation had simply run its course and the slowdown in mortality reduction is a natural consequence of the tapering off of an earlier period of rapid innovation. Alternatively, changes in the medical system may be partly to blame. Cost sharing required to access medical care increased markedly in the 2010s (Kaiser Family Foundation, 2019), and this could have deterred low socioeconomic status groups from accessing care.



The goal of this research is to understand the slowdown in heart disease mortality reductions in the US, especially among low education groups, and to understand the economic, social, and medical changes that are driving it.

1. ***Public Health Benefit:*** *In one paragraph, how does your research benefit public health?*

The results will have major implications for how we understand mortality trends in the US, and what policy can do to improve health. For example, differential changes in risk factor management among high and low education groups would support policies that address better control of these risks. In contrast, differential survival given risk might indicate the need to focus on access to medical care.

#### Data Requirements:

*Please address the four items below and provide an explanation for “yes/checked” responses from the RDC Proposal Summary Information.*

#### Survey, Years, Files:

*List survey name(s) and survey years you would like to access. For examples, NHIS 2005-2007 Household, Person and Sample Adult Files, NAMCS 2005-2006 Provider and Patient Visit Files, or NHANES 2005-2006 Examination and Demographic Files.*

### NHANES III and continuous NHANES 1999-2016

NHANES III Adult Demographic, Laboratory, and Examination Files

Continuous NHANES through 2015-16 Demographic, Examination, Laboratory and Questionnaire Files

* Demographic variables and sample weights
* Physical examination measurements

### NHIS 1999-2013

Sample Adult and Person Files

* Demographic variables and sample weights

Hopelessness and worthlessness variables (variable names – HOPELESS and WORTHLS)

We are requesting two pieces of restricted information: (1) mortality data including month and year of death to match with NHANES III and the continuous NHANES, and (2) geographic information for NHANES III, the continuous NHANES, and NHIS 1999-2013. The geographic information will be used to merge on economic and social factors that might explain differential trends in cardiovascular disease mortality by socioeconomic status. After the merge, we do not need to keep the state or county identifier. However, we would like to create dummy county identifiers (i.e., 1, 2, …) so that we can include county fixed effects in our models.

#### Restricted-use Data:

*List and describe the restricted-use variables that you will need. These variables must be listed in the Data Dictionary section of this proposal. Explain why each variable is needed and how you will include them in your analysis. Specify how geographic variables, if applicable, will be used to merge files, analyze the data and/or be presented in output.*

### 2015 Restricted Use Linked Mortality Data, to link to NHANES III and continuous NHANES

SEQN – to merge with public use files ELIGSTAT – to adjust sample weights MORTSTAT

UCOD\_113 – want to know if COD is related to cardiovascular diseases. UCOD\_358 – want to know if COD is related to cardiovascular diseases. UCOD\_39 – want to know if COD is related to cardiovascular diseases. HYPTERTEN – analysis includes hypertension statistics.

DODMONTH – to calculate death within 5 years DODYEAR – to calculate death within 5 years MECMONTH – to calculate death within 5 years MECYEAR – to calculate death within 5 years

### Geographic information for NHANES

NHANES III: SEQN (to merge with public use files), STATE2K, CNTY2K – 2000 state and county codes.

Continuous NHANES from 1999-2000 to 2015-2016: SEQN (to merge with public use files), STATE2K, CNTY2K.

### Geographic information for NHIS

HHX, FMX, PX and FPX (to merge with public use files), STATE, COUNTY – state and county codes

#### Non-NCHS Data:

*If you plan to provide data from another source (such as Census or EPA) to merge to the restricted-use data, please describe the source, list the filename(s), and provide a general description of the non-NCHS data. The variables from the non-NCHS data must be listed in the Data Dictionary section.*

### American Community Study (ACS) 5-year estimates, 2005-2009

Unemployment rate in the county, by education group

Prime age male (25-54) employment-to-population ratio in the county, by education group Percent of adult population (18-64) uninsured in the county, by education group

### Census Data

Population change between 1980 and 2000, by county

#### Merge Variables:

*In detail, please describe the merge procedures needed to produce your analytic dataset(s). Highlight the variables used in the merge routine in the Data Dictionary. Leave blank if not applicable (e.g. NHDS, NAMCS/NHAMCS, Mortality, Natality, and DHHS Hosted Data Users). Note: The RDC Analyst will merge the data for you.*

* + 1. *“What variable(s) will be used to merge the public and restricted NCHS data files?*
    2. *What variable(s) will be used to merge the NCHS data files with any non-NCHS data?*

### NHANES III (1988-1994)

* + - 1. merge mortality information from restricted access file to NHANES III by SEQN from 1988 to 1994.
      2. merge state and county identifiers to NHANES III by SEQN from 1988 to 1994.
      3. merge economic information from ACS by county and education group\* and population change by county from Census.
      4. merge feelings of hopelessness and worthlessness from NHIS by county and education group\*.

### Continuous NHANES (1999-2016)

1. merge mortality information from restricted access file to continuous NHANES up to 2009-10 by SEQN.
2. merge state and county identifiers to continuous NHANES up to 2015-16 by SEQN.
3. merge economic information from ACS by county and education group\* and population change by county from Census.
4. merge feelings of hopelessness and worthlessness from NHIS by county and education group\*.

### NHIS 1999-2014

1) merge county to NHIS by PUBLICID (created using variables SRVY\_YR, HHX, FMX, PX, and FPX)

\* <High school; High school grad / GED; some college; college grad

#### Methodology:

*We highly recommend you familiarize yourself with the analytic guidelines of the data you intend to use. Any deviations from the methodology suggested in the guidelines will require explanation as it may pose a disclosure risk. Please address these three items in your proposal:*

#### Unit or Level of Analysis and Subpopulation(s):

*There can be many levels of analysis: be as detailed as possible in your description. A common example for an analysis using NHANES data is where the unit of analysis is the person while the subpopulation is adults ages 18-64. A common example involving geography is when the researcher aggregates persons to the state level to make comparisons between state policies.*

Unit of analysis – individual

Subpopulation 1 – people aged 45-74 who are white, black, or Hispanic, except those with implausible values for cardiovascular disease risk factors (BMI, blood pressure, cholesterol level, HbA1c) and those with missing covariate data (age, sex, education).

Subpopulation 2 – subpopulation 1 restricted to non-Hispanic whites

* 1. ***Analysis Plan:*** *Provide an overall analysis plan that specifies what analytic procedures or models you will use, such as prevalence estimates, logistic regression, or log-linear modeling. Also, list the specific statistical package procedures you will use.*

Our analysis will have two parts. The first part will determine whether the adverse trend in heart disease mortality rates for low education groups is due in larger part to adverse trends in risk factors such as smoking and poor blood pressure control or instead to adverse changes in death rates given risk. This analysis will use data from NHANES III and the continuous NHANES. To measure socioeconomic status, we will form four groups by education: ≤high school; high school/GED; some college; and college graduate. This delineation of socioeconomic status by education is standard in the literature. Depending on what we find, we may aggregate several of these groups, for example into college graduate or not.

The measure of cardiovascular disease risk we derive will use risk factors from the Framingham Heart Study (D’Agostino, et al., 2008). It combines age, sex, total and high-density lipoprotein, cholesterol, systolic blood pressure, smoking and diabetes status. All the data necessary to do this are in the public use NHANES data releases.

We will start by plotting trends in cardiovascular risk and cardiovascular death within 5 years, separately by education. The template for these plots is shown in exhibit 1. Panels (a) and (b) are for subpopulation 1, encompassing all races and ethnicities; panels (c) and (d) restrict the analy- sis to subpopulation 2, non-Hispanic whites. Within these two groups, the left chart (panels (a) and (c)) shows the trend in risk and the right panel shows the trend in actual CVD mortality. Risk factor trends are shown through 2018, the most recent year of public use data available. Death within five years is shown through 2010, since the data on death are available through 2015.

Differential trends in disease risk would be shown as a divergence in the predicted mortality risk between the low and high education groups in panels (a) and (c) that occurs after 2000. Differen- tial trends in mortality rates conditional on disease risk would be shown as a divergence in out- comes in panels (b) and (d) after 2000 between the low and high education groups that is not ap- parent in panels (a) and (c).

The second part of the analysis is to understand why outcomes have changed differentially for higher and lower SES groups. The specific outcome we will examine will depend on the results of aim 1. We illustrate the analysis assuming that what we find from the earlier analysis is that cholesterol levels have been declining more rapidly for higher education groups than for lower education groups post-2000. Our analytic approach would readily extend to other situations such as findings that mortality conditional on risk has diverged.

We posit three classes of explanations for why cholesterol or other trends might differ: (1) eco- nomic changes have adversely affected the health of less educated groups; (2) health insurance is less accessible for less educated groups than for better educated groups; and (3) despair is more common among less educated groups and this adversely affects the health. These theories are ob- viously related, for example economic change may lead to despair and reduced insurance cover- age. Still, they imply that different types of interventions may be most effective in improving health. The specific variables we will utilize to test for these theories are as follows:

Economic change

* EPOPea / URea – the employment-to-population ratio and unemployment rate among prime age men (ages 25-54) in education group e and area a, on average between 2005 and 2009, as measured in the American Community Study.
* POPCHa – population change between 1990 and 2010, as measured by the Census.

Health Insurance

* UIea – the share of people in education group e and area a who are uninsured, on average between 2005 and 2009, as measured in the American Community Study

Despair

* Dea – the share of people in education group e and area a who report feeling hopeless or worthless on average between 1999 and 2013, as measured in the NHIS

Our methodology for testing these theories is as follows. Consider a person i with education level e who lives in area a at time t. For now, consider the time period t as being any year from 1999- 2016. They have outcome Yieat (cholesterol, obesity, taking medications, etc.). For ease of presentation, we consider only one explanatory variable, in this case the share of the population that is uninsured. The models we estimate will include all of the variables together. The specifi- cation will be of the form:

Yieat = β1e\*UIea\*Year + **X**ieat\***β** + θa + θe + θt + eieat (1)

The X variables are demographics: age, sex, and race/ethnicity when using subpopulation 1; age and sex when using subpopulation 2. The coefficients θa, θe, and θt are dummy variables for ar- eas, education groups, and year. These pick up differences in the outcome across areas and de- mographic groups, and over time. Note that we wish to include county dummy variables so we will want to create a fake county variable, i.e. (county=1, 2, 3, …). Year is a time trend (i.e., 1999=0, 2000=1, …)

The coefficients β1e are the key coefficient of interests. For each education group e, they show the trend in outcomes over time for that education group, controlling for demographics. The spe- cific hypothesis we will test is whether β1lowSES = β1highSES, that is trends in outcomes are the same for high and low SES individuals. For an outcome for which high values are worse (e.g., cholesterol), a finding that β1lowSES > β1highSES would indicate that health has deteriorated more for lower SES individuals than for high SES individuals in areas where uninsurance rates for low income people are higher.

Exhibit 2 and 3 show how the results will be presented. Exhibit 2 presents sample means; exhibit 3 shows regression coefficients. It is worth noting that we will not publish information about health in any particular state or county. We will use geographic information solely to form co- variates in regression models.

One concern about the findings from regression (1) is that outcomes might have been trending differently in areas with high and low rates of uninsured for separate reasons. While we cannot rule this out entirely without identifying every other potential variable to include, we can exam- ine the theory using data from the NHANES III through the early round of the continuous

NHANES. In the 1990s, heart disease mortality reductions were relatively similar across educa- tion groups. Thus, the share of the population that is uninsured in the 2000s should not predict cardiovascular disease trends in the 1990s.

We can test this by pooling the NHANES III and the continuous NHANES 1999-2000. Using a data set of these two groups, we will estimate a model similar to that in equation (1), though Year will be a dummy variable for 1999-2000 data vs. 1988-94, rather than a time trend. The output will be as shown in the second column of exhibit 2. A failure to reject that β1lowSES = β1highSES would indicate that health was not deteriorating more in the 1990s for areas that would subse- quently experience job loss, high unemployment, and despair. If β1lowSES > β1highSES, that would indicate that those areas were trending adversely prior to the 2000s, and therefore that there were ‘pre-trends’ in poor outcomes in areas that subsequently did worse economically and demograph- ically.

* 1. ***Complex Survey Design:*** *Indicate how you will address sample weights, design variables, and other adjustments for the use of complex survey data, if applicable, using the statistical software listed in the General Information area. A detailed description per weight, design variables, and other adjustments are required and central to understanding the limitation of the data. This is a critical element to describe and is central to the proposal review process.*

We will adjust the standard errors for the complex sample design, using STATA’s “svy” commands. We will use examination weight variables to weight individuals in measuring risk of cardiovascular disease death. We will follow the NCHS guidelines on weighting linked data created from the NHANES interview and examination data and NHANES mortality data. We will create adjusted survey weights to account for linkage-ineligibility and non-matches by using NHANES sample weights within different subgroups to estimate population counts of the whole sample (linkage-eligible, linkage-ineligible, and non-matches). Then we pool linkage-eligible respondents and adjust the NHANES sample weight so that the total sum of the adjusted weights in each subgroup is the same as that for the whole sample; the adjusted weights for the respondents who are not linkage eligible are 0. This is an example of one such approach to adjust weights for ineligible respondents.

Example of our proposed code is as follows:

svyset sdmvpsu [pw=wtmec2yr], strata(sdmvstra) svy: regress cholesterol UI\_year …

#### Output:

*Please describe in full detail all of the output you need to look at in order to ensure that the data file created for you is correct and complete. NCHS will only provide you with a SAS proc contents as the default if you do not specify any additional output needed.*

*Please describe in full detail all of the output you would like to take out of the RDC that is needed for the final product you wish to create. Please be very specific in describing your output needs (i.e., state groupings, levels of output and how restricted-use data will be displayed) as this section is necessary for the Review Committee to assess disclosure risk of your output. Your examples should reflect the geographic variables you will use in your output. Note: Any output produced in the RDC that is not listed in your proposal will not be allowed out of the RDC.*

* 1. ***Output Needed to Confirm Accuracy of Analytic Dataset:*** *Since your RDC Analyst will create your analytic dataset, please describe what output you need to review to ensure that your dataset was constructed accurately and is complete (e.g., need to review univariate frequencies of certain variables – please specify).*

We would like to see the merge statistics. We will check that:

* In the mortality data, all death information should merge to someone in the NHANES file.
* In the geographic data, all geographic areas in NHANES should match to economic variables, health insurance, and despair measures.
  1. ***Table Shells:*** *Include detailed examples of all table shells, models, and/or graphs with titles. Indicate the subsample and unit of analysis used in each type of table, model, or graphs. Your proposal will not be approved without this information.*

### Exhibit 1: Trends in Cardiovascular Risk and Cardiovascular Mortality by SES\* All Race/Ethnic Groups

1. CVD Risk

4

3

Percent

2

1

0

1. Five Year CVD Death

4

3

Percent

2

1

0



No college degree College grad

No college degree College grad

### Non-Hispanic Whites

1. CVD Risk

4

3

Percent

2

1

1. Five Year CVD Death

4

3

Percent

2

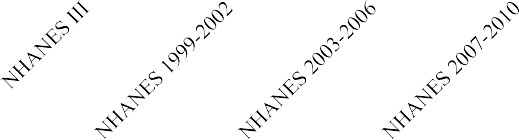
1

0



No college degree College grad

0



No college degree College grad

Note: The sample is people aged 45-74 in the indicated year who are not missing on demographics, risk variables, and merged to mortality information. Disease risk is formed from the Framingham Heart Study equation. The data are weighted to national totals. For simplicity, we show two education groups, though our final analysis may have up to four groups.

\* Data are hypothetical.

|  |  |
| --- | --- |
| **Exhibit 2: Summary Statistics** | |
| **Variable** | **Continuous NHANES Mean (std dev)** |
| **Demographics** |  |
| Age |  |
| Male (%) |  |
| Non-Hispanic white (%) |  |
| Non-Hispanic black (%) |  |
| Hispanic |  |
| **Cardiovascular disease risk** |  |
| Predicted 5 year mortality |  |
| Systolic / diastolic blood pressure |  |
| Total cholesterol |  |
| LDL cholesterol |  |
| HbA1c |  |
| BMI |  |
| Current smoker / former smoker (%) |  |
| **Cardiovascular outcomes** |  |
| Cardiovascular death in next 5 years (%) |  |
|  |  |
| **Economic variables (%, average for county)** |  |
| Employment-to-population ratio, 2005-2009 |  |
| Unemployment rate, 2005-2009 |  |
| Population change, 1990-2010 |  |
|  |  |
| **Health insurance (average for county)** |  |
| Uninsured 2005-2009 (%) |  |
|  |  |
| **Despair (%, average for county)** |  |
| Percent feeling helpless or worthless, 1999-2014 |  |
| Note: All tabulations are from the continuous NHANES, 1999-2016, with the exception of cardiovascular disease death in the next five years, which is meas- ured for 1999-2010 years. Data are for people aged 45-74 at the time of the MEC sample. Data are weighted to national totals. Standard errors are reported for continuous variables and are adjusted for complex survey design. | |

|  |  |  |
| --- | --- | --- |
| **Exhibit 3: Regressions for Cholesterol Levels** | | |
| **Coefficient on variable interacted with time trend:** | **Continuous NHANES (1999-2016)** | **NHANES III**  **and continuous NHANES (1988-2000)** |
| **Panel A: All Race/Ethnicities** | | |
| Employment-to-population rate |  |  |
| Unemployment rate |  |  |
| Population change |  |  |
| Percent uninsured |  |  |
| Despair |  |  |
| Other controls | Age, sex, race/ethnicity | Age, sex, race/ethnicity |
| N |  |  |
| R2 |  |  |
| **Panel B: Non-Hispanic White Population** | | |
| Employment-to-population rate |  |  |
| Unemployment rate |  |  |
| Population change |  |  |
| Percent uninsured |  |  |
| Despair |  |  |
| Other controls | Age, sex | Age, sex |
| N |  |  |
| R2 |  |  |
| Note: The sample is people aged 45-74 who are not missing on demographics and risk variables. Regression also include five-year age-sex dummy variables, and dummy variables for race/ethnicity in the upper panel. The data are weighted to national totals. Standard errors account for the complex survey design. | | |

* 1. ***Presentation of Results:*** *Describe how you will present the results (in a report, publication in a peer-reviewed journal, presentation at a scientific meeting, or used for internal policy analysis).*

We intend to write one or more research papers. They will be submitted for publication in peer- reviewed journals and presented at scientific meetings.

#### Data Dictionary:

*Include a data dictionary for each data source. Provide a public-use and restricted-use data dictionary for NCHS survey data. This should be a listing of variables you would like in your dataset. Also, provide a data dictionary for any non-NCHS data you want to use. See instructions and examples for* [*creating the data dictionary.*](http://www.cdc.gov/rdc/B3Prosal/PP323.htm) *When asking for multiple years of data, make sure to reflect the public-use file layout for each year as variable names can change over years. Include all explanations in Section E. Data Requirements.*

***Note: NHDS, NAMCS/NHAMCS, Mortality, Natality, and DHHS Hosted Data Users:*** *Provide a single data dictionary that includes all the variables (public-use and restricted-use) you would like extracted for your analytic data set.*

In addition to the data dictionary here, the attached excel spreadsheets show the data we request.

Proposal Format Updated 07/2020

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Public Use Variables: NHANES 1999-2018** | | | | | | | | | | | |
| Variable Name | Variable description | 1999-  2000 | 2001-  2002 | 2003-  2004 | 2005-  2006 | 2007-  2008 | 2009-  2010 | 2011-  2012 | 2013-  2014 | 2015-  2016 | 2017-  2018 |
| **Demographic File** | | | | | | | | | | | |
| seqn \* | Respondent sequence number (merge to restricted data) | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| wtint2yr | Interviewed Sample Persons. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| wtmec2yr | Both Interviewed and MEC Examined Sample Persons. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| ridageyr | Best age in years of the sample person at time of HH screening. Individuals 80 and over are topcoded at 80 years of age. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| riagendr | Gender of the sample person | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| ridreth1 | Recode of reported race and ethnicity information. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| ridreth2 | Linked NH3 Race and Ethnicity Recode. | Y | Y | Y | N | N | N | N | N | N | N |
| dmdeduc2 \* | What is the highest grade or level of school {you have/SP has} completed or the highest degree {you have/s/he has} received? |  |  | Y | Y | Y | Y | Y | Y | Y | Y |
|  | Y | Y |  |  |  |  |  |  |  |  |
| dmdhredu | What is the highest grade or level of school {you have/NON\_SP HEAD has} received? | Y | Y | Y | Y | Y | Y | Y | Y | Y | N |
| dmqmilit | Did {you/SP} ever serve in the Armed Forces of the United States? | Y | Y | Y | Y | N | N | N | N | N | N |
| indhhinc | Total household income (reported as a range value in dollars) | Y | Y | Y | Y | N | N | N | N | N | N |
| indhhin2 | Total household income (reported as a range value in dollars) | N | N | N | N | Y | Y | Y | Y | Y | Y |
| indfminc | Total family income (reported as a range value in dol- lars) | Y | Y | Y | Y | N | N | N | N | N | N |
| indfmin2 | Total family income (reported as a range value in dol- lars) | N | N | N | N | Y | Y | Y | Y | Y | Y |
| indfmpir | Poverty income ratio (PIR) - a ratio of family income to poverty threshold | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| DMDBORN | In what country {were you/was SP} born? | Y | Y | Y | Y | N | N | N | N | N | N |
| DMDBORN2 | In what country {were you/was SP} born? | N | N | N | N | Y | Y | N | N | N | N |
| DMDBORN4 | In what country {were you/was SP} born? | N | N | N | N | N | N | Y | Y | Y | Y |
| DMDMARTL | Marital Status | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable Name | Variable description | 1999-  2000 | 2001-  2002 | 2003-  2004 | 2005-  2006 | 2007-  2008 | 2009-  2010 | 2011-  2012 | 2013-  2014 | 2015-  2016 | 2017-  2018 |
| **Examination File** | | | | | | | | | | | |
| bmxwt | Weight (kg) | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bmxbmi | Body Mass Index (kg/m\*\*2) | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxsy1 | Systolic: Blood pressure (first reading) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxsy2 | Systolic: Blood pressure second reading) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxsy3 | Systolic: Blood pressure third reading) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxsy4 | Systolic: Blood pressure fourth reading if necessary) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxdi1 | Diastolic Blood pressure (first reading) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxdi2 | Diastolic Blood pressure second reading) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxdi3 | Diastolic Blood pressure third reading) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxdi4 | Diastolic Blood pressure fourth reading if necessary) mm Hg | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| bpxsar | Systolic blood pressure average | Y | Y | Y | N | N | N | N | N | N | N |
| bpxdar | Diastolic blood pressure average | Y | Y | Y | N | N | N | N | N | N | N |
| **Laboratory File** |  |  |  |  |  |  |  |  |  |  |  |
| lbxtc | Total cholesterol (mg/dL) | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Lbxhdd | Direct HDL-Cholesterol (mg/dL) | N | N | Y | N | N | N | N | N | N | N |
| Lbdhdd | Direct HDL-Cholesterol (mg/dL) | N | N | N | Y | Y | Y | Y | Y | Y | Y |
| lbdhdl | HDL-cholesterol (mg/dL) | Y | Y | N | N | N | N | N | N | N | N |
| lbdtrsi | Triglyceride (mmol/L) | Y | Y | Y | Y | Y | Y | Y | Y | Y | N |
| lbdldl | LDL-cholesterol (mg/dL) | Y | Y | Y | Y | Y | Y | Y | Y | Y | N |
| lbxgh | Glycohemoglobin (%) | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| **Questionnaire Files** | | | | | | | | | | | |
| **SMQ** |  |  |  |  |  |  |  |  |  |  |  |
| smq020 | Smoked at least 100 cigarettes in life | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| smq040 | Do you now smoke cigarettes? | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| smq050q | How long has it been since {you/SP} quit smoking cig- arettes? | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable Name | Variable description | 1999-  2000 | 2001-  2002 | 2003-  2004 | 2005-  2006 | 2007-  2008 | 2009-  2010 | 2011-  2012 | 2013-  2014 | 2015-  2016 |
| **OCQ** |  |  |  |  |  |  |  |  |  |  |
| OCD150 | Type of work done last week | N | Y | Y | Y | Y | Y | Y | Y | Y |
| ocq180 | How many hours did {you/SP} work last week at all jobs or busi- nesses? | Y | N | N | N | Y | Y | Y | Y | Y |
| ocd230 | What kind of business or industry is this? (For example: TV and radio management, retail shoe store, state labor department, farm.) | Y | Y | Y | N | N | N | N | N | N |
| OCD231 | What kind of business or industry is this? (For example: TV and radio management, retail shoe store, state labor department, farm.) | N | N | N | Y | Y | Y | Y | Y | N |
| ocd240 | (SP Interview Version) What kind of work {were you/was SP} doing? (For example: farming, mail clerk, computer specialist.) (Family Interview Version) What kind of work {were/was}  {you/NON-SP HEAD/NON-SP SPOUSE} doing? (For example: farming, mail clerk, computer specialist.) | Y | Y | Y | N | N | N | N | N | N |
| OCQ150 | Type of work done last week | Y | N | N | N | N | N | N | N | N |
| ocq260 | Looking at the card, which of these best describes this job or work situation? | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| ocq290q | The next questions are about conditions {you/SP} may experience and equipment {you/he/she} may use at {EMPLOYER} as a(n)  {OCCUPATION}. At this job or business, how many hours per day can {you/SP} smell the smoke from other people's cigarettes, cigars, and/or pipes? | Y | Y | Y | Y | Y | Y | N | N | N |
| ocq390g | Thinking of all the paid jobs or businesses {you/SP} ever had, what kind of work {were you/was s/he} doing the longest? (For example, electrical engineer, stock clerk, typist, farmer.) | Y | N | N | N | N | N | N | N | N |
| ocd390 | Thinking of all the paid jobs or businesses {you/SP} ever had, what kind of work {were you/was s/he} doing the longest? (For example, electrical engineer, stock clerk, typist, farmer.) | Y | Y | Y | N | N | N | N | N | N |
|  | Thinking of all the paid jobs or businesses {you/SP} ever had, what kind of work {were you/was s/he} doing the longest? (For example, electrical engineer, stock clerk, typist, farmer.) | N | Y | Y | Y | Y | Y | Y | Y | Y |
| ocd390G |  |  |  |  |  |  |  |  |  |
| Variable Name | Variable description | 1999-  2000 | 2001-  2002 | 2003-  2004 | 2005-  2006 | 2007-  2008 | 2009-  2010 | 2011-  2012 | 2013-  2014 | 2015-  2016 |

Thinking of all the paid jobs or businesses {you/SP} ever had, What kind of business or industry is this? (For example: a TV or,

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| what kind of work {were you/was s/he} doing the longest? (For N | N | N | Y | Y | Y | Y | Y | N |
| OCD392 example, electrical engineer, stock clerk). |  |  |  |  |  |  |  |  |
| radio management, retail shoe store, state labor department, N  OCD391 farm.) | N | N | N | Y | Y | Y | Y | N |
| **WHQ**  whd010 How tall {are you/is SP} without shoes? Y | Y | Y | Y | Y | Y | Y | Y | Y |
| whd020 How much {do you/does SP} weigh without clothes or shoes? Y | Y | Y | Y | Y | Y | Y | Y | Y |
| whd050 How much did {you/SP} weigh a year ago? Y | Y | Y | Y | Y | Y | Y | Y | Y |
| whd110 How much did {you/SP} weigh 10 years ago? [If you don't know Y | Y | Y | Y | Y | Y | Y | Y | Y |
| How much did {you/SP} weigh at age 25? [If you don't know |  |  |  |  |  |  |  |  |
| whd120 {your/his/her} exact weight, please make your best guess.] If ( Y | Y | Y | Y | Y | Y | Y | Y | Y |
| fore (your/her) pregnancy? |  |  |  |  |  |  |  |  |
| whd140 Up to the present time, what is the most {you have/SP has} ever Y | Y | Y | Y | Y | Y | Y | Y | Y |
| sdmvpsu Masked Variance Unit Pseudo-PSU variable for variance estima- Y | Y | Y | Y | Y | Y | Y | Y | Y |
| sdmvstra Masked Variance Unit Pseudo-Stratum variable for variance esti- Y | Y | Y | Y | Y | Y | Y | Y | Y |
| **BPQ** |  |  |  |  |  |  |  |  |
| bpq020 Ever told you had high blood pressure Y | Y | Y | Y | Y | Y | Y | Y | Y |
| BPQ040A Taking prescription for hypertension Y | Y | Y | Y | Y | Y | Y | Y | Y |
| BPQ040B Told to control weight for hypertension Y | Y | Y | N | N | N | N | N | N |
| BPQ040C Told to reduce sodium for hypertension Y | Y | Y | N | N | N | N | N | N |
| BPQ040D Told to exercise more for hypertension Y | Y | Y | N | N | N | N | N | N |
| 1999- | 2001- | 2003- | 2005- | 2007- | 2009- | 2011- | 2013- | 2015- |
| Variable Name Variable description 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 | 2014 | 2016 |
| BPQ040E Told to reduce alcohol for hypertension Y | Y | Y | N | N | N | N | N | N |
| BPQ043A Told to stop smoking for hypertension Y | Y | Y | N | N | N | N | N | N |
| BPQ043B Told to increase potassium for hyprtnsn Y | Y | Y | N | N | N | N | N | N |
| BPQ043C Told of other diet changes for hyprtnsn Y | Y | Y | N | N | N | N | N | N |

{your/his/her} exact weight, please make your best guess.]

you were/she was) pregnant, how much did (you/she) weigh be-

weighed?

tion

mation

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BPQ080 | Doctor told you - high cholesterol level | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| BPQ100D | Now taking prescribed medicine | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| BPQ090A | Told to eat less fat for cholesterol | Y | Y | Y | Y | Y | Y | N | N | N |
| BPQ090B | Told to reduce weight for cholesterol | Y | Y | Y | Y | Y | Y | N | N | N |
| BPQ090C | Told to exercise more for cholesterol | Y | Y | Y | Y | Y | Y | N | N | N |
| BPQ090D | Told to take prescriptn for cholesterol | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| BPQ050B | Now controlling weight | Y | Y | Y | N | N | N | N | N | N |
| BPQ050C | Now reducing salt/sodium | Y | Y | Y | N | N | N | N | N | N |
| BPQ050D | Now exercising more | Y | Y | Y | N | N | N | N | N | N |
| BPQ050E | Now reducing alcohol consumption | Y | Y | Y | N | N | N | N | N | N |
| BPQ100A | Now eating fewer high fat foods | Y | Y | Y | Y | Y | Y | N | N | N |
| BPQ100B | Now controlling weight | Y | Y | Y | Y | Y | Y | N | N | N |
| BPQ100C | Now increasing exercise | Y | Y | Y | Y | Y | Y | N | N | N |
| BPQ100D | Now taking prescribed medicine | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| BPQ110A | Eating fewer high fat foods on own | Y | N | N | N | N | N | N | N | N |
| BPQ110B | Controlling weight on own | Y | N | N | N | N | N | N | N | N |
| BPQ110C | Increasing exercise on own | Y | N | N | N | N | N | N | N | N |
| **DIQ** |  |  |  |  |  |  |  |  |  |  |
| DIQ010 | Doctor told you have diabetes | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| DIQ050 | Taking insulin now | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| DIQ070 | Take diabetic pills to lower blood sugar | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| **MCQ** |  |  |  |  |  |  |  |  |  |  |
|  | Has a doctor or other health professional ever told {you/SP} that  {you/s/he} . . .had a heart attack (also called myocardial infarction (my-o-car-dee-al in-fark-shun))? | N | N | N | N | N | N | Y | Y | Y |
| MCQ160e |  |  |  |  |  |  |  |  |  |
|  | Has a doctor or other health professional ever told {you/SP} that  {you/s/he} . . .had a heart attack (also called myocardial infarc- tion)? | Y | Y | Y | Y | Y | Y | N | N | N |
| MCQ160E |  |  |  |  |  |  |  |  |  |
| Variable Name | Variable description | 1999-  2000 | 2001-  2002 | 2003-  2004 | 2005-  2006 | 2007-  2008 | 2009-  2010 | 2011-  2012 | 2013-  2014 | 2015-  2016 |
| MCQ160B | Has a doctor or other health professional ever told {you/SP} that  {you/s/he} . . .had congestive heart failure? | Y | Y | Y | Y | Y | Y | N | N | N |
| MCQ160b | Has a doctor or other health professional ever told {you/SP} that  {you/s/he} . . .had congestive heart failure? | N | N | N | N | N | N | Y | Y | Y |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MCQ220 | Ever told you had cancer or malignancy | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| mcq230a | What type of cancer? | N | N | N | N | N | N | Y | Y | Y |
| mcq230A | What kind of cancer was it? | Y | Y | Y | Y | Y | Y | N | N | N |
| mcq230b | What type of cancer? | N | N | N | N | N | N | Y | Y | Y |
| mcq230B | What kind of cancer was it? | Y | Y | Y | Y | Y | Y | N | N | N |
| mcq230c | What type of cancer? | N | N | N | N | N | N | Y | Y | Y |
| mcq230C | What kind of cancer was it? | Y | Y | Y | Y | Y | Y | N | N | N |
| mcq230d | What type of cancer? | N | N | N | N | N | N | Y | Y | Y |
| mcq230D | What kind of cancer was it? | Y | Y | Y | Y | Y | Y | N | N | N |

**Restricted Variables: 2015 Linked Mortality File (NHANES 1999-2014 Mortality 2015)**

Variable Name Variable Label

SEQN \* “NHANES Respondent Sequence Number” (merge to public use data)

ELIGSTAT "Eligibility Status for Mortality Follow-up" MORTSTAT "Final Mortality Status"

UCOD\_113 “CD-10 Underlying Cause of Death 113 Groups Recode (all years)”

UCOD\_358 “CD-10 Underlying Cause of Death 358 Groups Recode (availa- ble 1999 forward”

UCOD\_39 “CD-10 Underlying Cause of Death 39 Groups Recode (available 1999 forward)”

DIABETES "Diabetes flag from Multiple Cause of Death" HYPTERTEN "Hypertension flag from Multiple Cause of Death"

permth\_int "Number of Person-Months of Follow-up from NHANES Inter- view date"

permth\_exm "Number of Person-Months of Follow-up from NHANES Mobile Examination Center (MEC) Date"

DODMONTH “Month of Death”

DODYEAR “Year of Death” MECMONTH “Month of MEC Exam” MECYEAR “Year of MEC Exam”

**Public - Use Variables: NHANES III**

|  |  |
| --- | --- |
| **Household Adult File**  SEQN Sequence number  WTPFQX6 Final Interview Weight  WTPFEX6 Final examination (MEC only) weight WTPFHX6 Final MEC + home examination weight WTPFQX1 Phase 1 interviewed sample final wgt WTPFEX1 Phase 1 MEC examined sample final wgt WTPFHX1 Phase 1 M+H examined sample final wgt WTPFQX2 Phase 2 interviewed sample final wgt WTPFEX2 Phase 2 MEC examined sample final wgt WTPFHX2 Phase 2 M+H examined sample final wgt | |
| DMARETHN | Race-ethnicity |
| DMARACER | Race |
| DMAETHNR | Ethnicity |
| HSAGEIR | Age at interview (Screener) |
| HSAGEU | Age at interview-unit (Screener) |
| HSAITMOR | Age in months (Screener) |
| HFAGERR | Respondent age in years |
| HSSEX | Sex |
| SDPPSU6 | Total NHANES III pseudo-PSU |
| SDPSTRA6 | Total NHANES III pseudo-stratum |
| HFA8R \* | Highest grade or yr of school completed |
| DMPCNTYR | County code |
| DMPFIPSR | FIPS code for State |
| DMPMETRO | Rural/urban code based on USDA code |
| DMPCREGN | Census region, weighting(Texas in south) |
| HAZA8AK1 | K1 for first BP measure (systolic, mmHg) |
| HAZA8AK5 | K5 for first BP measure diastolic mmHg) |
| HAZA8BK1 | K1 for second BP measure (systolic, mmHg) |
| HAZA8BK5 | K5 for second BP measure diastolic mmHg) |
| HAZA8CK1 | K1 for third BP measure (systolic, mmHg) |
| HAZA8CK5 | K5 for third BP measure diastolic mmHg) |
| HAZA8DK1 | K1 for fourth BP measure (systolic, mmHg) |
| HAZA8DK5 | K5 for fourth BP measure diastolic mmHg) |
| HAZA11A | Average K1 BP (survey years 5 and 6) |

|  |  |
| --- | --- |
| HAZA11AR | Average K1 BP after data edits |
| HAZA11B | Average K5 BP (survey years 5-6) |
| HAZA11BR | Average K5 BP after data edits |
| HAZNOK1R | Number of BP's used for average K1 |
| HAZNOK5R | Number of BP's used for average K5 |
| HAD1 | Ever been told you have sugar/diabetes |
| HAD6 | Are you now taking insulin |
| HAD7S | How often you take insulin - times/week |
| HAD10 | Are you now taking diabetes pill |
| HAE2 | Doctor ever told had hypertension/HBP |
| HAE3 | Told 2+ times you had hypertension/HBP |
| HAE4A | Doctor told take prescribed med for HBP |
| HAE4B | Doctor told control/lose weight for HBP . |
| HAE4C | Doctor told cut salt intake for HBP |
| HAE4D | Doctor told do anything else for HBP |
| HAE4D1 | Doctor told exercise more for HBP |
| HAE4D2 | Doctor told restrict alcohol for HBP |
| HAE4D3 | Doctor told stop smoking for HBP ... |
| HAE4D4 | Doctor told reduce tension for HBP |
| HAE4D5 | Doctor told change diet for HBP |
| HAE4D6 | Doctor told make other changes for HBP |
| HAE5A | Now taking prescribed medicine for HBP |
| HAE5B | Now controlling or losing weight for HBP . |
| HAE5C | Now using less salt or sodium for HBP |
| HAE5D1 | Now exercising for high blood pressure |
| HAE5D2 | Now restricting alcohol for HBP |
| HAE5D3 | Now quit smoking for high blood pressure |
| HAE5D4 | Now reduced tension for HBP |
| HAE5D5 | Now made diet changes for HBP |
| HAE5D6 | Now made other changes for HBP |
| HAE7 | Doctor told blood cholesterol level high |
| HAE8A | Doctor told eat less fat for HBC |
| HAE8B | Doctor told lose weight for HBC |
| HAE8C | Doctor told exercise for HBC |
| HAE8D | Doctor told take medicine for HBC |
| HAE9A | Now eat less high fat foods for HBC |

|  |  |
| --- | --- |
| HAE9B | Now losing weight to lower cholesterol |
| HAE9C | Now exercising to lower cholesterol |
| HAE9D | Take prescribed med to lower cholesterol |
| HAS1 | Past 2 wks, did you work at job/business . |
| HAS2 | Past 2 weeks, did you have job/business |
| HAS8R | What kind of business or industry -- rcd . |
| HAS9R | What kind of work were you doing -- rcd . |
| HAS11 | Class of worker |
| HAS12S | How long in months worked at that job |
| HAV12R | Where were you born -- recode |
| HFA13 | Military service |
| HFA12 | Marital Status |
| HAR1 | Have you smoked 100+ cigarettes in life |
| HAR3 | Do you smoke cigarettes now |
| HAR10 | How many years not smoked since started |
| HAC1C | Doctor told: congestive heart failure |
| HAC1D | Doctor ever told you had: stroke |
| HAC1N | Doctor ever told you had: skin cancer |
| HAC1O | Doctor ever told you had: other cancer |
| HAF10 | Doctor ever told you had a heart attack |
| HAC4A | Doctor told relatives they had diabetes |
| HAC4B | Any relatives had heart attack before 50 |
| HAC5B3 | Did mother have heart attack |
| HAC5B4 | Did father have heart attack |
| **Examination File** |  |
| BMPBMI | Body mass index |
| PEP6G1 | K1, systolic, for 1st BP (mmHg)(age 5+) |
| PEP6H1 | K1, systolic, for 2nd BP (mmHg)(age5+) |
| PEP6I1 | K1, systolic, for 3rd BP (mmHg)(age 5+) |
| PEPMNK1R | Overall average K1, systolic, BP(age 5+) |
| PEP6G3 | K5, diastolic, for 1st BP (mmHg)(age 5+) |
| PEP6H3 | K5, diastolic, for 2nd BP (mmHg)(age 5+) |
| PEP6I3 | K5, diastolic, for 3rd BP (mmHg)(age 5+) |
| PEPMNK5R | Overall average K5, diastolic, BP(age5+) |

|  |  |
| --- | --- |
| **Laboratory File** |  |
| TGP | Serum triglycerides (mg/dL) |
| LCP | Serum LDL cholesterol (mg/dL) |
| HDP | Serum HDL cholesterol (mg/dL) |
| TCP | Serum cholesterol (mg/dL) |
| GHP | Glycated hemoglobin: (%) |
| **Restricted Use Variables: 2015 Linked Mortality File (NHANES III)** | |
| Variable Name | Variable Label |
| SEQN \* | “NHANES Respondent Sequence Number” (merge to public use data) |
| ELIGSTAT | "Eligibility Status for Mortality Follow-up" |
| MORTSTAT | "Final Mortality Status" |
| UCOD\_113 | “CD-10 Underlying Cause of Death 113 Groups Recode (all years)” |
| UCOD\_358 | “CD-10 Underlying Cause of Death 358 Groups Recode (availa- ble 1999 forward” |
| UCOD\_39 | “CD-10 Underlying Cause of Death 39 Groups Recode (available 1999 forward)” |
| DIABETES | "Diabetes flag from Multiple Cause of Death" |
| HYPTERTEN | "Hypertension flag from Multiple Cause of Death" |
| permth\_int | "Number of Person-Months of Follow-up from NHANES Inter- view date" |
| permth\_exm | "Number of Person-Months of Follow-up from NHANES Mobile Examination Center (MEC) Date" |
| DODMONTH | “Month of Death” |
| DODYEAR | “Year of Death” |
| MECMONTH | “Month of MEC Exam” |
| MECYEAR | “Year of MEC Exam” |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **NHANES Geocodes** | | | | | | | | | | | |
| **GEO\_2000** |  |  |  |  |  |  |  |  |  |  |  |
| Variable Name | Variable description | 1999-  2000 | 2001-  2002 | 2003-  2004 | 2005-  2006 | 2007-  2008 | 2009-  2010 | 2011-  2012 | 2013-  2014 | 2015-  2016 | 2017-  2018 |
| SEQN \* | “NHANES Respondent Sequence Number ” (merge to public use data) | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| STATE2K \* | **“**Census 2000 FIPS State Code (2-digit numeric with leading zeros significant)” | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| CNTY2K \* | **“**Census 2000 FIPS County Code (3-digit numeric with leading zeros significant)” | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| **\* Merge variable** | |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Public Use Variables: NHIS 1999-2013** | | | | | | | | | | | | | | | | |
| **Public - Sample Adult File** | | | | | | | | | | | | | | | | |
| Variable Name | Variable description | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
| SRVY\_YR | NHIS Survey year | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| HHX \* | Household serial number | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| FMX \* | Family serial number | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| PX \* | Person serial number | X | X | X | X | X |  |  |  |  |  |  |  |  |  |  |
| FPX \* | Person serial number |  |  |  |  |  | X | X | X | X | X | X | X | X | X | X |
| HOPELESS | How often felt hopeless, past 30 days | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| WORTHLS | How often felt worthless past 30 days | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| **Restricted Geocodes - In-House files for Sample Adult** | | | | | | | | | | | | | | | | |
| Variable Name | Variable description | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
| PUBLICID | Merge variable | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| STATE | State of residence | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| COUNTY | County of residence | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| \* Merge variables (used to create PUBLICID) | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Proposal Format Updated 07/2020

1. *Literature References: Please provide a list of up to 10 references that are relevant to your project.*

Case, Anne, and Angus Deaton. 2015. “Rising Morbidity and Mortality in Midlife among White Non- Hispanic Americans in the 21st Century.” *Proceedings of the National Academy of Sciences* 112, no. 49: 15078–83.

Case, Anne, and Angus Deaton. 2017. “Mortality and Morbidity in the 21st Century.” *Brookings Papers on Economic Activity,* 397-443.

Case, Anne, and Angus Deaton. 2017. *Deaths of Despair and the Future of Capitalism,* Princeton University Press, 2020.

Kaiser Family Foundation, 2019. *Employer Health Benefits: 2019 Annual Survey,* Washington, D.C.: Kaiser Family Foundation.

Pierce, Justin R., and Peter K. Schott. 2020. “Trade Liberalization and Mortality: Evidence from U.S. Counties.” *American Economic Review: Insights,* 2(1), March, 47-64.

Ralph B. D’Agostino, Sr, PhD, Ramachandran S. Vasan, MD, Michael J. Pencina, PhD, Philip A. Wolf, MD, Mark Cobain, PhD, Joseph M. Massaro, PhD, and William B. Kannel, MD. 2008. General “Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study.” *American Heart Association*, 743-753.

1. *Resumes/C.V.: Please include a 2-page C.V. for each member of the research team (not as attachments).*

### David M. Cutler

Department of Economics, Harvard University 226 Littauer Center - 1805 Cambridge Street Cambridge, MA 02138

Phone: (617) 496-5216

[dcutler@harvard.edu](mailto:dcutler@fas.harvard.edu) [Website](https://scholar.harvard.edu/cutler)

# Employment

2014-2019: Harvard College Professor, Harvard University

2005-: Otto Eckstein Professor of Applied Economics, Department of Economics and Kennedy School of Government, Harvard University

2003-2008: Social Sciences Dean, Faculty of Arts and Sciences, Harvard University

1997-2005: Professor of Economics, Department of Economics and Kennedy School of Government, Harvard University

1995-1997: John L. Loeb Associate Professor of Social Sciences, Harvard University

1993: On leave as Senior Staff Economist, Council of Economic Advisers and Director, National Economic Council

1991-1995: Assistant Professor of Economics, Harvard University

***Other Affiliations*** Member[, Institute of Medicine](http://www.iom.edu/) Member, [National Academy of Social Insurance](http://www.nasi.org/)

Research Associate[, National Bureau of Economic Research,](http://www.nber.org/) Aging, Health Care, Public Economics, and Productivity programs

# Honors and Awards

2006: American Society of Health Economists Medal, Outstanding Health Economist Age 40 and Under 2004: David Kershaw Prize, Association for Public Policy and Management

2004: John Eisenberg Mentoring Award, Agency for Health Care Quality and Research

2003: Eugene Garfield Award, Research!America, for ["The Return to Biomedical Research: Treatment](http://www.press.uchicago.edu/cgi-bin/hfs.cgi/00/15531.ctl) [and Behavioral Effects"](http://www.press.uchicago.edu/cgi-bin/hfs.cgi/00/15531.ctl)

2000: Kenneth Arrow Award, Best Paper in Health Economics, for ["How Does Managed Care Do It?"](http://www.economics.harvard.edu/faculty/cutler/files/How%20Does%20Managed%20Care%20Do%20It.pdf)

1999: Griliches Prize, best paper in *Quarterly Journal of Economics*, for “Are Medical Prices Declining?”

1999: Outstanding Mentor Award, Harvard University Graduate School of Arts and Sciences

# Education

Ph.D. (Economics), M.I.T., September 1991.

A.B. (Economics, Summa Cum Laude) Harvard, 1987.

# Professional Service

President-elect, American Society for Health Economists

Former Member, Board of Directors, International Health Economics Association Former Editor, *Journal of Health Economics*

# Books

[*Your Money or Your Life: Strong Medicine for America's Health Care System*](http://www.oup-usa.org/isbn/0195160428.html), Oxford University Press, 2004.

*The Quality Cure: How Focusing on Health Care Quality Can Save Your Life and Lower Spending Too,*

University of California Press, 2014.

# Selected Articles

“The Association Between Income and Life Expectancy in the United States, 2001-2014,” *JAMA*, 315(16), April 2016, 1750-66 (with Raj Chetty, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, and Augustin Bergeron).

“When Does Education Matter? The Protective Effect of Education for Cohorts Graduating in Bad Times,” *Social Science & Medicine*, 127, February 2015, 63-73 (with Wei Huang and Adriana Lleras- Muney).

“Education And Health: Insights from International Comparisons” in Anthony J. Culyer, ed.,

*Encyclopedia of Health Economics*, Vol 1. San Diego: Elsevier; 2014. pp. 232-245.

“More Americans Living Longer with Cardiovascular Disease Will Increase Costs while Lowering Quality of Life,” *Health Affairs*, 32(10), October 2013, 1706-1714 (with Ankur Pandya, Thomas A. Gaziano, and Milton C. Weinstein).

“Rising Educational Gradients in Mortality: The Role of Behavioral Risk Factors” *Journal of Health Economics*, 30(6), December 2011, 1174-87 (with Fabian Lange, Ellen Meara, Seth Richards-Shubik, and Christopher J. Ruhm).

“Understanding Differences in Health Behavior by Education” *Journal of Health Economics*, 29(1), January 2010, 1-28 (with Adriana Lleras-Muney).

[“Value of Medical Innovation in the United States: 1960-2000,"](http://content.nejm.org/cgi/content/abstract/355/9/920) *New England Journal of Medicine*, (355), 2006, 920-927 (with Allison B. Rosen and Sandeep Vijan).

### Bryan Kim

National Bureau of Economic Research 1050 Massachusetts Ave, Cambridge, MA 02138

Phone: (949) 274-2363

[kimcwa@bc.edu](mailto:kimcwa@bc.edu)

# Employment and Experience

September 2020 – Present: Research Assistant, National Bureau of Economic Research

Using Stata and R to work with NHANES data from the National Center for Health Statistics.

June 2020 – August 2020: Research Intern, Federal Reserve Bank of Dallas

Used Stata to build a dataset with confidential mortgage data from the Federal Housing Administration.

June 2020 – August 2020: Research Assistant, MIT

June 2019 – August 2019: Data Analysis Intern, Center for Health Information and Analysis

Analyzed and processed Massachusetts claims and provider data for the Annual Report on the Performance of the Massachusetts Health Care System

# Education

A.B. (Mathematics and Economics) Boston College, 2022 (Expected)

# Technical Skills

## Programming Languages: R, Python, Swift.

Computing & Statistics: Stata, Mathematica, Matlab, Tableau. Other: Github, SQL.