Distribution Agreement

Signature:

Date:
April 27th, 2020

Sydney Brady

Potential overestimation of the influence of unmeasured confounding in non-randomized studies: a comparison of bias-analysis methods

By

Sydney Brady
Master of Public Health

Department of Epidemiology

Timothy L. Lash, DSc, MPH

Committee Chair

Potential overestimation of the influence of unmeasured confounding in non-randomized studies: a comparison of bias-analysis methods

By

Sydney L. Brady

B.S., Clemson University, 2017

Thesis Committee Chair: Timothy L. Lash, DSc, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in the Department of Epidemiology
2020

Abstract

Potential overestimation of the influence of unmeasured confounding in non-randomized studies: a comparison of bias-analysis methods

By Sydney L. Brady

Bias is a common concern among non-randomized public health studies. Thus, it is important to accurately estimate the magnitude of bias affecting a study. There are several methods that have been used to quantify the degree of bias, such as calculating an E-value or producing bias-adjusted estimates of association. We compared different approaches for addressing uncontrolled confounding to estimates of association obtained from the systematic removal of potential confounders in an analysis of the association between hypertension and cardiovascular disease (CVD) mortality.

We studied 16,220 National Health and Nutrition Examination Study (NHANES III) participants who were successfully linked to the 2015 Public-use Linked Mortality File. The final study population excluded those with prior confirmed history of CVD. We fit Cox regression models to estimate the association between hypertension and CVD mortality. Confounders we considered were age, race/ethnicity, sex, BMI, alcohol use, diet, hypercholesterolemia, health insurance, education, diabetes, exercise, tobacco use, and household income. Each confounder assessment was fit with a separate Cox model.

The crude association between hypertension and CVD mortality was HR=10.83, whereas the fully-adjusted association was HR=1.03. The ratio of single confounder-adjusted HRs to fully adjusted HRs ranged from 0.96 to 2.48. The E-value for the fully-adjusted model was 1.22 and the relative risk due to confounding computed by bias analysis methods ranged from 0.93 to 2.40.

The crude association between hypertension and CVD mortality was substantially confounded. Bias adjustment methods consistently overestimated the strength of confounding by each variable treated independently, presumably because these methods do not account for the covariance between adjustment variables. Use of E-values as bounds on confounding overestimated the actual strength of confounding by all adjustment variables.

Potential overestimation of the influence of unmeasured confounding in non-randomized studies: a comparison of bias-analysis methods

By

Sydney L. Brady

B.S., Clemson University, 2017

Thesis Committee Chair: Timothy L. Lash, DSc, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in Department of Epidemiology
2020

Introduction

Non-randomized studies are a fundamental part of epidemiology. A non-randomized study allows researchers to evaluate hypotheses about causation between exposures or risk factors and outcomes of interest without having to intervene in the observed population. Randomized control trials (RCT) are another study design that analyzes the relationship between exposures and outcomes. Unlike non-randomized studies, RCTs use the randomization process to allocate groups to an exposure or risk factor to study the impact of the exposure on the assigned group. This process of randomization creates an expectation that the comparison groups will have a similar distribution of potential confounding factors. Because all the groups, in expectation, have similar distribution of these factors, the amount of bias and confounding is reduced, thus increasing the study's internal validity.[1] However, RCTs are not always feasible or ethical. Additionally, RCTs have limited generalizability due to strict inclusion criteria and the specific demographic makeup of the population studied.[1] Therefore, non-randomized studies provide a more ethical and feasible way to examine the relationship between an exposure or risk factor and the outcome of interest.[1]

Because there is no randomization process in observational studies, the issue of uncontrolled confounding is more concerning than it would be in RCTs. Although factors comprising the population in observational studies cannot be changed, there are statistical methods that can account for the confounding attributable to the results of these studies. In order to accurately account for the exposure of interest in an observational study, a sufficient set of confounders must be controlled. However, unmeasured confounders may still exist even after controlling for all measured confounders, leading to residual bias.

Methods have been developed to help quantify the degree to which unmeasured confounders may influence a study's results. E-values and targeted bias analysis as well as other methods have been commonly used to calculate a bias-adjusted estimate of association to account for the unmeasured confounder.[2] These methods assume that the unmeasured confounder is independent of other measured confounders incorporated into the adjusted estimate, which may lead to an overestimation of the impact of unmeasured confounding on the reported results.[3]

Methods accounting for unmeasured confounding will be demonstrated and compared using an applied example with cardiovascular disease. Cardiovascular disease (CVD) is one of the leading causes of death in the United States.[4] There are many established risk factors that contribute to the development of CVD and subsequent CVD-specific mortality, including hypertension. To estimate the independent contribution of hypertension on the risk of CVD mortality, non-randomized studies must control for a sufficient set of confounders. With these potential confounders, there are many opportunities for bias due to unmeasured confounding to arise when estimating the association between hypertension and CVD-specific mortality. We examine the effect of hypertension on CVD-specific mortality in the National Health and Nutrition Examination Survey III (NHANES III) data linked to 2015 Public-Use Mortality Data. Our aim is to determine if quantitative bias analysis, used to address simulated unmeasured confounding, overestimates the true bias, thus helping to improve validity of unmeasured confounder calculations in non-randomized studies.

Methods

NHANES III Data and Study Design

The National Health and Nutrition Examination Study (NHANES III) was conducted between 1988 and 1994 by the National Center for Health Statistics of the Centers of Disease Control and Prevention. NHANES III uses a complex, multistage sample design that provides national estimates of health and nutritional status of the noninstitutionalized United States population. Survey participants received a detailed, in-person interview at home, followed by a physical examination at a mobile examination center.[5] Of the 33,994 total participants that completed the in-home survey portion of NHANES, 30,818 completed a physical examination at the medical examination center. Participants with a confirmed prior heart attack (N = 938), diagnosed congestive heart failure (N = 755), and those with history of stroke (N = 648), were excluded from analysis. After merging of NHANES III participants eligible for mortality follow-up, the final population for the comparative analyses included 16,220 records with 1,279 cardiovascular deaths.

2015 Public-Use Linked Mortality File

The National Center for Health Statistics (NCHS) has publicly available mortality data linked to the NHANES III survey. This public-use mortality data contains variables for adult participants only. The public-use linked mortality file provides mortality follow-up data from the date of survey participation through December 31, 2015. Participants with sufficient identifying data were eligible for mortality follow-up, and any participant that did not meet the minimum data requirements was considered ineligible for record linkage.

If a death occurred before 1999, the specific cause of death was classified by the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD-9) guidelines. If a death occurred after 1998, the specific cause of death was classified by the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD-10).[6] To account for all coded deaths during the survey mortality period, a death variable was created that recoded all deaths before 1999 originally coded under ICD-9 to be comparable to deaths coded under ICD-10. Any death coded as the primary cause of death occurring from heart disease was classified as the CVD-specific mortality variable.

Variables

Cardiovascular Disease (CVD) Mortality

Of the total recorded deaths in the final population, 1,279 were classified as a cardiovascular death, coded by ICD10. The person-months of follow-up time were constructed from the 2015 public-use file using the NHANES III interview month and year, and the month and year of death. For respondents assumed alive, person-months of follow-up were calculated from the NHANES III interview month and year to December 31, 2015.

Age

Age was divided into quintiles. Participants were binned to 18-26, 27-36, 37-50, 50-68, or older than 68 years old.

Race/Ethnicity

Race/ethnicity was divided into four groups, according to the NHANES classification. These groups were non-Hispanic white, non-Hispanic black, Mexican-American, and other.[7]

Hypercholesterolemia

Participants who had a total cholesterol level of 240 mg/dL or higher, or those who said they were currently taking medicine to control high cholesterol, were classified as having hypercholesterolemia according to the Centers for Disease Control (CDC) cholesterol guidelines.[8]

Body Mass Index (BMI)

The CDC BMI guidelines were used to create BMI categories. Height and weight were measured with a standardized protocol to calculate BMI (kg/m^2). Those with a BMI less than 18.5 kg/m^2 were classified as being underweight. Those with a BMI greater than or equal to 18.5 kg/m^2 and less than 25 were classified as being normal weight. Those with a BMI greater than or equal to 25.0 kg/m^2 and less than 30.0 kg/m^2 were classified as being overweight. Those with a BMI greater than or equal to 30.0 kg/m^2 were classified as being obese.[9]

Alcohol

Alcohol consumption was defined by the Substance Abuse and Mental Health Services Administration (SAMHSA) categories. Heavy drinkers were defined as binge drinking on 5 or more days in the past month, or more than 2 drinks per day. Moderate drinkers were defined as 1 to 2 drinks per day, and light or non-drinkers were defined as 1 or less than 1 drink per day.[10]

Tobacco use

Tobacco use was classified as those who self-identified as a non-smoker or former smoker, current smoker, or other (current use of smokeless tobacco, pipe, or cigars) at time of the in-home NHANES III survey.

Diet

The Healthy Eating Index (HEI) score was calculated from participants' responses to the 24-hour food recall survey within NHANES III. The HEI score takes into account fat, saturated fat, cholesterol, sodium, grain, fruit, vegetables, dairy, and meat. The HEI score is a standardized method to quantify how the diet conforms to the Dietary Guidelines. The higher the HEI score, the better the diet conforms to the guidelines.[11]

Hypertension

Hypertension was defined into three categories: normal, pre-hypertensive, and hypertensive as defined by the CDC guidelines.[12] Participants with a systolic value of 120 mmHG or less and a diastolic value of 80 mmHg or less were classified as having normal blood pressure. Participants with a systolic value of 150 mmHG and a diastolic value of 90 mmHg or higher were classified as having hypertension. Additionally, participants who were currently taking medicine to control hypertension were classified as hypertensive. Participants who were between the values of normal and hypertensive were classified as pre-hypertensive, and they were excluded from the analysis for this study.

Health Insurance

Health insurance was categorized as current enrollment in any type of health insurance coverage or currently having no health insurance coverage at time of interview.

Education

Education was categorized as less than high school (0-11 years), high school (12 years), and more than high school (13+ years), according to the analytic guidelines recommended by the National Center for Health Statistics on analyzing NHANES III variables.[7]

Diabetes

Diabetes was categorized as diabetic, pre-diabetic, or normal, according to the American Diabetic Association [13] categories. The Oral Glucose Tolerance Test (OGTT) was administered to participants to determine diabetes category. Those with blood sugar level of 200 mg/dl or higher were considered diabetic. Those with blood sugar level 140-199 mg/dl were considered pre-diabetic, and those with blood sugar level less than 140mg/dl were considered normal.

Household Income

Income was categorized as less than $10,000, $10,000 to $29,000, $30,000 to $49,000, and over $50,000 for total household income in a year, according to the analytic guidelines recommended by the National Center for Health Statistics on analyzing NHANES III variables.[7]

Exercise

Exercise was categorized by number of times per month the participant engaged in any physical activity. No activity, 1–4 times, 5–12 times, 13–30 times, and 31 times or more were the grouped categories for self-reported exercise.[14]

Statistical Analysis

All analyses were conducted in SAS 9.4. The Cox proportional hazards model was used to calculate the hazard ratio (HR) from time of NHANES III interview, in months, to the end of the follow-up period within the public-use 2015 mortality data on December 31st, 2015. The outcome of interest was death due to cardiovascular disease; the exposure of interest was hypertension versus no hypertension, and the potential confounders included age, sex, race/ethnicity, smoking, alcohol use, diabetes, exercise, household income, health insurance, BMI, educational attainment, diet, and hypercholesterolemia.

The final model included the following confounders: age, sex, race/ethnicity, smoking, alcohol use, diabetes, exercise, household income, BMI, educational attainment, diet, and hypercholesterolemia. The health insurance variable was excluded from the model due to evidence of collinearity with age. All remaining variables met the Cox proportional hazards assumptions. The equation to calculate the E-value was as follows:

E-value = HR + sqrt{HR*(HR-1)}[1]

Because the outcome of interest was relatively rare (<15%), the HR was used to approximate the RR for the E-value equation, in accordance with recommendations of the E-value authors.[1] For bias analysis, the following equation was used:

$$HR_{adj} = HR_{obs}\frac{HR_{CD}p_0+(1-p_0)}{HR_{CD}p_1+(1-p_1)}\ .$$

HR_{obs} is the observed HR associating hypertension with cardiovascular death without adjustment for the confounder. HR_{CD} is the HR associating the confounder with the disease, p_0 is the proportion of participants with the confounder among those without hypertension in the NHANES III data, and p_1 is the proportion of subjects with the confounder among those with hypertension in the NHANES III data. The resulting HR_{adj} is the adjusted HR associating hypertension with cardiovascular death, adjusted for the potential confounder.[3]

Results

Table 1:Unweighted population characteristics from NHANES III data merged with 2015 Public Use Mortality Data

Variable		**Overall population: N = 16,220**	**Participants with hypertension at NHANES survey: N = 2,707**
Age-mean (%)		46.28 (20)	62.74 (15)
Sex-no. (%)			
	Male	7,199 (44)	1,141 (42)
	Female	9,021 (56)	1,566 (58)
Race/ethnicity*-no. (%)			
	Non-hispanic white	6,754 (42)	1,310 (48)
	Non-hispanic black	4,444 (28)	917 (34)
	Mexican-American	4,378 (27)	412 (15)
	Other	644 (4)	68 (3)
BMI (kg/m2)-no.(%)			
	Underweight	385 (2)	32 (1)
	Normal	5,968 (37)	535 (20)
	Overweight	4,878 (30)	909 (34)
	Obese	3,486 (22)	907 (33)
	Missing	1,503 (9)	324 (12)
Alcohol-no. (%)			
	Non-drinker/light	3,093 (19)	402 (15)
	Moderate	1,115 (7)	126 (5)
	Heavy	2,192 (14)	229 (8)
	Missing	9,820 (60)	1950 (72)
Exercise, times per month-no. (%)			
	0	3,961 (24)	804 (30)
	1-4	3,136 (19)	530 (19)
	5-12	2,271 (14)	310 (11)
	13-30	6,020 (37)	939 (35)
	>31	475 (3)	70 (3)
	Missing	357 (3)	54 (2)
Tobacco use-no.(%)			
	Non-smoker or former smoker	3,381 (20)	792 (29)
	Current smoker	3,986 (25)	494 (18)
	Other tobacco products (Snuff, cigars, tobacco)	454 (3)	105 (4)
	Missing	8,399 (52)	1316 (49)
Education-no.(%)			
	Less than high school	6,476 (40)	1,320 (49)
	High school education	4,999 (31)	737 (27)
	More than high school	4,571 (28)	631 (23)
	Missing	174 (1)	19 (1)
Household Income-no. (%)			
	Less than $10,000	2,884 (18)	602 (22)
	$10,000-$29,000	6,551 (40)	1,084 (40)
	$30,000-$49,000	3,012 (19)	427 (16)
	Over $50,000	2,077 (13)	293 (11)
	Missing	1,696 (10)	301 (11)
Health Insurance-no. (%)			
	Any type of health insurance coverage	12,685 (78)	2,446 (90)
	No health insurance coverage	2,807 (17)	206 (8)
	Missing	728 (5)	55 (2)
Hypertension (mmHg)-no. (%)			
	Normal	7,736 (74)	.
	Hypertensive	2,707 (26)	.
HEI score-mean (std dev)		62.66 (13)	64.55 (14)
Mortality from CVD-no. (%)			
	Yes	1,279 (8)	460 (17)
	No	14,941 (92)	2,247 (83)
Diabetic-no. (%)			
	Normal	9,470 (58)	1,123 (41)
	Pre-diabetic	2,553 (16)	620 (23)
	Diabetic	1,073 (7)	432 (16)
	Missing	3,124 (19)	532 (20)
Hypercholesterolemia (mg/ll)-no. (%)			
	Hypercholesterolemia	2,787 (17)	811 (30)
	Normal	11,117 (69)	1,422 (52)
	Missing	2,316 (14)	474 (18)
*Race/ethnicity terminology and category recommended by NHANES III			

Table 2: Hazard ratios (HR) and E-values associating hypertension with CVD death with potential confounders removed

Observations used	**Variables in the model**	**HR**	**LL, 95% CI**	**UL, 95% CI**	**E-value**	**E-value LL 95% CI**	**E-value UL 95% CI**	**p-value**
1,942	Hypertension, age, race/ethnicity, sex, BMI, HEI score, household income, alcohol, hypercholesterolemia, education, diabetes, smoking, exercise	1.03	0.48	2.21	1.22	1.00	3.84	0.9305
1,981	Hypertension, age, race/ethnicity, sex, BMI, HEI score, household income, alcohol, hypercholesterolemia, education, diabetes, smoking	0.96	0.44	2.08	1.11	1.00	3.58	0.916
3,403	Hypertension, age, race/ethnicity, sex, BMI, HEI score, household income, alcohol, hypercholesterolemia, education, diabetes	1.24	0.70	2.20	1.78	1.00	3.82	0.457
3,605	Hypertension, age, race/ethnicity, sex, BMI, HEI score, household income, alcohol, hypercholesterolemia, education	1.28	0.73	2.26	1.89	1.00	3.95	0.3362
3,620	Hypertension, age, race/ethnicity, sex, BMI, HEI score, household income, alcohol, hypercholesterolemia	1.30	0.74	2.30	1.93	1.00	4.02	0.3007
3,783	Hypertension, age, race/ethnicity, sex, BMI, HEI score, household income, alcohol	1.40	0.80	2.43	2.14	1.00	4.29	0.2293
8,399	Hypertension, age, race/ethnicity, sex, BMI, HEI score, household income	2.28	1.73	3.01	3.99	2.85	5.48	<0.0001
9,221	Hypertension, age, race/ethnicity, sex, BMI, HEI score	2.22	1.68	2.93	3.86	2.74	5.31	<0.0001
9,671	Hypertension, age, race/ethnicity, sex, BMI	2.13	1.63	2.79	3.69	2.64	5.03	<0.0001
9,671	Hypertension, age, race/ethnicity, sex	2.15	1.64	2.83	3.72	2.65	5.10	<0.0001
9,671	Hypertension, age, race/ethnicity	2.18	1.66	2.85	3.78	2.71	5.15	<0.0001
9,671	Hypertension, age	2.15	1.64	2.83	3.73	2.66	5.10	<0.0001
9,671	Hypertension	10.83	8.64	13.59	21.16	16.76	26.66	<0.0001

Table 3: Bias adjusted HR and HR of each individual confounder

Confounder removed	**HR of each removed confounder, one at a time**	**Bias adjusted HR**	**Ratio between fully adjusted model and confounder removed**
Fully adjusted HR: 1.03	.	.	.
Age	2.48	1.58	2.40
Sex	1.05	2.42	1.02
Race/ethnicity	1.07	2.17	1.03
BMI	1.36	2.09	1.32
Alcohol	1.59	2.31	1.53
Exercise	0.96	0.95	0.93
Tobacco	1.34	1.69	1.30
Education	1.03	1.30	1.00
Household income	1.05	2.35	1.02
Diet	0.97	2.13	0.94
Hypercholesterolemia	0.97	1.18	0.94
Diabetes	1.03	1.15	1.00

The overall study poulation of the NHANES III dataset merged with the 2015 Public Use Mortality dataset had a mean age of 46.28 ± 20 years, was 44% male, and was 42% non-Hispanic white (Table 1). Among the overall population, 52% were obese or overweight, 24% reported no exercise the past month, 25% were current smokers at time of interview, 18% reported less than $10,000 as their yearly household income, 40% had less than a high school education, 26% had hypertension, 7% were considered diabetic, 17% had hypercholesterolemia, and 8% died from CVD (Table 1).

Of those who were diagnosed as being hypertensive at time of the survey, the mean age was 62.74 ± 15 years, 42% were male, and 48% were non-Hispanic white (Table 1). Among those who were hypertensive, 67% were obese or overweight, 30% reported no exercise the past month, 18% were current smokers at time of interview, 22% reported less than $10,000 as their yearly household income, 49% had less than a high school education, 16% were considered diabetic, 30% had hypercholesterolemia, and 17% died from CVD (Table 1). The group with hypertension had higher percentages of participants being overweight or obese, smoking, lower education, diabetic, and hypercholesterolemia.

The fully adjusted hazard ratio (HR) model included all potential confounders and was 1.03 (0.48, 2.21) (Table 2). The crude hazard ratio, including none of the potential confounders, was 10.83 (8.64, 13.59). As each confounder was removed from the fully adjusted model, the HR increased towards the crude. Additionally, more observations were used in the Cox proporitonal analysis as potential confounders were dropped from the model, because participants with missing values for the dropped confounder are then included. The calculated E-value for each HR was larger than the corresponding HR (Table 1). For example, in the model that contained hypertension and age, the HR was 2.15, and the E-value was 3.73 (Table 1).

Models in Table 3 contained all of the potential confounders except the specified confounder that was removed in the model. Each potential confounder was removed while keeping all of the other covariates in the model. The resulting HR was the HR of the model without the specified confounder. The highest HR occurred in the model with age removed, 2.48, and the lowest HR occurred with exercise removed, 0.96 (Table 3). The ratio between the fully adjusted HR containing all potential confouders was compared against each removed confounder. The highest ratio was between age and the fully adjusted model, 2.40, and the lowest ratio was between exercise and the fully adjusted model, 0.93 (Table 3).

Discussion

Unmeasured confounding is an important factor to be cognizant of when estimating the true effect measure between an exposure and an outcome. Using bias adjustment estimates described earlier as well as E-value estimations is one approach to account for the unmeasured confounders. However, through this applied example, these estimates demonstrated to have overestimated the magnitude of confounding when addressing the uncontrolled confounders.

The bias adjustment estimates consistently overestimated the strength of confounding by each variable treated independently, with age and exercise being an exception. These bias adjustment estimates do not account for the covariance between adjustment variables, and this may be one explanation for why the bias adjusted results are overestimated. Additionally, the use of E-values as bounds on confounding overestimated the actual strength of confounding by all adjustment variables.

The strengths of this analysis included the coverage and sample size provided by the NHANES III study design. This study design allowed for over sampling of underrepresented populations, helping to provide more data and better power in the modeling analyses. Another strength included the data availability; all data were collected and made publicly available, making this analysis more readily reproducible.

One limitation to this analysis was the presence of missing data. For variables alcohol and tobacco use, 60% and 52% of the data was missing, respectively. Likewise, diabetes data had 19% missing values, hypercholesterolemia had 14% missing, and household income information was 10% missing, as seen in Table 1. One explanation for the missing data could be explained by voluntary self-reporting of the data. Because habits such as alcohol and smoking have a negative connotation, participants feel social pressure to not admit to these habits.

Another source of missing data could come from misclassification or incorrect readings of biologic factors, such as hypercholesteremia and diabetes mentioned earlier. If the biologic measurement was refused or unable to be measured, the result was classified as a missing value. There may also be potential for additional unmeasured confounders that were not accounted for in this analysis.

Another limitation to this analysis was death certificate linkage to the participants. In order to successfully link to the correct NHANES III participant, the participant had to provide sufficient identifying data. Those who could not provide this were excluded from analyses, and this served as potential selection bias among the population who could not be linked.

Overall, these results show that adjustment for just one or two confounders often largely accounts for the confounding of an association. Uncontrolled confounders likely covary with the controlled confounders, so adding them to the model only slightly changes the strength of association. Nonetheless, bias analysis methods to adjust for unmeasured confounders treat the unmeasured variables as uncorrelated with the measured and controlled confounders, thus overestimating the bias due to uncontrolled confounding.

References

1. Lu CY. Observational studies: a review of study designs, challenges and strategies to reduce confounding. Int J Clin Pract. 2009; 63:691-7.

2. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017; 167:268-74.

3. Lash TL, Fox, M.P., and Fink, A.K. Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer; 2009.

4. CDC. Heart Disease Facts. 2019 [cited 2020 March 1st]; Available from: https://www.cdc.gov/heartdisease/facts.htm.

5. (DHHS) USDoHaHS. Third National Health and Nutrition Examination Survey, 1988-1994, NHANES III Laboratory Data File (CD-ROM). Public Use Data File Documentation Number 76200. In: Prevention CfDCa, editor. Hyattsville, MD: Available from National Technical Information Service (NTIS). Springfield, VA.
Acrobat. PDF format; includes access software: Adobe Systems, Inc. Acrobat Reader 2.1; 1996.

6. Wheatcroft G CC, Lochner KA. Comparative Analysis of the NHANES III Public-Use and Restricted-Use Linked Mortality Files. National Center for Health Statistics. 2007.

7. Statistics NCfH. Analytic and Reporting Guidlines: The
Third National Health and Nutrition Examination Survey, NHANES III (1988-94). In: Prevention CfDCa, editor. Hyattsville, Maryland: Centers for Disease Control and Prevention; 1996. p. 49.

8. CDC. High Cholesterol Facts. 2019.

9. CDC. Defining Adult Overweight and Obesity. 2017.

10. NIH. Drinking Levels Defined.

11. Statistics TNCfH. NHANES III Healthy Eating Index Data File, NHANES III, 1988-94. In: (CDC) CfDCaP, editor.2000.

12. CDC. Measuring Blood Pressure. 2019.

13. Association AD. Diabetes Diagnosis. Available from:https://www.diabetes.org/a1c/diagnosis.

14. Scragg R, Camargo CA, Jr. Frequency of leisure-time physical activity and serum 25-hydroxyvitamin D levels in the US population: results from the Third National Health and Nutrition Examination Survey. Am J Epidemiol. 2008; 168:577-86; discussion 87 -91.

15. VanderWeele TJ, Hernan MA, Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias. Epidemiology. 2008; 19:720-8.

16. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 2011; 22:42-52.

17. Ritchey MD, Wall HK, George MG, Wright JS. US trends in premature heart disease mortality over the past 50 years: Where do we go from here? Trends Cardiovasc Med. 2019.

18. Cohen HW, Hailpern SM, Alderman MH. Sodium intake and mortality follow-up in the Third National Health and Nutrition Examination Survey (NHANES III). J Gen Intern Med. 2008; 23:1297-302.

19. Gu Q, Burt VL, Paulose-Ram R, Yoon S, Gillum RF. High blood pressure and cardiovascular disease mortality risk among U.S. adults: the third National Health and Nutrition Examination Survey mortality follow-up study. Ann Epidemiol. 2008; 18:302-9.

20 . Inc. SI. The SURVEYPHREG Procedure. Cary, NC: SAS Institute Inc; 2017.

21. Prevention CfDCa. Specifying Weighting Parameters. 2013 [cited 2020 April 09 2020];Available from: https://www.cdc.gov/nchs/tutorials/nhanes/SurveyDesign/Weighting/intro_iii.htm.