
National Health Statistics Reports

Number 155 ■ May 24, 2021

Comparative Analysis of the National Health and Nutrition Examination Survey Public-use and Restricted-use Linked Mortality Files

by Lisa B. Mirel, Cindy Zhang, Christine S. Cox, Yeats Ye, Suad El Burai Félix, and Cordell Golden

Abstract

Objectives—Linking national survey data with administrative data sources enables researchers to conduct analyses that would not be possible with each data source alone. Recently, the Data Linkage Program at the National Center for Health Statistics (NCHS) released updated Linked Mortality Files, including the National Health and Nutrition Examination Survey data linked to the National Death Index mortality files. Two versions of the files were released: restricted-use files available through NCHS and Federal Statistical Research Data Centers and public-use files. To reduce the reidentification risk, statistical disclosure limitation methods were applied to the public-use files before they were released. This included limiting the amount of mortality information available and perturbing cause of death and follow-up time for select records.

Methods—To assess the comparability of the restricted-use and public-use files, relative hazard ratios for all-cause and cause-specific mortality using Cox proportional hazards models were estimated and compared.

Results—The comparative analysis found that the two data files yield similar descriptive and model results.

Keywords: survival analysis • hazard rate • data integration • data quality • National Death Index

Introduction

As the nation's principal health statistics agency, the mission of the National Center for Health Statistics (NCHS) is to provide statistical information that can be used to guide actions and policy to improve the health of the American people. In addition to collecting and disseminating the nation's official vital statistics, NCHS

conducts several population-based health surveys. Through the NCHS Data Linkage Program, data from these surveys have been linked to vital and other administrative data. The NCHS Data Linkage Program was established to maximize the scientific value of the data collected in the NCHS population-based surveys and to enable researchers to use longitudinal data from administrative databases or mortality data in

combination with survey data to examine factors that influence disability, health care utilization, morbidity, and mortality among different U.S. subpopulations.

NCHS recently completed an update for the National Health and Nutrition Examination Survey (NHANES) linkage to the National Death Index (NDI) to obtain mortality status and cause of death through December 31, 2015. To protect the confidentiality of the NHANES participants, restricted-use versions of the NHANES linked mortality files (LMFs) were made available only through the NCHS and Federal Statistical Research Data Centers (RDCs). To complement the restricted-use files and increase data access, NCHS developed public-use versions of the LMFs.

The public-use data release included the addition of perturbed data for two elements: date of death and underlying cause of death. Data perturbation is a privacy-protecting technique that relies on statistical methods to add random "noise" to confidential data (1). Perturbed (or synthetic) data for a subset of records were added to reduce reidentification risk to survey participants, maximizing the amount of mortality data included in the public-use release and limiting the amount of synthetic data introduced to the data file.



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics



This report describes a comparative analysis of the public-use and restricted-use NHANES LMFs. Cox proportional hazards models were used to estimate the relative hazard ratios for a standard set of sociodemographic covariates for all-cause and cause-specific mortality using the public-use and restricted-use NHANES LMFs. For each covariate, relative percent differences were calculated for the hazard ratios estimated with the two separate files. Then point estimates produced with the public-use LMF were compared with confidence interval estimates produced with the restricted-use LMF. NCHS conducted this comparative analysis to assess the comparability of the two versions of LMFs.

Methods

Description of continuous NHANES

NCHS has administered NHANES, a nationally representative survey consisting of about 5,000 persons from 15 different counties each year, continuously since 1999. (NHANES, starting with the 1999–2000 cycle, is called “continuous” because data collection is performed year-round but is released in 2-year cycles for confidentiality reasons.) The survey includes a standardized physical examination, laboratory tests, and questionnaires that cover various health-related topics. NHANES includes an interview in the household, followed by an examination in a mobile examination center (MEC). The NHANES sample represents the noninstitutionalized civilian U.S. population that is selected using a complex, multistage probability design (2). Data from the 1999 through 2014 NHANES cycles have been linked to mortality data through 2015.

Description of NHANES III

Prior to becoming a continuous survey, NHANES was conducted periodically, with the last periodic survey, NHANES III, conducted between 1988 and 1994. NHANES III was designed to provide national estimates of health and nutritional status of the U.S. civilian,

noninstitutionalized population aged 2 months and over. Like the continuous survey, NHANES III included a standardized physical examination, laboratory tests, and questionnaires that covered various health-related topics (3). The NHANES III data have also been linked to mortality data through 2015.

Description of NDI

NDI is a centralized database of death record information on file in jurisdictional vital records or statistics offices and maintained by NCHS (available from: <https://www.cdc.gov/nchs/ndi/index.htm>). These data can be used to identify each person who has died in the United States and U.S. military overseas and that person's cause(s) and manner of death. Deaths are categorized using the *International Classification of Diseases, Ninth Revision* (ICD–9) and, from 1999 onwards, the *International Classification of Diseases, 10th Revision* (ICD–10) for underlying and multiple causes of death.

Description of NHANES linked mortality data

NCHS' Research Ethics Review Board (ERB) approved the linkage of these files. The NCHS Research ERB, also known as an Institutional Review Board or IRB, is an administrative body of scientists and nonscientists that was established to protect the rights and welfare of human research subjects. Mortality status for eligible NHANES participants was determined primarily through probabilistic record matching with NDI. NHANES participants were considered eligible for mortality follow-up if they provided sufficient identifying information at the time of interview or MEC follow-up. Each participant's survey record was screened to determine if it contained at least one of the following combinations of identifying data elements:

- Social Security Number (SSN) (nine digits or last four digits), last name, first name
- SSN (nine digits or last four digits), sex, month of birth, day of birth, year of birth

- Last name, first name, month of birth, year of birth

Any survey participant record that did not meet the minimum data requirements was ineligible for record linkage. On average across all survey cycles, about 99.9% of survey participants were eligible for the mortality record linkage. For this report, the continuous NHANES 1999–2014 data were split into two analytic cohorts (1999–2006 and 2007–2014) due to the number of cause-of-death categories available on their respective public-use LMFs. NCHS' Disclosure Review Board approved the release of the privacy-protected public-use file.

For a complete description of the matching methodology, see: https://www.cdc.gov/nchs/data/datalinkage/LMF2015_Methodology_Analytic_Considerations.pdf (4).

Restricted-use LMFs

The restricted-use LMFs include detailed mortality information for all eligible survey participants, including children. The restricted-use files include the following variables: survey participant eligibility status, mortality status, age at death, age last known alive, date of death (month, day, and year), underlying and multiple causes of death, date of birth (month, day, and year), and NHANES interview and examination dates (month, day, and year). Additional information obtained from the death certificate like literal text and match status indicators can be accessed by researchers within an RDC.

Public-use LMFs

Due to confidentiality protections, the public-use files include only eligible survey participants aged 18 years and over at the time of the interview and a limited set of mortality variables. In addition, the public-use versions of the mortality variables were subjected to data perturbation techniques to reduce the risk of participant reidentification. Synthetic data were substituted for the actual date of death and underlying cause-of-death data for selected decedent records. Vital status (e.g., whether the

person was deceased by the end of the follow-up) was not perturbed. Variables provided in the public-use NHANES LMFs include survey participant eligibility status, vital status, person months (an estimate of the person's time at risk in months) of follow-up, and the top nine leading causes of death based on "Deaths: Leading Causes for 2015" (5) for NHANES III and NHANES 1999–2006. The top nine causes of death included diseases of the heart; malignant neoplasms (cancer); chronic lower respiratory disease; cerebrovascular disease; accidents (unintentional injury); Alzheimer disease; diabetes mellitus; nephritis, nephrotic syndrome and nephrosis; and influenza and pneumonia. All other causes of death were grouped together and placed in a residual category. For NHANES 2007–2014, only two causes of death—diseases of the heart and malignant neoplasms (cancer)—and an all other causes residual category were released in the public-use LMF due to an increased disclosure risk for those survey cycles. In addition, two variables were created and provided in all versions of the NHANES LMFs to indicate the presence of diabetes or hypertension in the multiple cause-of-death codes, because these conditions are generally reported as contributing, rather than underlying, causes of death. No additional death certificate information is available in the public-use files. A summary of the two files is presented in [Table 1](#).

Reidentification risk assessment and data perturbation

Similar to previous files, reidentification risk for deceased participants was assessed by combining different public-use sources. The names of these public data sources are intentionally omitted here to reduce reidentification risk. For NHANES decedent records, public-use survey data and the proposed public-use LMF were matched on sociodemographic variables to existing publicly available data sources. This match identified potential records that were at increased risk of reidentification (6). For each publicly available data source, unique records were identified based on the combination

of available data elements and then compared with the unique records identified from each data source. All NHANES decedent records that were correctly matched to unique records in these public data sources were considered to be at risk of reidentification. After identifying the NHANES decedent records at increased risk of reidentification, a data perturbation plan was created to reduce the risk of reidentification and allow for the release of an NHANES linked mortality public-use file. All NHANES decedent records considered at risk of reidentification were subject to data perturbation and were randomly assigned to have either the date of death or the underlying cause of death perturbed. Vital status (e.g., whether the person was deceased by the end of follow-up) was not perturbed. Because the linkage was based on probabilistic techniques, those who linked were assumed deceased and those who did not were assumed alive.

Sample selection

For this report, NHANES III, NHANES 1999–2006, and NHANES 2007–2014 were analyzed as separate cohorts. The continuous NHANES 1999–2014 data were split into two cohorts due to the difference in the number of causes of death available on the public-use files. As noted above, the public-use NHANES 1999–2006 LMFs included nine causes of death; however, only two causes were provided in the public-use NHANES 2007–2014 LMFs. To compare results obtained from the restricted-use and public-use data sets, the public-use Household Adults file for NHANES III (1988–1994) and the public-use Demographic files for continuous NHANES 1999–2006 and 2007–2014 were merged with the corresponding public-use and restricted-use mortality files, respectively, to create the analytic samples. All analyses were limited to those eligible for mortality follow-up who were at least age 25 at the time of the NHANES interview; non-Hispanic white, non-Hispanic black, Mexican American (for NHANES III and NHANES 1999–2006), or Hispanic (for NHANES 2007–2014); and had

no missing values for education level, marital status, or cause of death.

Outcome measurement

All-cause and cause-specific mortality in the public-use and restricted-use NHANES LMFs were examined. Follow-up time was constructed in two ways, depending on the file. The public-use files contain less detailed date information (person months of follow-up only) than the restricted-use files and include perturbed information for person months for selected decedents, which creates slight differences when comparing the duration of follow-up calculations between the two files. For the public-use files, duration of follow-up was based on person months of follow-up from the date of the interview. For those who were assumed deceased, the person-month variable was calculated from the date of the interview to the date of death. For those who were assumed alive, the person-month variable was calculated as the date of the interview to the end of the follow-up period, which was December 31, 2015. For the restricted-use files, duration of follow-up was calculated using complete information on the month, day, and year of the NHANES interview and the month, day, and year of death. For participants assumed alive, the follow-up period ended on December 31, 2015.

In addition to all-cause mortality, nine leading causes of death in the United States in 2015 were examined for NHANES III and NHANES 1999–2006 (7). The ICD codes used to categorize the nine causes of death are documented elsewhere (8). For NHANES 2007–2014, only all-cause mortality, diseases of the heart, and cancer mortality were examined. The NHANES LMFs include both ICD–9 and ICD–10 cause-of-death coding for all U.S. deaths. To maintain the same cause-of-death codes across all years in the study period, the leading causes of death were based on the ICD–10 underlying cause-of-death 113 group recode, which recodes all deaths occurring prior to 1999 into ICD–10 codes (9). Although the code numbers are the same for all years of mortality data, the coding rules for determining underlying cause of death differ for

deaths that occurred prior to 1999 under ICD–9 codes (9) and those that occurred in later years under ICD–10. The analyses presented in this report do not control for the transition in coding rules between ICD–9 and ICD–10 because that transition does not affect the comparisons of interest in this report.

Covariates

All models included a standard set of sociodemographic characteristics as reported at the time of the NHANES interview: age (in continuous years), sex, race and ethnicity (non-Hispanic white, non-Hispanic black, Mexican American [NHANES III and NHANES 1999–2006], or Hispanic [NHANES 2007–2014]), educational attainment (less than high school, high school diploma or GED, or some college and above), and marital status (married or not married). If survey participants responded, “don’t know,” refused to answer, or had missing data for any of these covariates, they were excluded from the analytic sample. These criteria reduced the eligible sample for the three cohorts by 5.2%, 10.7%, and 10.1% for NHANES III, NHANES 1999–2006, and NHANES 2007–2014, respectively.

Data analysis

Comparisons of estimated relative risks between the public-use and restricted-use files were conducted within each defined cohort. Comparisons across cohorts were not conducted. Cox proportional hazards models were used to estimate the relative risk for the covariates for all-cause and cause-specific mortality. Cox proportional hazards models are used to assess the association of survival times and covariates (10). All relative risk estimates were calculated with the survival procedure in SAS 9.4 callable SUDAAN to account for the complex survey design of NHANES (11). The Efron method was used to handle tied failure times (12).

For all estimates, a relative difference was calculated as the value of the difference between the restricted-use and public-use estimates divided by the restricted-use estimate. This difference was then multiplied by 100 to compute

the percent relative difference. All estimates were assessed by the magnitude of the percent relative difference. The following criteria were used to assess similarities and differences in the hazard ratios for the selected sociodemographic and health characteristics between the public-use and restricted-use files. This method is based on criteria used in a previous report (13). If the hazard ratio for each category (e.g., male or female) of a characteristic (e.g., sex) was equal for both files, it was described as being the same.

1. If the relative difference between the restricted-use and public-use hazard ratios was between 0.0 to 10% for all categories of the characteristic, the estimate was described as being “consistent or similar” across the two files. The relative difference could be positive or negative.
2. If the relative difference between the hazard ratios for at least one category of the characteristic was greater than 10%, the estimates were described as being “different (higher/more or lower/less)” with respect to that characteristic.

In addition, estimates from the public-use file were assessed to see if they fell within the confidence interval of the estimate from the restricted-use file. Because the estimates from the two files were not statistically independent, and measures of correlation between the two populations were not readily available, comparisons were not statistically tested.

Descriptive statistics for the analytic samples are presented in [Table 2](#). The percent distribution of cause of death for those who were assumed deceased is presented in [Table 3](#). Hazard ratios and relative percent difference between the restricted-use and public-use files for all-cause mortality and cause-specific mortality are presented in [Tables 4–13](#).

Results

[Table 2](#) shows the unweighted sample counts (*n*), weighted percent or means, and standard errors for the covariates used in the analyses. The descriptive statistics for covariates in [Table 2](#) are the same for both the public-use and restricted-use files, as

these variables were not subject to data perturbation. In all subsequent analyses, any differences between the results for the two files were associated with cause and date of death because those are the variables that were subject to perturbation.

The average age of the analytic sample at baseline was 47.7 years in NHANES III, 49.3 in NHANES 1999–2006, and 50.1 in NHANES 2007–2014. For all three cohorts, about 52% of the analytic sample was female, and the majority were non-Hispanic white (83.7% in NHANES III, 80.4% in NHANES 1999–2006, and 73.7% in NHANES 2007–2014). The modal educational category was some college and above (41.3% in NHANES III, 54.8% in NHANES 1999–2006, and 59.1% in NHANES 2007–2014), and the majority of the sample was married at the time of the NHANES interview (65.9% in NHANES III, 62.1% in NHANES 1999–2006, and 59.1% in NHANES 2007–2014).

The number and weighted percentage of persons in the analytic sample who were identified as deceased (33.7% in NHANES III, 16.4% in NHANES 1999–2006, and 5.5% in NHANES 2007–2014) were identical in the public-use and restricted-use files, because vital status of individuals was not changed in the perturbation process. For NHANES III, NHANES 1999–2006, and NHANES 2007–2014, the mean person month of follow-up (weighted) between the public-use and restricted-use file was similar (242.8 compared with 242.9 in NHANES III, 141.1 compared with 140.6 in NHANES 1999–2006, and 59.9 compared with 57.9 in NHANES 2007–2014 [data not shown]).

[Table 3](#) shows the cause-specific percent distributions for the nine leading causes of death studied. Overall, the distributions were similar when comparing the two files. The distributions of causes of death were similar between the public-use and restricted-use files for NHANES III, where the percent differences were all under 1.5%. For NHANES 1999–2006, the distribution of causes of death was similar for all causes of death. Note that the relative percent difference was less than 10% in all comparisons. Due to limited sample

sizes in NHANES 2007–2014, only deaths from diseases of the heart and cancer were available as causes of death on the public-use file. The distribution of these two causes was similar in the two files. Note that the rankings of each cause of death remained the same within the restricted-use and public-use files for each cohort.

All-cause mortality model results

Table 4 displays hazard ratios from two Cox proportional hazards models of all-cause mortality: one estimated from the public-use files and one estimated from the restricted-use files for all three cohorts. Differences between restricted-use and public-use derived estimates were evaluated separately for each cohort but not across cohorts. While fact of death (i.e., vital status) was not changed in the public-use files, differences were observed in the duration of follow-up (survival time) between the public-use and restricted files due to the perturbation of date of death for selected decedents in the public-use files.

Nevertheless, the results of all-cause mortality models were consistent in showing an association between risk for mortality and age, sex, education, and marital status for all three cohorts. For example, males, persons with less than a high school education or a high school diploma or GED only, and never married individuals had higher risks of mortality compared with their respective comparison groups in both the public-use and restricted-use models across all three cohorts, although the magnitude of the hazard ratios varied somewhat across cohorts. All estimates have a relative percent difference of less than 5%, and the majority were under 1%. Variations were observed among the estimates, according to the cohort of NHANES, which may be due to the differing follow-up times of the cohorts. For each covariate, all point estimates from the public-use files fell within the confidence interval estimates from the restricted-use files. Moreover, the magnitude of the relative risks and the width of the 95% confidence intervals were similar for estimates from the public-use and restricted-use files.

Cause-specific mortality model results

Tables 5 and 6 show that the pattern of results for heart disease and cancer mortality was similar for all estimates in NHANES III and NHANES 1999–2006 (except for education in NHANES 1999–2006) between the public-use and restricted-use files. Also, the point estimates from the public-use files fell within the confidence interval estimates of the restricted-use files. For NHANES 2007–2014, similar estimates were noted between the public-use and restricted-use files for age, race and ethnicity, and marital status for both heart disease and cancer. However, differences were noted in this cohort for sex and education. This was likely due to the lower number of deaths for this cohort and greater variation due to the perturbation process. Nevertheless, all estimates from the NHANES 2007–2014 public-use models fell within the confidence intervals of the restricted-use files.

Tables 7–13 display the results of Cox proportional hazards models with the seven other specific underlying causes of death as an outcome for NHANES III and NHANES 1999–2006. Data from NHANES 2007–2014 include two causes of death (heart disease and cancer) only and were not presented for the additional causes of death in Tables 7–13. Each cause-specific table provides a comparison of the model results from the public-use version and the restricted-use version of the files. Overall, the hazard ratios for the variables in the models from the two files presented similar results, except for education and marital status (for accidents [unintentional injury] only) in the following NHANES 1999–2006 models: chronic lower respiratory disease; accidents (unintentional injury); Alzheimer disease; and nephritis, nephrotic syndrome and nephrosis disease. The relative percent difference between the public-use and restricted-use estimates tended to be smaller for NHANES III than for NHANES 1999–2006. For all specific causes of death, the point estimates from the public-use model fell within the confidence intervals for the restricted-use models. However, for select covariates in the public-use compared

with restricted-use models for accidents (unintentional injury); diabetes; nephritis, nephrotic syndrome and nephrosis disease; and influenza and pneumonia mortality, the statistical significance of an estimate differed between the two files. For example, as noted in Table 12, for NHANES 1999–2006 those with less than a high school education were twice as likely to die from nephritis, nephrotic syndrome and nephrosis disease compared with those with an education attainment of some college or above. However, the hazard ratio (HR) was statistically significant in the restricted-use data model (HR = 2.1, 95% confidence interval [CI]: 1.01–4.31) but not in the public-use data model (HR = 1.95, 95% CI: 0.95–3.98).

Discussion

This report describes the results obtained from the comparison of public-use and restricted-use NHANES LMF analyses for three analytic cohorts, with mortality follow-up through 2015. This report supports earlier findings comparing the public-use and restricted-use LMFs for the National Health Interview Survey (6,14). In the public-use version of the data files, a limited amount of information for selected decedents was perturbed. Further, less detailed date of death information is included in the public-use version compared with the restricted-use files, where no information has been perturbed, and complete information is included on date of death.

The comparative analysis found that the two data files yielded similar descriptive and model results when examining all-cause mortality. Because the perturbation process in the public-use files did not affect the vital status of any individuals in the files, differences in results between the two files when examining overall (all-cause) mortality arose because the public-use files included perturbed information for date of death and less specific information regarding date of death for individuals compared with the restricted-use files. The comparisons for all-cause mortality across a standard set of sociodemographic characteristics between the public- and restricted-use LMFs showed similar

results, based on the criteria used to evaluate results.

The comparative analysis of cause-specific mortality across the public-use and restricted-use versions of the NHANES LMFs also yielded consistent model results, even for less common causes of death such as diabetes; nephritis, nephrotic syndrome and nephrosis; and influenza and pneumonia, which each represented 3% or less of deaths for the two cohorts with those causes available on the public-use LMFs. However, differences in the statistical significance of certain covariates may be observed when working with the privacy-protected public-use files as opposed to the restricted-use files. As presented in this report, for certain causes of death, some instances were noted where a given covariate did not remain statistically significant between the public-use compared with restricted-use LMF models. This was likely a function of the more recent cohorts having fewer deaths and being more affected by the perturbation process.

These findings offer a resource for researchers who are considering using the most recent public-use LMFs to study mortality among linkage-eligible adult NHANES participants. However, all potential users are urged to exercise caution in using the public-use files when examining the mortality patterns of small subgroups of the population, such as numerically small racial or ethnic minority groups, very old individuals, or young adults. This is particularly true when cause-specific analyses of such numerically small demographic subgroups are performed in the more recent NHANES cycles, where overall lower numbers of decedents exist due to shorter follow-up periods. If smaller subgroup analyses are desired, the restricted-use files are available through the NCHS RDC.

In summary, the release of a public-use version of the NHANES LMFs provides the public health, social science, demographic, and medical communities with a large data set that is readily available, nationally representative, and rich in detail for both all-cause and cause-specific mortality outcomes. The public-use files may serve as a resource for researchers and policymakers in further

understanding adult mortality trends and patterns.

References

1. Wilson RL, Rosen PA. Protecting data through ‘perturbation’ techniques: The impact on knowledge discovery in databases. *J Database Manag* 14(2):14–26. 2003.
2. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: Sample design, 2011–2014. National Center for Health Statistics. *Vital Health Stat* 2(162). 2014.
3. Ezzati TM, Massey JT, Waksberg J, Chu A, Maurer KR. Sample design: Third National Health and Nutrition Examination Survey. National Center for Health Statistics. *Vital Health Stat* 2(113). 1992.
4. National Center for Health Statistics. The linkage of National Center for Health Statistics survey data to the National Death Index—2015 linked mortality file (LMF): Methodology overview and analytic considerations. Hyattsville, MD. 2019. Available from: https://www.cdc.gov/nchs/data/datalinkage/LMF2015_Methodology_Analytic_Considerations.pdf.
5. Heron M. Deaths: Leading causes for 2015. National Vital Statistics Reports; vol 66 no 5. Hyattsville, MD: National Center for Health Statistics. 2017.
6. Lochner K, Hummer RA, Bartee S, Wheatcroft G, Cox C. The public-use National Health Interview Survey linked mortality files: Methods of reidentification risk avoidance and comparative analysis. *Am J Epidemiol* 168(3):336–44. 2008.
7. Murphy SL, Xu JQ, Kochanek KD, Curtin SC, Arias E. Deaths: Final data for 2015. National Vital Statistics Reports; vol 66 no 6. Hyattsville, MD: National Center for Health Statistics. 2017.
8. National Center for Health Statistics. Public-use 2015 LMF data dictionary. Hyattsville, MD. 2019. Available from: <https://www.cdc.gov/nchs/data/datalinkage/public-use-2015-linked-mortality-files-data-dictionary.pdf>.
9. Anderson RN, Miniño AM, Hoyert DL, Rosenberg HM. Comparability of cause of death between ICD–9 and ICD–10: Preliminary estimates. National Vital Statistics Reports; vol 49 no 2. Hyattsville, MD: National Center for Health Statistics. 2001.
10. Cox DR. Regression models and life-tables (with discussion). *J R Stat Soc Series B Stat Methodol* 34(2):187–202. 1972.
11. RTI International. SUDAAN (Release 11.0) [computer software]. 2012.
12. Hertz-Picciotto I, Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics*. 53(3):1151–6. 1997.
13. Lloyd PC, Helms VE, Simon AE, Golden C, Brittain J, Call E, et al. Linkage of 1999–2012 National Health Interview Survey and National Health and Nutrition Examination Survey data to U.S. Department of Housing and Urban Development administrative records. National Center for Health Statistics. *Vital Health Stat* 1(60). 2017.
14. Mirel LB, El Burai Felix S, Zhang C, Golden C, Cox CS. Comparative analysis of the National Health Interview Survey public-use and restricted-use linked mortality files. National Health Statistics Reports; no 143. Hyattsville, MD: National Center for Health Statistics. 2020.

Table 1. List of key differences between public-use and restricted-use linked mortality files

File detail	Restricted-use	Public-use
Dates	Exact date of death, birth date, and interview date	Person months of follow-up
Cause of death	Detailed UCOD and MCOD information	Nine UCODs grouped and recoded before 2007 and three UCODs grouped and recoded from 2007–2014. Two MCOD indicators: diabetes and hypertension
Participants.	Both adults and children	Adults aged 18 and over
Perturbed data on death.	No perturbation	Perturbed information: Cause of death for select decedents and follow-up time (vital status is not perturbed)

NOTES: UCOD is underlying cause of death. MCOD is multiple cause of death.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 2. Descriptive statistics for adults aged 25 and over: National Health and Nutrition Examination Survey III, National Health and Nutrition Examination Survey, 1999–2006, and National Health and Nutrition Examination Survey, 2007–2014, linked mortality files (follow-up through 2015)

Characteristic	NHANES III			NHANES 1999–2006			NHANES 2007–2014		
	Unweighted (n)	Weighted percent or mean	Standard error	Unweighted (n)	Weighted percent or mean	Standard error	Unweighted (n)	Weighted percent or mean	Standard error
Vital status: Assumed deceased.....	7,358	33.7	1.10	4,107	16.4	0.45	1,573	5.5	0.23
Age (mean).....	15,989	47.7	0.45	16,342	49.3	0.30	19,216	50.1	0.22
Age group:									
25–44.....	6,718	51.0	1.28	6,167	43.8	0.92	6,887	39.9	0.71
45–64.....	4,275	29.4	0.70	5,098	36.5	0.76	7,102	40.0	0.58
65–79.....	3,271	15.6	0.72	3,351	14.8	0.44	3,742	14.9	0.34
80 and over.....	1,725	4.0	0.33	1,726	4.9	0.25	1,485	5.2	0.24
Sex:									
Male.....	7,480	47.4	0.45	7,873	47.9	0.37	9,340	47.9	0.38
Female.....	8,509	52.6	0.45	8,469	52.1	0.37	9,876	52.1	0.38
Race and ethnicity:									
Non-Hispanic white.....	7,568	83.7	0.78	9,166	80.4	1.24	9,487	73.7	1.70
Non-Hispanic black.....	4,389	11.4	0.65	3,487	11.9	0.98	4,471	11.9	0.92
Mexican American.....	4,032	4.9	0.38	3,689	7.7	0.76	---	---	---
All Hispanic.....	---	---	---	---	---	---	5,258	14.3	1.28
Education level:									
Less than high school.....	6,791	24.6	0.99	5,184	18.9	0.71	5,429	18.2	0.79
High school diploma or GED.....	4,720	34.1	0.82	3,882	26.2	0.70	4,484	22.7	0.64
Some college and above.....	4,478	41.3	1.33	7,276	54.8	1.11	9,303	59.1	1.14
Marital status:									
Married.....	9,550	65.9	0.82	9,540	62.1	0.89	10,232	59.1	0.75
Not married.....	6,439	34.1	0.82	6,802	37.9	0.89	8,984	40.9	0.75

--- Data not available.

NOTE: NHANES is National Health and Nutrition Examination Survey.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 3. Distribution of cause of death among adults aged 25 and over who are assumed deceased: National Health and Nutrition Examination Survey III, National Health and Nutrition Examination Survey, 1999–2006, and National Health and Nutrition Examination Survey, 2007–2014, linked mortality files (follow-up through 2015)

Cause-specific deaths ¹	NHANES III (n = 7,358)					NHANES, 1999–2006 (n = 4,107)					NHANES, 2007–2014 (n = 1,573)				
	Public-use		Restricted-use		Relative percent difference	Public-use		Restricted-use		Relative percent difference	Public-use		Restricted-use		Relative percent difference
	Weighted percent	Standard error	Weighted percent	Standard error		Weighted percent	Standard error	Weighted percent	Standard error		Weighted percent	Standard error	Weighted percent	Standard error	
Diseases of the heart	22.9	0.70	22.6	0.72	–1.4	17.0	0.84	17.3	0.86	2.1	16.7	1.12	17.4	1.13	4.4
Cancer, all sites	24.0	0.80	24.3	0.83	1.2	22.9	0.96	24.3	0.99	5.8	23.6	1.47	25.7	1.59	8.2
Chronic lower respiratory disease	5.6	0.32	5.7	0.32	0.9	5.9	0.42	6.5	0.45	8.3	---	---	---
Accidents (unintentional injuries)	2.8	0.28	2.8	0.28	0.4	3.9	0.45	4.2	0.48	6.4	---	---	---
Cerebrovascular disease	6.1	0.41	6.2	0.41	1.5	5.1	0.40	5.3	0.43	4.7	---	---	---
Alzheimer disease	2.3	0.22	2.3	0.22	–0.4	3.6	0.31	3.7	0.30	4.4	---	---	---
Diabetes mellitus	2.8	0.34	2.8	0.34	0.4	2.8	0.31	2.9	0.33	4.6	---	---	---
Nephritis, nephrotic syndrome and nephrosis	1.1	0.17	1.1	0.17	0.9	2.2	0.27	2.3	0.28	5.1	---	---	---
Influenza and pneumonia	2.6	0.26	2.6	0.26	0.8	2.0	0.26	2.1	0.26	3.6	---	---	---

... Category not applicable (see Table 1).

--- Data not available.

¹Underlying cause-of-death codes are based on the *International Classification of Diseases, 10th Revision*, recoded into 113 selected causes. Weighted percentages for cause-specific deaths are based on the sample of decedents.

NOTE: NHANES is National Health and Nutrition Examination Survey.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 4. Hazard ratios for all-cause mortality: National Health and Nutrition Examination Survey III, National Health and Nutrition Examination Survey, 1999–2006, and National Health and Nutrition Examination Survey, 2007–2014, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							NHANES 2007–2014						
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use				Public-use			Restricted-use		Relative percent difference
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)		
Age, years	1.09	1.09	1.10	1.09	1.09	1.10	0.1	1.09	1.09	1.09	1.09	1.09	1.09	0.0	1.09	1.08	1.09	1.09	1.08	1.09	0.0
Sex (female):																					
Male	1.54	1.44	1.64	1.55	1.45	1.65	0.5	1.68	1.53	1.86	1.69	1.54	1.86	0.4	1.74	1.54	1.95	1.74	1.55	1.95	0.1
Race and ethnicity (non-Hispanic white):																					
Non-Hispanic black	1.16	1.05	1.28	1.18	1.08	1.29	1.6	1.09	0.98	1.21	1.07	0.96	1.20	-1.5	1.09	0.92	1.29	1.06	0.92	1.23	-2.6
Mexican American	0.98	0.88	1.09	0.97	0.87	1.09	-0.8	1.12	0.96	1.30	1.10	0.95	1.28	-1.7	---	---	---	---	---	---	---
All Hispanic	---	---	---	---	---	---	---	---	---	---	---	---	---	---	0.90	0.72	1.12	0.87	0.70	1.08	-3.7
Education (some college and above):																					
Less than high school	1.54	1.41	1.68	1.52	1.39	1.66	-1.2	1.65	1.45	1.87	1.64	1.43	1.87	-0.5	1.64	1.40	1.93	1.65	1.41	1.93	0.6
High school diploma or GED	1.35	1.25	1.46	1.33	1.23	1.44	-1.3	1.27	1.14	1.42	1.27	1.14	1.42	0.1	1.29	1.10	1.50	1.29	1.10	1.50	-0.1
Marital status (married):																					
Not married	1.29	1.19	1.41	1.30	1.20	1.42	0.7	1.55	1.43	1.68	1.55	1.43	1.69	0.2	1.77	1.52	2.06	1.78	1.53	2.07	0.6

--- Data not available.

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 5. Hazard ratios for diseases of the heart mortality: National Health and Nutrition Examination Survey III, National Health and Nutrition Examination Survey, 1999–2006, and National Health and Nutrition Examination Survey, 2007–2014, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							NHANES 2007–2014								
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	
Age, years	1.11	1.10	1.12	1.11	1.10	1.12	0.0	1.10	1.09	1.11	1.10	1.09	1.11	0.0	1.10	1.08	1.12	1.10	1.08	1.12	0.2		
Sex (female):																							
Male	1.94	1.75	2.16	1.94	1.73	2.18	−0.2	2.28	1.91	2.73	2.12	1.76	2.55	−7.7	2.64	2.13	3.28	2.22	1.82	2.72	−18.8		
Race and ethnicity (non-Hispanic white):																							
Non-Hispanic black	1.12	0.94	1.32	1.12	0.95	1.31	−0.1	1.27	1.03	1.56	1.21	0.99	1.49	−4.9	1.08	0.75	1.55	1.01	0.72	1.41	−6.9		
Mexican American	0.90	0.71	1.14	0.82	0.65	1.03	−9.4	1.03	0.75	1.41	0.94	0.71	1.26	−9.0	---	---	---	---	---	---	---		
All Hispanic	---	---	---	---	---	---	---	---	---	---	---	---	---	---	0.71	0.47	1.08	0.69	0.44	1.09	−3.0		
Education (some college and above):																							
Less than high school	1.52	1.27	1.82	1.56	1.32	1.84	2.2	1.51	1.16	1.95	1.51	1.16	1.95	0.0	2.02	1.39	2.93	1.81	1.30	2.53	−11.3		
High school diploma or GED	1.38	1.17	1.64	1.44	1.21	1.72	3.8	1.00	0.76	1.32	1.12	0.85	1.48	11.1	1.88	1.29	2.74	1.59	1.09	2.31	−18.6		
Marital status (married):																							
Not married	1.33	1.11	1.59	1.31	1.10	1.57	−1.1	1.64	1.29	2.08	1.72	1.39	2.14	4.9	2.45	1.77	3.38	2.56	1.83	3.59	4.5		

--- Data not available.

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 6. Hazard ratios for malignant neoplasms (cancer) mortality: National Health and Nutrition Examination Survey III, National Health and Nutrition Examination Survey, 1999–2006, and National Health and Nutrition Examination Survey, 2007–2014, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							NHANES 2007–2014								
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)		Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference		
Age, years.	1.07	1.06	1.08	1.07	1.06	1.08	0.0	1.07	1.07	1.08	1.07	1.07	1.08	0.1	1.07	1.06	1.09	1.08	1.07	1.09	0.6		
Sex (female):																							
Male	1.43	1.21	1.68	1.44	1.22	1.69	0.6	1.74	1.41	2.13	1.79	1.47	2.18	3.1	1.76	1.32	2.35	2.06	1.61	2.65	14.5		
Race and ethnicity (non-Hispanic white):																							
Non-Hispanic black	1.17	0.97	1.41	1.15	0.96	1.37	-2.3	1.33	1.07	1.66	1.27	1.03	1.57	-5.1	1.16	0.88	1.53	1.06	0.82	1.36	-9.8		
Mexican American.	0.92	0.74	1.14	0.93	0.75	1.16	1.3	0.93	0.69	1.25	0.96	0.71	1.28	2.8	---	---	---	---	---	---	---		
All Hispanic	---	---	---	---	---	---	---	---	---	---	---	---	---	---	0.92	0.65	1.30	0.92	0.67	1.27	0.2		
Education (some college and above):																							
Less than high school	1.60	1.23	2.08	1.58	1.21	2.06	-1.4	1.50	1.17	1.93	1.47	1.15	1.89	-2.0	1.61	1.19	2.18	1.62	1.21	2.18	0.6		
High school diploma or GED.	1.39	1.12	1.71	1.33	1.08	1.63	-4.6	1.16	0.92	1.47	1.15	0.91	1.45	-1.1	1.14	0.81	1.59	1.28	0.93	1.76	11.0		
Marital status (married):																							
Not married	1.07	0.90	1.27	1.10	0.93	1.30	3.1	1.15	0.95	1.38	1.14	0.96	1.35	-1.2	1.30	0.96	1.77	1.32	0.99	1.75	1.0		

--- Data not available.

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 7. Hazard ratios for chronic lower respiratory disease mortality: National Health and Nutrition Examination Survey III and National Health and Nutrition Examination Survey, 1999–2006, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006						
	Public-use			Restricted-use				Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference
Age, years.....	1.09	1.08	1.10	1.09	1.08	1.10	0.1	1.10	1.09	1.12	1.10	1.09	1.12	–0.2
Sex (female):														
Male	1.68	1.23	2.31	1.72	1.25	2.36	2.0	1.65	1.18	2.31	1.53	1.07	2.18	–7.8
Race and ethnicity (non-Hispanic white):														
Non-Hispanic black	0.51	0.36	0.72	0.54	0.38	0.75	5.2	0.48	0.31	0.76	0.45	0.29	0.69	–8.1
Mexican American.....	0.25	0.16	0.39	0.27	0.17	0.42	7.3	0.23	0.14	0.40	0.25	0.15	0.42	6.1
Education (some college and above):														
Less than high school	2.56	1.66	3.93	2.53	1.66	3.87	–0.8	2.64	1.74	4.00	2.66	1.79	3.97	0.9
High school diploma or GED.....	2.35	1.49	3.70	2.35	1.49	3.71	0.0	1.22	0.81	1.83	1.45	0.97	2.18	16.1
Marital status (married):														
Not married	1.50	1.09	2.06	1.51	1.09	2.10	0.9	1.72	1.17	2.54	1.57	1.07	2.29	–9.9

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 8. Hazard ratios for cerebrovascular disease mortality: National Health and Nutrition Examination Survey III and National Health and Nutrition Examination Survey, 1999–2006, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	
Age, years.....	1.12	1.11	1.13	1.12	1.11	1.13	0.1	1.12	1.10	1.14	1.12	1.10	1.14	0.1	
Sex (female):															
Male	1.03	0.78	1.35	1.03	0.78	1.36	0.1	1.63	1.08	2.45	1.64	1.10	2.46	0.7	
Race and ethnicity (non-Hispanic white):															
Non-Hispanic black	1.59	1.20	2.12	1.68	2.27	2.21	4.8	1.15	0.75	1.76	1.10	0.72	1.67	-4.9	
Mexican American.....	1.48	1.06	2.06	1.45	1.05	2.00	-2.2	1.43	0.94	2.17	1.37	0.91	2.08	-3.8	
Education (some college and above):															
Less than high school	0.91	0.66	1.25	0.90	0.66	1.24	-0.5	1.53	1.00	2.35	1.54	1.03	2.29	0.3	
High school diploma or GED.....	1.10	0.79	1.54	1.12	0.81	1.53	1.3	1.17	0.76	1.82	1.18	0.76	1.82	0.1	
Marital status (married):															
Not married	1.24	0.89	1.72	1.23	0.89	1.69	-1.0	1.12	0.74	1.71	1.15	0.76	1.73	2.5	

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 9. Hazard ratios for accident (unintentional injury) mortality: National Health and Nutrition Examination Survey III and National Health and Nutrition Examination Survey, 1999–2006, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	
Age, years.....	1.03	1.00	1.05	1.03	1.01	1.05	0.1	1.01	1.00	1.03	1.02	1.00	1.04	0.4	
Sex (female):															
Male	1.77	1.08	2.89	1.78	1.09	2.90	0.5	1.22	0.79	1.88	1.16	0.72	1.87	-5.2	
Race and ethnicity (non-Hispanic white):															
Non-Hispanic black	0.82	0.45	1.50	0.85	0.47	1.56	4.0	0.45	0.22	0.91	0.44	0.22	0.87	-3.8	
Mexican American.....	1.07	0.62	1.85	1.02	0.58	1.79	-4.8	1.78	0.80	3.97	1.68	0.78	3.60	-6.3	
Education (some college and above):															
Less than high school	3.32	1.72	6.42	3.22	1.69	6.16	-3.0	2.37	1.08	5.17	2.08	0.96	4.50	-14.0	
High school diploma or GED.....	1.25	0.65	2.38	1.21	0.64	2.30	-3.1	1.58	0.91	2.74	1.44	0.86	2.42	-9.5	
Marital status (married):															
Not married	1.66	10.60	2.59	1.67	1.07	2.60	0.7	2.39	1.41	4.05	2.16	1.32	3.55	-10.4	

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 10. Hazard ratios for Alzheimer disease mortality: National Health and Nutrition Examination Survey III and National Health and Nutrition Examination Survey, 1999–2006, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	
Age, years.....	1.17	1.15	1.19	1.17	1.15	1.19	0.0	1.20	1.17	1.23	1.20	1.17	1.23	0.0	
Sex (female):															
Male	1.08	0.67	1.75	1.09	0.67	1.76	0.5	0.90	0.57	1.41	0.88	0.57	1.38	-1.7	
Race and ethnicity (non-Hispanic white):															
Non-Hispanic black	0.65	0.37	1.15	0.68	0.39	1.19	3.9	0.82	0.46	1.46	0.75	0.42	1.32	-9.8	
Mexican American.....	0.68	0.40	1.15	0.65	0.37	1.13	-5.4	0.49	0.27	0.88	0.53	0.31	0.90	7.4	
Education (some college and above):															
Less than high school	1.60	1.02	2.51	1.56	1.00	2.44	-2.3	0.82	0.54	1.26	0.75	0.49	1.15	-10.1	
High school diploma or GED.....	0.95	0.46	1.95	0.92	0.45	1.88	-3.1	1.48	0.85	2.55	1.34	0.79	2.30	-9.9	
Marital status (married):															
Not married	1.07	0.66	1.72	1.08	0.67	1.74	0.8	0.88	0.54	1.43	0.94	0.58	1.53	6.8	

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 11. Hazard ratios for diabetes mellitus mortality: National Health and Nutrition Examination Survey III and National Health and Nutrition Examination Survey, 1999–2006, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006						
	Public-use			Restricted-use				Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference
Age, years.....	1.08	1.07	1.10	1.08	1.07	1.10	0.0	1.08	1.05	1.10	1.07	1.05	1.09	–0.1
Sex (female):														
Male	1.77	1.15	2.72	1.78	1.16	2.74	0.6	1.26	0.84	1.90	1.39	0.94	2.06	9.1
Race and ethnicity (non-Hispanic white):														
Non-Hispanic black	2.07	1.36	3.15	2.16	1.44	3.25	4.2	1.73	1.00	2.99	1.60	0.94	2.73	–8.0
Mexican American.....	1.72	1.10	2.67	1.65	1.06	2.58	–3.7	1.65	0.79	3.46	1.78	0.90	3.51	7.0
Education (some college and above):														
Less than high school	3.50	1.81	6.74	3.44	1.79	6.62	–1.7	2.84	1.48	5.44	2.92	1.56	5.46	2.7
High school diploma or GED.....	3.82	1.61	9.07	3.72	1.58	8.80	–2.7	1.35	0.70	2.61	1.46	0.78	2.71	7.0
Marital status (married):														
Not married	1.22	0.83	1.78	1.24	0.84	1.82	1.7	1.15	0.64	2.06	1.24	0.71	2.18	7.4

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey linked mortality files.

Table 12. Hazard ratios for nephritis, nephrotic syndrome and nephrosis disease mortality: National Health and Nutrition Examination Survey III and National Health and Nutrition Examination Survey, 1999–2006, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	
Age, years.....	1.11	1.09	1.14	1.12	1.09	1.14	0.0	1.11	1.08	1.14	1.10	1.08	1.13	–0.5	
Sex (female):															
Male	1.25	0.65	2.43	1.28	0.67	2.45	2.3	2.26	1.29	3.99	2.34	1.35	4.04	3.2	
Race and ethnicity (non-Hispanic white):															
Non-Hispanic black	2.70	1.58	4.61	2.74	1.61	4.66	1.4	2.01	1.06	3.82	1.94	1.01	3.72	–3.6	
Mexican American.....	1.99	0.93	4.22	2.18	1.04	4.61	9.1	0.92	0.39	2.13	0.89	0.37	2.14	–2.6	
Education (some college and above):															
Less than high school	1.34	0.74	2.42	1.34	0.74	2.42	–0.3	1.95	0.95	3.98	2.09	1.01	4.31	6.7	
High school diploma or GED.....	1.28	0.57	2.87	1.26	0.56	2.84	–1.5	1.47	0.77	2.79	1.75	0.88	3.50	16.2	
Marital status (married):															
Not married	0.89	0.49	1.62	0.90	0.50	1.61	0.2	2.68	1.50	4.80	2.47	1.41	4.33	–8.7	

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

Table 13. Hazard ratios for influenza and pneumonia mortality: National Health and Nutrition Examination Survey III and National Health and Nutrition Examination Survey, 1999–2006, linked mortality files (follow-up through 2015)

Characteristic	NHANES III							NHANES 1999–2006							
	Public-use			Restricted-use				Relative percent difference	Public-use			Restricted-use			
	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio		Lower bound (95% CI)	Upper bound (95% CI)	Hazard ratio	Lower bound (95% CI)	Upper bound (95% CI)	Relative percent difference	
Age, years.....	1.14	1.12	1.16	1.14	1.12	1.16	0.0	1.14	1.11	1.17	1.14	1.11	1.17	0.1	
Sex (female):															
Male	2.12	1.41	3.17	2.12	1.42	3.16	0.2	1.78	1.11	2.85	1.68	1.06	2.68	-5.7	
Race and ethnicity (non-Hispanic white):															
Non-Hispanic black	1.06	0.63	1.79	1.09	0.65	1.83	2.7	1.32	0.73	2.38	1.27	0.71	2.29	-3.6	
Mexican American.....	0.81	0.44	1.50	0.84	0.47	1.50	3.1	1.64	0.69	3.86	1.69	0.77	3.73	3.2	
Education (some college and above):															
Less than high school	2.07	1.15	3.73	1.97	1.13	3.43	-5.3	1.35	0.79	2.32	1.35	0.79	2.31	-0.1	
High school diploma or GED.....	1.86	1.00	3.49	1.76	0.97	3.19	-6.1	1.18	0.55	2.51	1.30	0.62	2.72	9.8	
Marital status (married):															
Not married	1.48	0.91	2.41	1.47	0.91	0.24	-0.9	1.45	0.93	2.25	1.38	0.90	2.13	-4.8	

NOTES: NHANES is National Health and Nutrition Examination Survey. CI is confidence interval. Hazard ratios are estimated from a Cox proportional hazards model. All models adjust for sample weights and the National Health and Nutrition Examination Survey complex survey design using the SUDAAN software program (11.0). Values in parentheses are reference categories.

SOURCE: National Center for Health Statistics, National Health and Nutrition Examination Survey, linked mortality files.

**U.S. DEPARTMENT OF
HEALTH & HUMAN SERVICES**

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782-2064

FIRST CLASS MAIL
POSTAGE & FEES PAID
CDC/NCHS
PERMIT NO. G-284

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300

For more NCHS NHSRs, visit:
<https://www.cdc.gov/nchs/products/nhsr.htm>.



National Health Statistics Reports ■ Number 155 ■ May 24, 2021

Suggested citation

Mirel LB, Zhang C, Cox CS, Ye Y, El Burai Félix S, Golden C. Comparative analysis of the National Health and Nutrition Examination Survey public-use and restricted-use linked mortality files. National Health Statistics Reports; no 155. Hyattsville, MD: National Center for Health Statistics. 2021. DOI: <https://doi.org/10.15620/cdc:104744>.

Copyright information

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

National Center for Health Statistics

Brian C. Moyer, Ph.D., *Director*
Amy M. Branum, Ph.D., *Acting Associate
Director for Science*

Division of Analysis and Epidemiology

Irma E. Arispe, Ph.D., *Director*
Kevin C. Heslin, Ph.D., *Associate Director
for Science*

For e-mail updates on NCHS publication releases, subscribe online at: <https://www.cdc.gov/nchs/email-updates.htm>.
For questions or general information about NCHS: Tel: 1-800-CDC-INFO (1-800-232-4636) • TTY: 1-888-232-6348
Internet: <https://www.cdc.gov/nchs> • Online request form: <https://www.cdc.gov/info> • CS323656