

## **Comparative Analysis of the NHANES III Public-Use and Restricted-Use Linked Mortality Files**

Suggested citation: Wheatcroft G, Cox CS, Lochner KA. “Comparative analysis of the NHANES III public-use and restricted-use linked mortality files.” National Center for Health Statistics. 2007. Hyattsville, Maryland. (Available at the following address: [http://www.cdc.gov/nchs/data/datalinkage/nh3\\_mort\\_compare\\_2007\\_final.pdf](http://www.cdc.gov/nchs/data/datalinkage/nh3_mort_compare_2007_final.pdf)).

### **Introduction**

Federally sponsored health surveys are a critical source of information on public health in the United States. Yet, access to nationally representative survey data that provide a longitudinal mortality component is limited. The National Center for Health Statistics (NCHS) periodically conducts mortality follow-up studies through record linkage to administrative data sources for its major population-based surveys (see [NCHS Data Linkage Activities](#)). These linked mortality files fill research gaps by creating data resources that contain high quality socio-demographic, health, and mortality information for nationally representative U.S. samples.

In 2004, NCHS completed the first mortality follow-up study for the third National Health and Nutrition Examination Survey ([NHANES III](#) conducted during 1988-94). Due to requirements to protect the confidentiality of the NHANES III participants, the NHANES III Linked Mortality File was made available only through the [NCHS Research Data Center \(RDC\)](#). Recognizing that this would place restrictions on researchers’ use of a unique data resource, NCHS developed a data perturbation plan to allow for a public-use release of the linked mortality data. The data release plan was developed to eliminate re-identification of survey participants, maximize the amount of data to be included in the release, and limit the amount of synthetic data introduced to the data file. In 2007, NCHS released a public-use version of the NHANES III Linked Mortality File.<sup>1</sup>

---

<sup>1</sup> NCHS also has released public-use files for the NHIS Linked Mortality Files, with a similar [comparative report](#) of the public-use and restricted-use versions.

This report describes a comparative analysis of the public-use and restricted-use NHANES III Linked Mortality Files. We used Cox proportional hazards models to compare the relative hazards for a standard set of socio-demographic covariates for all-cause as well as cause-specific mortality risk. NCHS is conducting this comparative analysis to demonstrate the comparability between the two versions of the linked mortality files.

#### *Description of Linked Mortality Data Resources*

The linkage of the NHANES III to death certificate data found in the [National Death Index \(NDI\)](#) was conducted for the NHANES III survey, with mortality follow-up through December 31, 2000. Mortality information is based upon the results from a probabilistic match between NHANES III and NDI death certificate records, with additional NCHS review and verification of death certificates for a select sample of the NHANES III participants, who had a single best NDI record match chosen. For more information on the matching methodology, please refer to the [NHANES III Linked Mortality File Matching Methodology Report](#). All NHANES III participants are included on the linked mortality file, but only participants 17 years and older at the time of the survey were eligible for mortality follow-up. NCHS has created two versions of the new NHANES III Linked Mortality File: a [restricted-use file](#) that includes more detailed mortality information and a [public-use file](#) that includes a limited set of mortality variables.

The restricted-use file includes the following variables: survey respondent eligibility status, mortality status, age when last presumed alive, age at death, date of death (month, day and year), underlying and multiple cause-of-death, and NHANES III interview and exam dates (month, day, and year). The public-use file was subjected to standard data perturbation techniques that introduce statistical noise into the data set, in order to reduce the risk of respondent re-identification. Synthetic data were substituted for the actual date and underlying cause-of-death data for selected decedent records. Information regarding vital status was not perturbed. Variables provided on the public-use NHANES III Linked Mortality File include: survey respondent eligibility status, mortality status,

person-months of follow-up from NHANES III interview date, person-months of follow-up from NHANES III exam date, and underlying cause-of-death 113 group recode. In addition, three variables were created to indicate the presence of diabetes, hypertension, or hip fracture in the multiple cause-of-death codes, as these conditions are more often reported as contributing, rather than underlying, causes of death.

## **Methods**

### *Analytic sample*

To effectively compare the public-use and restricted-use data sets, we merged the public-use NHANES III interview file with the accompanying public-use and restricted-use mortality files, respectively, to create the analytic samples. We restricted all analyses to those eligible for mortality follow-up, who were at least 25 years of age at the time of the NHANES III interview, and were non-Hispanic white, non-Hispanic black, or Mexican American, with no missing values for education level, and with person-months of follow-up greater than zero.

### *Outcome measurement*

We examined all-cause and cause-specific mortality in the public-use and restricted-use NHANES III Linked Mortality Files. Respondents who were not identified as dying by the end of the follow-up period were assumed to be alive. For decedents in the public-use file, we used the person-months of follow-up from interview variable, which was constructed for the public-use file using the NHANES III interview month and year, and the month and year of death. Respondents who died in the same month and year as their NHANES III interview were given zero months of follow-up time as their duration. These individuals were dropped from the analytic sample in our comparative analysis. For respondents assumed alive, person-months of follow-up were calculated from the NHANES III interview month and year to December 31, 2000. For decedents in the restricted-use file, person-months of follow-up were calculated using complete information on the month, day, and year of the NHANES III interview and the month, day, and year of death and for respondents assumed alive, the month, day, and year of interview until the end of the follow-up period (December 31, 2000). Thus, minor

differences in model results may affect the comparative results below because of the differing calculations of the duration of follow-up time that were used for the two files.

In addition to all-cause mortality, we examined five causes of death that are among the leading causes of death in the United States.<sup>1</sup> The NHANES III Linked Mortality file encompasses both the Ninth Revision of the *International Classification of Diseases* (ICD-9) and the Tenth Revision (ICD-10) cause of death coding for all U.S. deaths. In order to have the same cause of death codes across all years in the study period, we used the ICD-10 underlying cause-of-death 113-group recode, which recodes all deaths occurring prior to 1999 into ICD-10 codes.<sup>2</sup> However, even though 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 and those that occurred in later years under ICD-10. The analyses presented in this paper 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 paper.

The cause-specific death categories include the following [Underlying Cause-of-Death Recoded 113 Groups](#) (UCOD-113): heart disease (55-68), ischemic heart disease (59-61), cancer from all sites (20-44), lung cancer (27), and cerebrovascular disease (70).

### *Covariates*

All models included a standard set of socio-demographic characteristics, which were collected at the time of NHANES III interview: age in continuous years, sex, race/ethnicity (non-Hispanic black, non-Hispanic white, Mexican American), and educational attainment (less than high school, high school diploma, more than high school).

### *Data Analysis*

We used Cox proportional hazards models to compare the relative hazards for the covariates for all-cause as well as cause-specific mortality risk. All relative hazards were calculated with the survival procedure in Software for Survey Data Analysis (SUDAAN),

version 9.0.1 to take into account the complex survey design of the NHANES III.<sup>3</sup> The Efron method was used for handling tied failure times.<sup>4</sup> Due to an insufficient number of deaths for Mexican Americans, the cause-specific mortality analyses are restricted to non-Hispanic whites and non-Hispanic blacks.

## **Results**

### *Descriptive Results*

The public-use and restricted-use NHANES III Linked Mortality Files each contain 33,994 records and 3,384 deaths. After excluding NHANES III respondents ineligible for mortality follow-up (n = 13,970) and individuals with missing or zero years of mortality follow-up, the final sample for the comparative analyses included 16,048 records ([Table 1](#)) and 3,209 deaths. [Table 2.1](#) shows the unweighted sample sizes and weighted percentage distributions for the covariates used in the analysis. Note that these descriptive statistics for covariates do not differ between the public-use and restricted-use files because the only differences between the two files are associated with the variables taken from the mortality file. Using the weighted distributions of covariates for this sample, the average age is 47.8 years and fewer than two percent of respondents are aged 85 or above. Females outnumber males, and non-Hispanic whites make up 84 percent of the sample while non-Hispanic blacks (11.4 percent) and Mexican Americans (4.9 percent) account for considerably small proportions.

[Table 2.2](#) shows the comparative descriptive statistics for mortality outcome variables in the public-use and restricted-use files, respectively. Note first that the total number and percentage of persons who were identified as dying in each of the two files (n = 3,209; Percent = 12.1) is identical. As mentioned above, this illustrates that the vital status of individuals was not changed for anyone as a result of the perturbation process for the public-use file. At the same time, however, there are some modest differences in the cause of death distributions when comparing the public-use and restricted-use files. While the cause-specific percentage distributions are, overall, quite similar when comparing the two files, there are some differences that should be noted. For example,

the number of deaths attributed to heart disease (n = 1,188) and ischemic heart disease (n = 344) in the public-use file is greater than the number of deaths attributed to heart disease (n = 1,158) and ischemic heart disease (n = 336) in the restricted-use file, while the number of deaths attributed to all-cancer in the public-use file is smaller (n = 689) than the number of deaths in the restricted-use file (n = 698). Similarly, there are modest differences when looking at the numbers and percentage distribution of deaths for the other underlying causes when comparing the two files (not shown).

#### *All-Cause Mortality Model Results*

[Table 3.1](#) displays results from two Cox proportional hazards models of all-cause mortality: one estimated from the public-use file and one estimated from the restricted-use file. The results of both models are consistent. Age is very strongly and positively related to the risk of adult mortality, and mortality risk is substantively and significantly different by racial/ethnic and educational groups. For example, non-Hispanic blacks and persons with less than a high school education display the highest risks of mortality compared to their respective counterpart subgroups. Moreover, coefficients and standard errors are nearly identical when comparing the results from the public-use and restricted-use files, with only slight differences occurring in the third and fourth decimal places of the coefficients and standard errors, respectively. Recall that there are differences in the way that the duration of follow-up variable was calculated for these two versions of the NHANES III Linked Mortality Files; thus, these very slight differences in model results for all-cause mortality are due to the differences in the duration of follow-up variable.

The results of all-cause Cox proportional hazards models of adult mortality that are estimated separately by sex are shown in [Table 3.2](#). For each sex, results from the public-use and restricted-use files are shown. The sex-specific models yield consistent results, with nearly identical coefficients and standard errors when the public-use and restricted-use files are compared.

Finally, [Table 3.3](#) shows the results of separate proportional hazards models for non-Hispanic whites, non-Hispanic blacks, and Mexican Americans, respectively. Again,

there are only tiny and non-substantive differences when comparing the results from the public-use and restricted-use files for each of the three racial/ethnic groups. For each group, covariates exhibit relationships with all-cause mortality that are consistent with what one would expect from the U.S. adult mortality literature; for example, males exhibit higher mortality risks than females in each racial/ethnic group and persons with less than a high school education demonstrate higher mortality risks over the follow-up period.

#### *Cause-Specific Mortality Model Results*

[Tables 4.1](#) through [4.5](#) display the results of the Cox proportional hazards models for five specific underlying causes-of-death. Each cause-specific table provides a comparison of the model results from the public-use version and the restricted-use version of the NHANES III Linked Mortality File. As previously mentioned, these cause-specific results are limited to adult individuals, who are 25 years of age or older at time of interview, and who are identified as non-Hispanic white or non-Hispanic black. Females and males are included in these cause-specific models, with a dummy variable for sex in each model. Those with missing data on education or cause of death and person-months of follow-up equal to zero were excluded. Some of the specific causes (e.g., lung cancer, ischemic heart disease) are subsets of a larger underlying cause category (e.g., all-cancer mortality, heart disease).

A comparison of the results for the public-use and restricted-use files for each of the five causes yields no substantive differences in conclusions, and coefficients and standard errors that are very similar. That is, when comparing public-use and restricted-use coefficients and significance tests for each cause, the conclusions to be reached from the models are identical and there are only minor differences when comparing the actual coefficients and standards errors of the models.

To illustrate an example of the consistency between results from the public-use data and restricted-use data, [Table 4.3](#) provides comparative models that specify all-cancer mortality as the outcome variable. Mortality risk increases just over seven percent for

each additional year of age in both the public-use data model and the restricted-use data model. Males experience 30 percent higher cancer mortality risk than females over the course of the follow-up period according to the public-use data and 33 percent higher cancer mortality risk than females according to the restricted-use data.

## **Discussion**

This report describes analyses comparing results obtained from the public-use and restricted-use versions of the NHANES III Linked Mortality File. In the public-use version of the data file, a limited amount of information for decedents was perturbed. Further, there is less detailed mortality information in the public-use version, compared to the restricted-use file, where no information has been perturbed and there is complete information on date of death; including month, day and year.

The comparative analysis finds that the two data files yield very similar descriptive and model results. This is particularly true when examining all-cause mortality. Because the perturbation process in the public-use file did not affect the vital status of any individuals in the file, the only differences in results between the two files when examining overall (all-cause) mortality arose because the public-use file has less specific information available regarding timing of death for individuals who died compared to the restricted-use file. In the end, the differences that resulted from the comparisons of all-cause mortality between the public-use file and restricted-use file, then, were very minor.

The comparative analysis of cause-specific mortality across the public-use and restricted-use versions of the NHANES III Linked Mortality File also yielded only very slight differences in model results. The frequency distributions that were shown for cause of death for the public-use and restricted-use versions of the NHANES III Linked Mortality File demonstrated that the perturbation process in the public-use file did have a minor impact on the number of persons identified as dying from each cause as well as the overall distribution of deaths. This should be kept in mind when conducting cause-specific analyses of the public-use file. Nevertheless, the coefficients and standard errors in the cause-specific models that we have estimated demonstrate that such differences in



the identification of causes of death for some cases result in only very slight, and statistically and substantively insignificant, changes in the comparative results. Moreover, no differences in conclusions could be reached based on these cause-specific models when comparing the public-use and restricted-use data sets.

Our findings should provide analysts with the confidence to use the perturbed public-use version of the NHANES III Linked Mortality File. In addition, there are some analytic considerations that should be noted by all potential users. First, the NHANES III Linked Mortality File requires the use of statistical software that takes into account the complex survey design of the NHANES III. We used the statistical software package SUDAAN 9.0.1 because it fits Cox proportional hazard models to sample surveys. Caution in using the public-use file is urged when examining the mortality patterns of small subgroups of the population, such as numerically small racial/ethnic minority groups, very old individuals, or young adults. This is particularly the case when cause-specific analyses of such numerically small demographic subgroups are performed. Caution also is urged when conducting analyses that allow participants to age into varying age strata over the follow-up period. The availability of more precise and detailed age and follow-up information on the restricted-use file could lead to different samples being obtained in the various age strata. Researchers using the public-use data file for such analyses are strongly encouraged to confirm their findings with the restricted-use data.

In sum, the new public-use version of the NHANES III Linked Mortality File provides the public health, social science, demographic, and medical communities with a data set that is easily available, very large, nationally representative, and rich in detail for both mortality covariates and specificity in outcomes. The public-use file is an important resource for researchers and policymakers in furthering our understanding of adult mortality trends and patterns.

## References

1. National Center for Health Statistics. [Health, United States, 2006](#). Hyattsville, MD: 2006.
2. Anderson RN, Minino AM, Hoyert DL, Rosenberg HM. [Comparability of cause of death between ICD-9 and ICD-10: Preliminary estimates](#). National Center for Health Statistics. *National Vital Statistics Reports* 49. 2001;No.2.
3. SUDAAN: Software for the Statistical Analysis of Correlated Data, 9.01. RTI International.
4. Hertz-Picciotto I; Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics*. 1997;53:1151-1156.

**Table 1. Sample selection criteria for the comparative analyses of the NHANES III Linked Mortality Files**

	Number of records	Excluded cases
Total records	33,994	
Sample selection criteria		
Eligible for mortality follow-up <sup>1</sup>		13,970
Age 25 years or older at time of NHANES III interview		3,165
Race/ethnicity restricted to non-Hispanic whites, non-Hispanic blacks, and Mexican Americans. Non missing values for education		801
Non missing values for underlying cause of death 113-group recode		2
Person-months of follow-up greater than zero		8
Total number of excluded cases		17,946
<b>Final cases included in sample</b>	<b>16,048</b>	

Notes:

<sup>1</sup>Only NHANES III participants 17 years of age and older with sufficient linkage information were eligible for mortality follow-up.

**Table 2.1. Sample characteristics, NHANES III household adult data file: n = 16,048**

	Unweighted n	Weighted percentage
Age in years, mean	52.7	47.8
Age in years (grouped)		
25-44	6,728	51.0
45-64	4,288	29.4
65-84	4,376	18.1
85+	656	1.6
Sex		
Male	7,510	47.4
Female	8,538	52.6
Race/Ethnicity		
non-Hispanic white	7,580	83.7
non-Hispanic black	4,405	11.4
Mexican American	4,063	4.9
Education level		
Less than high school	6,831	24.6
High school	4,731	34.1
More than high school	4,486	41.3

**Table 2.2. Sample characteristics, NHANES III linked mortality variables: n = 16,048**

	Public-use		Restricted-use	
	Unweighted n	Weighted percentage	Unweighted n	Weighted percentage
Follow-up period in months, mean	99.6	103.8	99.9	104.1
Assigned vital status				
Dead	3,209	12.1	3,209	12.1
Alive	12,839	87.9	12,839	87.9
Cause-specific deaths (113-group recode) <sup>1</sup>				
Diseases of the heart (55-68)	1,188	35.5	1,158	34.8
Ischemic heart disease (59-61)	344	11.4	336	11.3
Cancer, all sites (20-44)	689	25.0	698	25.3
Lung cancer (27)	180	7.6	179	7.6
Cerebrovascular diseases (70)	266	6.9	269	7.0

<sup>1</sup>Weighted percentages based upon sample of decedents (n =3,209).

**Table 3.1. Relative hazard for all-cause mortality: Comparative analyses of public-use and restricted-use NHANES III Linked Mortality Files (n = 16,048)**

	Public-use		Relative hazard	Restricted-use		Relative hazard
	Coefficient	SE		Coefficient	SE	
Age in years	0.086 ***	0.0031	1.090	0.086 ***	0.0031	1.090
Sex (Female)						
Male	0.381 ***	0.0469	1.464	0.381 ***	0.0468	1.464
Race/ethnicity (NHW)						
NHB	0.321 ***	0.0628	1.378	0.322 ***	0.0629	1.380
Mexican American	-0.007	0.0752	0.993	-0.008	0.0762	0.992
Education (More than high school)						
Less than high school	0.333 ***	0.0609	1.395	0.332 ***	0.0609	1.394
High school	0.245 **	0.0719	1.277	0.245 **	0.0719	1.278

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models adjust for sample weights and the NHANES III complex survey design using the SUDAAN software program (9.0.1).

NHW refers to non-Hispanic white; NHB refers to non-Hispanic black; the values in parentheses are reference categories.

\* p <.05; \*\* p <.01; \*\*\* p <.001.

**Table 3.2. Relative hazard for all-cause mortality by sex: Comparative analyses of public-use and restricted-use NHANES III Linked Mortality Files (n =16,048)**

	Men						Women					
	Public-use		Relative hazard	Restricted-use		Relative hazard	Public-use		Relative hazard	Restricted-use		Relative hazard
	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
Age in years	0.087 ***	0.0043	1.091	0.087 ***	0.0043	1.091	0.085 ***	0.0033	1.089	0.085 ***	0.0033	1.089
Race/ethnicity (NHW)												
NHB	0.441 ***	0.0883	1.555	0.443 ***	0.0885	1.558	0.204 **	0.0753	1.226	0.204 **	0.0755	1.227
Mexican American	0.050	0.1088	1.052	0.052	0.1086	1.053	-0.070	0.0952	0.933	-0.075	0.0960	0.928
Education (More than high school)												
Less than high school	0.303 ***	0.0854	1.354	0.302 ***	0.0853	1.352	0.355 **	0.1017	1.425	0.356 **	0.1019	1.427
High school	0.319 ***	0.0778	1.376	0.318 ***	0.0778	1.375	0.175	0.1172	1.192	0.178	0.1173	1.194

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models adjust for sample weights and the NHANES III complex survey design using the SUDAAN software program (9.0.1).

NHW refers to non-Hispanic white; NHB refers to non-Hispanic black; the values in parentheses are reference categories.

\* p <.05; \*\* p <.01; \*\*\* p <.001.

**Table 3.3. Relative hazard for all-cause mortality by race/ethnicity: Comparative analyses of public-use files and restricted-use NHANES III Linked Mortality File (n =16,048 )**

	non-Hispanic whites						non-Hispanic blacks					
	Public-use		Relative hazard	Restricted-use		Relative hazard	Public-use		Relative hazard	Restricted-use		Relative hazard
	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
Age in years	0.090 ***	0.0040	1.094	0.090 ***	0.0040	1.094	0.072 ***	0.0035	1.074	0.072 ***	0.0035	1.074
Sex (Female)												
Male	0.368 ***	0.0525	1.445	0.368 ***	0.0524	1.445	0.520 ***	0.0817	1.682	0.521 ***	0.0816	1.683
Education (More than high school)												
Less than high school	0.302 ***	0.0692	1.353	0.302 ***	0.0693	1.353	0.584 ***	0.1555	1.794	0.583 ***	0.1560	1.792
High school	0.213 *	0.0796	1.238	0.214 *	0.0798	1.239	0.556 **	0.1670	1.744	0.557 **	0.1671	1.745

	Mexican American					
	Public-use		Relative hazard	Restricted-use		Relative hazard
	Coefficient	SE		Coefficient	SE	
Age	0.062 ***	0.0033	1.064	0.062 ***	0.0033	1.064
Sex (Female)						
Male	0.321 *	0.1342	1.379	0.322 *	0.1338	1.380
Education (More than high school)						
Less than high school	0.478	0.3540	1.613	0.477	0.3540	1.611
High school	-0.026	0.3567	0.975	-0.026	0.3567	0.975

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models adjust for sample weights and the NHANES III complex survey design using the SUDAAN software program (9.0.1).

Values in parentheses are reference categories.

\* p <.05; \*\* p <.01; \*\*\* p <.001.



**Table 4.1. Relative hazard for heart disease mortality: Comparative analyses of public-use and restricted-use NHANES III Linked Mortality Files, non-Hispanic whites and non-Hispanic blacks only (n =11,985)**

	Public-use		Relative hazard	Restricted-use		Relative hazard
	Coefficient	SE		Coefficient	SE	
Age in years	0.103 ***	0.0051	1.109	0.103 ***	0.0051	1.108
Sex (Female)						
Male	0.592 ***	0.0640	1.808	0.587 ***	0.0636	1.799
Race/ethnicity (NHW)						
NHB	0.241 *	0.0900	1.272	0.238 *	0.0933	1.269
Education (More than high school)						
Less than high school	0.357 **	0.1147	1.429	0.395 ***	0.1093	1.484
High school	0.158	0.1086	1.171	0.192	0.1045	1.212

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models use sample weights and take into account the NHANES III complex survey design using the SUDAAN software program (9.0.1).

NHW refers to non-Hispanic white; NHB refers to non-Hispanic black; the values in parentheses are reference categories.

\* p <.05; \*\* p<.01; \*\*\* p <.001.

**Table 4.2. Relative hazard for ischemic heart disease mortality: Comparative analyses of public-use and restricted-use NHANES III Linked Mortality Files, non-Hispanic whites and non-Hispanic blacks only (n = 11,985)**

	<u>Public-use</u>		Relative hazard	<u>Restricted-use</u>		Relative hazard
	Coefficient	SE		Coefficient	SE	
Age in years	0.076 ***	0.0081	1.079	0.077 ***	0.0082	1.080
Sex (Female)						
Male	0.789 ***	0.1398	2.200	0.771 ***	0.1376	2.161
Race/ethnicity (NHW)						
NHB	-0.254	0.2186	0.776	-0.299	0.2214	0.742
Education (More than high school)						
Less than high school	0.725 **	0.2496	2.065	0.727 **	0.2555	2.070
High school	0.417	0.2523	1.517	0.418	0.2599	1.519

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models use sample weights and take into account the NHANES III complex survey design using the SUDAAN software program (9.0.1).

NHW refers to non-Hispanic white; NHB refers to non-Hispanic black; the values in parentheses are reference categories.

\* p <.05; \*\* p<.01; \*\*\* p <.001.

**Table 4.3. Relative hazard for all-cancer mortality: Comparative analyses of public-use and restricted-use NHANES III Linked Mortality Files, non-Hispanic whites and non-Hispanic blacks only (n = 11,985)**

	<u>Public-use</u>		Relative hazard	<u>Restricted-use</u>		Relative hazard
	Coefficient	SE		Coefficient	SE	
Age in years	0.072 ***	0.0052	1.074	0.072 ***	0.0054	1.075
Sex (Female)						
Male	0.266 *	0.1198	1.304	0.281 *	0.1218	1.325
Race/ethnicity (NHW)						
NHB	0.299 *	0.1234	1.348	0.275 *	0.1239	1.316
Education (More than high school)						
Less than high school	0.351	0.1860	1.421	0.343	0.1927	1.409
High school	0.547 **	0.1629	1.728	0.519 **	0.1628	1.680

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models use sample weights and take into account the NHANES III complex survey design using the SUDAAN software program (9.0.1).

NHW refers to non-Hispanic white; NHB refers to non-Hispanic black; the values in parentheses are reference categories.

\* p <.05; \*\* p<.01; \*\*\* p <.001.

**Table 4.4. Relative hazard for lung cancer mortality: Comparative analyses of public-use and restricted-use NHANES III Linked Mortality Files, non-Hispanic whites and non-Hispanic blacks only (n = 11,985)**

	<u>Public-use</u>		Relative hazard	<u>Restricted-use</u>		Relative hazard
	Coefficient	SE		Coefficient	SE	
Age in years	0.068 ***	0.0070	1.071	0.069 ***	0.0074	1.072
Sex (Female)						
Male	0.653 **	0.2096	1.921	0.694 **	0.2235	2.001
Race/ethnicity (NHW)						
NHB	0.441	0.2186	1.553	0.370	0.2104	1.447
Education (More than high school)						
Less than high school	0.314	0.3410	1.369	0.251	0.3731	1.285
High school	0.301	0.2654	1.351	0.234	0.2608	1.263

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models use sample weights and take into account the NHANES III complex survey design using the SUDAAN software program (9.0.1).

NHW refers to non-Hispanic white; NHB refers to non-Hispanic black; the values in parentheses are reference categories.

\* p <.05; \*\* p<.01; \*\*\* p <.001.

**Table 4.5. Relative hazard for cerebrovascular diseases mortality: Comparative analyses of public-use and restricted-use NHANES III Linked Mortality Files, non-Hispanic whites and non-Hispanic blacks only (n = 11,985)**

	Public-use		Relative hazard	Restricted-use		Relative hazard
	Coefficient	SE		Coefficient	SE	
Age in years	0.114 ***	0.0123	1.121	0.114 ***	0.0122	1.121
Sex (Female)						
Male	0.101	0.2201	1.106	0.098	0.2202	1.102
Race/ethnicity (NHW)						
NHB	0.404 *	0.1702	1.498	0.435 *	0.1689	1.545
Education (More than high school)						
Less than high school	-0.216	0.2910	0.806	-0.209	0.2896	0.811
High school	-0.139	0.2679	0.870	-0.137	0.2672	0.872

Notes:

Relative hazards are estimated from a Cox proportional hazards model.

All models use sample weights and take into account the NHANES III complex survey design using the SUDAAN software program (9.0.1).

NHW refers to non-Hispanic white; NHB refers to non-Hispanic black; the values in parentheses are reference categories.

\* p <.05; \*\* p<.01; \*\*\* p <.001.