Estimating smoking-attributable mortality in the United States: Geographic variation

Andrew Fenelon

Samuel H. Preston

Population Studies Center, University of Pennsylvania, 239 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19143 afenelon@sas.upenn.edu

Introduction

Tobacco is the largest single cause of premature death in the developed world and is growing in importance throughout the developing world. Recent estimates indicate that as many as 400,000 deaths annually in the United States (Mokdad, Marks, Stroup, & Gerberding, 2004) may result from cigarette smoking. At the individual level, cigarette smoking is strongly linked to lung cancer; but smoking also confers increased risk of death from other cancers, heart diseases, stroke, and pulmonary diseases, among other causes of death (Doll, Peto, Boreham, & Sutherland, 2004). Recent CDC estimates indicate that only around thirty percent of smoking-related deaths in the United States are caused by lung cancers.

Studies demonstrating the link between cigarette smoking and individual mortality typically involve detailed cohort smoking histories. The prospective study of British doctors beginning in 1951 (Doll et al., 2004) and the National Cancer Institute's Cancer Prevention Studies Cohorts I and II (CPS-I and CPS-II) beginning in 1959 and 1982, respectively, provide rich data on the excess risks associated with cigarette smoking from a number of causes of death. While cohort studies often provide the most persuasive evidence of the increased mortality risk related to cigarette smoking behavior, they suffer from a number of drawbacks. First, study cohorts may not be representative of the population-at-large and their experience may not reflect that of the general population of smokers. Second, smoking behavior categories may not reflect temporal changes in cohort smoking patterns, and many studies assume that baseline smoking status remains constant throughout the study. Finally, large-scale cohort studies require long periods of observation and detailed demographic information that may be unavailable for many relevant populations.

While cohort studies have been used to estimate the total number of deaths caused by cigarette smoking in a population (Rogers, Hummer, Krueger, & Pampel, 2005), others have attempted to deal with the problematic issues of cohort studies by using indirect methods. These studies, such as Peto, Lopez and colleagues (1992), use the lung cancer death rate as an indicator of the accumulated damage from smoking in the population. Based on estimates of the relative risk of smokers compared with non-smokers of mortality from various causes of death, one can calculate the total mortality that would not have occurred in the absence of smoking. Preston, Glei, and Wilmoth (2009) developed an alternative method which relies on the statistical relationship between lung cancer and other causes of death across countries and time periods. Their method makes fewer assumptions than Peto-Lopez and produces results that are highly similar, validating the robustness of both approaches. In this paper, we use the Preston, Glei, and Wilmoth method to calculate smoking-attributable mortality in the United States between 1990 and 2004. Rather than using the coefficients that they estimate from international data, however, we reestimate their equations using data from states of the United States. We then examine the extent to which cigarette smoking explains variation in adult mortality across U.S. and divisions. Finally, we compare the results of various methods for estimating smoking attributable mortality.

Background

Studies calculating the number of excess deaths due to cigarette smoking typically use an attributable-risk approach. They estimate the number of deaths that would not occur if smokers experienced the same death rates as non-smokers (Peto, Lopez, Boreham, Thun, & Heath, 1994). This requires some estimate of the increased risk conferred to smokers by their behavior; researchers have developed many different methods to calculate this excess risk and have applied them in a number of different settings.

The first set of methods could be termed *direct* methods, because the mortality differential between smokers and non-smokers is actually observed. These studies follow cohorts of smokers and non-smokers over periods of time to track the mortality experience of each respective population. In the United States, the most commonly cited study is the American Cancer Society Cancer Prevention Study Cohort II (CPS-II) which is composed of more than 1.2 million individuals followed since 1982. The Centers for Disease Control and Prevention (CDC) issues regular estimates of smoking-attributable mortality in the U.S. using relative risks from CPS-II (Adhikari, Kahende, Malarcher, Husten, & Asman, 2009). Based on excess mortality among current smokers and former smokers relative to non-smokers, they calculate mortality attributable to cigarette smoking by applying relative risks to current smoking status data. They find that 440,000 annual deaths can be attributed to cigarette smoking in the late 1990s and early 2000s, more than one-fifth of all adult mortality.

The CDC approach has received criticism from those arguing that the relative mortality risks of smokers compared with non-smokers reflect more than just the effect of smoking. The assumption that smokers would have the same mortality experience as non-smokers in the absence of smoking ignores other behavioral and socioeconomic differences between smokers and non-smokers that may confer increased risk. Rogers et al. (2005) use the National Health Interview Survey 1990 supplement to control for many covariates of smoking behavior, which decreases the estimated number of deaths to 338,000.

While direct methods that control for smoking confounders offer relatively robust assessments of the mortality consequences of smoking, such extensive data on smoking status and mortality is unavailable for most populations, and a more widely-applicable method is necessary. Furthermore, current-status smoking data used in these methods may not represent

effective measures of exposure to smoking-related mortality. Peto, Lopez, and colleagues (1992) developed an *indirect* method for calculating smoking attributable mortality that relies exclusively on vital statistics data. Assuming that smoking behavior is the only factor which increases the risk of *lung cancer* death of smokers relative to non-smokers, they use CPS-II non-smoker lung cancer death rates to calculate age-specific 'proportion exposed' that reflects the prevalence of smoking-related damage. They then import relative risks for various disease categories from CPS-II and apply them to exposure composition of the population. In order to correct for confounding, they decrease the relative risks from causes of death other than lung cancer by half. They produced estimates of smoking-attributable mortality for the U.S. among other developed countries for the year 2000 (Peto, Lopez, Boreham, Thun, & Heath, 2006). Although the lung cancer death rate does not directly measure the prevalence or intensity of smoking in a population, it may be a more reliable indicator of exposure to smoking-related damage than are self-reports and current-status surveys.

In practice the Peto-Lopez method is complex to implement and relies heavily on the relative risks from CPS-II. Moreover, it makes a rather arbitrary assumption regarding the confounding of smoking with other factors in producing cause-specific relative risks, simply halving them.[1] Preston, Glei and Wilmoth (2009) developed an alternative method to Peto-Lopez which makes fewer assumptions and does not rely heavily on the generalizability of CPS-II relative risks. Their method develops a statistical model to estimate the relationship between the lung cancer death rate and the death rate from other causes of death across developed countries between 1950 and 2003. This relation is then used to estimate the mortality impact of smoking. They use lung cancer death rates of non-smokers in CPS-II between 1982 and 1988

[1] Other studies using Peto-Lopez have relaxed this assumption, decreasing observed relative risks by different amounts or actually attempting to control for differences between smokers and non-smokers .

(M. Thun et al., 1997) to produce an estimate of lung cancer attributable-risk and apply this each population of interest. The method produces results that are highly similar to those of Peto and Lopez without relying on strong assumptions regarding the relationship between smoking and cause-specific mortality.

A key advantage of Preston, Glei, and Wilmoth's is its applicability to varied populations. It can be implemented anywhere vital statistics data exists for a number of populations or geographic units. In this paper, we apply the method using annual mortality data at the U.S. state level between 1990 and 2004. Based on the results of the model, we calculate smoking attributable mortality for the United States as a whole as well as for the nine Census divisions. Rates of cigarette smoking in the United States have been historically very high, decreasing only in recent years for males. But large regional differences in cigarette smoking behavior and related mortality remain. A few descriptive studies have examined the issue of geographic variation in smoking-attributable mortality in the United States (CDC) but none has fully explored the implications of smoking for geographic disparities in U.S. adult mortality. Following other recent studies demonstrating the potential for differences in smoking to explain mortality differentials (Jha et al., 2006; Preston & Wang, 2006), we demonstrate the impact of smoking-related mortality on regional patterns of mortality in the U.S. Finally, we compare attributable fraction estimates for the U.S. produced by a variety of different methods.

Data

We use vital statistics data for the fifty states annually between 1990 and 2004. Death data are available through the Multiple Cause-of-Death (MCD) public-use micro-data files released annually by the National Center for Health Statistics (NCHS). MCD files contain demographic, geographic, and cause-of-death information about all deaths occurring in the

United States. Population denominators for death rate calculations come from bridged-race files available from the NCHS.[2] Mortality data by Census division are calculated by aggregating based on scheme in Table A1. We restrict our analyses to ages 50+ in order to capture the vast majority of smoking-related deaths.

Method

Lung cancer is a unique condition in that it is so closely tied to one behavioral risk factor. While other causes of death have been shown to be linked to smoking behavior, none is related as strongly as is lung cancers. In CPS-II, smoking was responsible for more than 90% of lung cancer deaths among men and more than 70% among women (M. Thun et al., 1997). In places where reliable cohort smoking histories are unavailable for most of the population, the age-specific death rate from lung cancer has been used as an indirect indicator of the accumulated damage from cigarette smoking (Peto et al., 1992; Preston et al., 2009). If we then assume that variation in lung cancer death rates is almost exclusively the result of variation in cigarette smoking, the use of the lung cancer death rate as a population-level index of the impact of smoking is robust (see Preston et al. 2009 for a discussion of these issues).

Statistical Model

We use a variation of the method developed by Preston et al. (2009) to calculate smoking attributable mortality in the United States for the years 1990, 2000, and 2004. We estimate the relationship between the age-specific lung cancer death rate and the log of the death rate from other causes of death annually between 1990 and 2004. We use Poisson regression to predict the logarithm of deaths to causes other than lung cancer in five-year age groups from 50 – 54 to 80+ as a function of the death rate from lung cancer

$$\ln(M_O) = \beta_L M_L + \beta_a X_a + \beta_S X_S + \beta_t X_t + \beta_{tS}(T \times X_S) + \beta_{aL}(M_L \times X_a)$$

[2] Electronically from http://www.cdc.gov/nchs/nvss/bridged_race.htm, accessed May, 2009.

where M_L and M_O are the death rate for lung cancer and other causes respectively in each state, year, and five-year age group. X_a, X_S, and X_t are dummy variables for age-group, state, and year, respectively, while are their corresponding coefficients. We also include interactions between state and year (treated as linear) and between lung cancer and age group. We use age-specific population counts as a statistical "offset" in the procedure to control for exposure to mortality. We estimate separate models for males and females to allow for highly distinct relationships between smoking and mortality. The coefficients of interest are β_L and β_{aL}, denoting the age-specific relationship between lung cancer and other causes of death ($\beta_L^{'} = \beta_L + \beta_{aL}$); these coefficients are used to calculate the attributable fraction.

Attributable Fraction

Lung cancer deaths attributable to cigarette smoking are estimated using values of lung cancer death rates among never-smokers, reported by Thun et al. (1997) from the CPS-II study between 1982 and 1988. The proportion of lung cancer deaths attributable to smoking is the ratio of smoking-related lung cancer death rate to the overall lung cancer death rate

$$A_L = \frac{M_L - M_L^*}{M_L}$$

where M_L^* is the lung cancer death rate among lifelong non-smokers, the expected death rate in the absence of smoking. While lung cancer mortality among never smokers does show some variation across populations (M. J. Thun et al., 2008), there is little evidence for long-term changes across periods (Rosenbaum, Sterling, & Weinkam, 1998). The relationship between the prevalence of cigarette smoking and lung cancer deaths among non-smokers is unclear.

We calculate mortality attributable to smoking for causes of death other than lung cancer based on the relationship between lung cancer and other causes across states. First, we calculate the predicted number of deaths expected based on the observed lung cancer death rate. Next, we

subtract the predicted number of deaths expected from the lung cancer death rates of non-smokers. We divide this difference by number of deaths expected from the observed rates to calculate the attributable fraction

$$A_O = \frac{e^{\beta_L^{'}(M_L)} - e^{\beta_L^{'}(M_L^*)}}{e^{\beta_L^{'}(M_L)}}$$

where $\beta_L^{'}$ is the model coefficient for lung cancer including age interactions ($\beta_L^{'} = \beta_L + \beta_{aL}$). The attributable fraction for total mortality is a weighted average of the attributable fractions for lung cancer and other causes

$$A = \frac{A_L D_L + A_O D_O}{D}$$

where D_L and D_O are deaths from lung cancer and other causes respectively and D is total deaths. In their application, Preston et al (2009) find that the estimated attributable fraction is generally robust to alternative specifications of age, time period, and interactions. Lung cancer also exhibits no statistically significant relationship with external causes of death which are assumed to be unrelated to smoking (Peto et al. 1992).

Variation in Mortality by U.S. Division

We estimate smoking-attributable mortality for ages 50+ for the United States as well as the nine U.S. Census Divisions. We calculate life expectancy at age 50 both including and excluding smoking-related deaths. Age-specific death rates in the absence of smoking (M_{abs}) include only those deaths not attributed smoking by our model

$$M_{abs} = \frac{D - D_A}{P}$$

where D_A the number of deaths attributed to smoking and P is is the number of person-years of exposure. We then recalculate life tables for each Division with smoking-related deaths removed.[3]

Results

Table 1 shows estimated coefficients of the relationship between lung cancer and other causes by age-group and sex. If exponentiated, they can be interpreted as the increase in the natural logarithm of the death rate of other causes resulting from an increase in the lung cancer death rate of one per thousand, all else being equal. Coefficients are smaller at higher ages, reflecting both higher death rates overall and more varied factors influencing mortality at higher ages. Assumed lung cancer death rates among lifelong non-smokers from CPS-II are presented in Table 2. Given that we assume smoking to be the single factor determining population variation in lung cancer death rates, the rates in Table 2 are intended to approximate conditions in which smoking was eliminated. The difference between these rates and observed lung cancer rates is used to calculate lung cancer attributable risk. We find that more than 80% of lung cancer deaths among women and more than 90% among men can be attributed to smoking.

Smoking attributable fractions for the U.S. and by Division are displayed in Table 3 for 1990, 2000, and 2004. Between 1990 and 2004, women experienced large increases in smoking-related mortality while men experienced decreases. In 1990, smoking was responsible for nearly 30% of deaths among men over age 50 compared with less than 15% among women. By 2004, it had decreased to 22.0% for men and increased to 19.0% for women. This finding is consistent with previous estimates of smoking-related mortality in the U.S. as well as studies demonstrating narrowing sex mortality differentials related to cohort changes in smoking behavior (Preston and

[3] We elect to simply 'remove' smoking deaths from the life table calculation as opposed to using 'cause-deleted' life tables in order to preserve the simplicity of interpretation. The results do not change substantively.

Wang). Historically high rates of smoking among American men have begun to decline, while rates of American women have become more similar to men's.

There was little variation in attributable fraction by division among women in 1990. By 2004, the Mountain division had begun to emerge with a substantially lower fraction, actually declining from 18.06% in 2000 to 16.53% and 2004. While rates of smoking have increased for women, divisional differences in the impact of smoking are particularly striking for men. In 1990, males in the East South Central and West South Central Divisions had attributable fractions greater than one-third; in the Mountain and Pacific Divisions they were about one-quarter. When considered alongside the high mortality already present the in the Southern states, the enormous impact of smoking on these areas is even more salient. While smoking-attributable mortality has declined in each division in recent years, areas in the Southern U.S. continue to be plagued by it. In 2004, smoking accounted for fully 30% of deaths in the East South Central Division compared with only 16% and 17% in the Mountain and Pacific Divisions respectively.

The impact of cigarette smoking on geographic variation in U.S. mortality is equally informative. Table 4 presents life expectancy at age 50 both including and excluding smoking-related deaths. These values give an indication of the number of years lost to smoking-related conditions among older Americans. In 1990, female life expectancy was two years shorter as a result of smoking while male life expectancy was nearly four years shorter; in the absence of smoking, male life expectancy at age 50 would rise by roughly 17%. This value increased to 2.9 years for women and decreased to 3.0 years for men by the year 2004. Among women in 2004, the greatest impact of smoking was found in the New England and Pacific Divisions (3.1 years lost) while the smallest was in the Mountain division (2.5 years). Among men, the East South

Central Division is particularly strongly affected by the smoking epidemic. In 1990, life expectancy in this division would increase by 4.8 years, or by 23%, in the absence of smoking. The division also lost nearly 4.0 years to smoking-related diseases in 2004. The Mountain and Pacific Divisions perform particularly well with respect to smoking, each losing only slightly more than 2 years in 2004.

Comparison with Alternative Methods

Researchers have developed a number of direct and indirect methods for estimating the number of deaths in a population attributable to cigarette smoking. Different methods place varying emphasis on assumptions regarding the relationship between smoking and mortality at the individual level, some assuming that all excess mortality among smokers is the result of smoking while others attempt to control for potential confounders. Table 5 shows estimated U.S. smoking-attributable fractions using many different methods. The first row shows our estimates of smoking-attributable fraction for U.S. men and women in 1990 and 2004. Row 2 shows estimates obtained using coefficients found by Preston, Glei, and Wilmoth (2009) using the same method across a sample of 20 developed countries. While our estimates are very similar to theirs for men, our estimates for women are substantially lower (0.190 vs. 0.247 in 2004). Row 3 shows estimates using the Peto-Lopez method reported in Peto et al. (2006) for ages 35+, which are generally similar to ours. Row 4 shows attributable fractions using the method employed by CDC, which are comparable to ours except for women in 2004 (0.153). The estimates made by Rogers et al. (2005) using NHIS data are substantially lower than our estimates for females (0.126) and quite similar to ours for males (0.212).

Discussion and Conclusion

Cigarette smoking is the single most important cause of premature death in the United States, accounting for more than one-fifth of all adult deaths in 2004. In the absence of deaths from smoking, U.S. life expectancy would be almost three years longer. Given its importance, researchers have developed a number of methods attempting to calculate the excess mortality burden resulting from cigarette smoking. Cohort studies that track individuals with respect to smoking behavior and mortality provide informative evidence about the link between cigarette use and mortality, but suffer from imprecise classification of smoking status and may ignore population-based exposure to smoking-related mortality. Moreover, detailed longitudinal studies of smoking may be unavailable for many relevant populations. To remedy some of these deficiencies, Peto et al. (1992) developed an indirect method which estimates the attributable fraction using lung cancer as a marker of smoking in developed-country populations. While their method makes some fairly arbitrary assumptions, it has been widely applied in populations where cohort smoking histories do not exist.

In this paper, we applied an alternative indirect method to estimating mortality attributable to cigarette smoking in the United States. We calculated attributable fractions in 1990, 2000, and 2004 for the U.S. as a whole as well as the nine Census Divisions in order to examine geographic differences in smoking-related mortality. Our estimates indicate that slightly more than 20% of all adult deaths in the United States in 2004 were caused by smoking, a figure which is highly consistent with previous estimates. We find further support for lung cancer as a robust indicator of population exposure to cigarette smoking and for the use of indirect methods to estimate smoking-attributable mortality.

Geographic differences in smoking-attributable mortality should reflect historical differences in cigarette smoking. Very little smoking prevalence data is available at subnational

levels of aggregation prior to 1985, but existing data match very closely to our patterns of estimated smoking-related mortality. Shopland et al. (1996) find the highest smoking prevalence for 1992-1993 in the East South Central, West South Central, and East North Central Divisions. The lowest prevalences are found in the Pacific, New England, and Mid-Atlantic Divisions. These regional patterns were also relatively constant between 1985 and 1992-1993 (Shopland, Niemcryk, & Marconi, 1992). Other studies have connected state-specific cancer patterns to past prevalence of smoking (Jemal et al., 2006). Evidence from the past half century indicates that heavy smoking areas also experience high mortality from smoking-related conditions, especially lung cancer. As the smoking epidemic took hold among men in the southern states, rates of male lung cancer death increased rapidly there, producing a high concentration of smoking related mortality (CDC, 2009; Devesa, Grauman, Blot, & Fraumeni, 1999).

Since the 1980s, U.S. life expectancy has lagged substantially behind that of its European counterparts, and smoking has been a significant factor (Preston et al. 2009). In the coming decades, smoking-related mortality promises to increase in importance as women smoke more heavily (Preston and Wang 2006). At the same time, countries experiencing more rapid declines in smoking among men may emerge as world longevity leaders. As these processes unfold, robust methods for estimating the mortality burden of cigarette smoking across a range of populations. In countries with relatively mature smoking epidemics, where the vast majority of lung cancer cases are attributable to smoking, our method provides reasonable and stable estimates of the impact of smoking on adult mortality. Within the United States, we find that our estimates of smoking-attributable mortality match very close to historical smoking prevalence data. The similarity of our results to those using other methods provides additional evidence for the validity of our approach.

Table 1: Model Coefficients for lung cancer death rate by age and sex

Age	Male	Female
50-54	0.319	0.370
55-59	0.213	0.291
60-64	0.145	0.151
65-69	0.105	0.108
70-74	0.077	0.060
75-79	0.058	0.050
80+	0.018	0.094

Estimated using Poisson regression in Equation (1). Includes controls and age interactions. The exponential of the above coefficients represents the proportional increase in the death rate for other causes associated with a one-per-thousand increase in the lung cancer death rate.

Table 2: Assumed lung cancer death rates of lifelong nonsmokers (per 1,000)

Age	Male	Female
50-54	0.06	0.06
55-59	0.05	0.07
60-64	0.12	0.12
65-69	0.22	0.17
70-74	0.35	0.31
75-79	0.52	0.33
80-84	0.89	0.58
85+	0.87	0.61

From Thun et al. (1997) for death rates of never smokers in the Cancer Prevention Study, Cohort II 1982 - 1988.

Table 3: Fraction of Deaths attributable to Cigarette Smoking by Division: 1990 - 2004

	Female			Male		
Division	1990	2000	2004	1990	2000	2004
New England	0.142	0.199	0.201	0.275	0.223	0.207
Mid-Atlantic	0.147	0.180	0.187	0.280	0.216	0.196
East North Central	0.136	0.178	0.191	0.297	0.244	0.232
West North Central	0.116	0.159	0.175	0.268	0.226	0.224
South Atlantic	0.153	0.190	0.196	0.325	0.257	0.238
East South Central	0.136	0.181	0.202	0.359	0.318	0.301
West South Central	0.144	0.183	0.195	0.330	0.262	0.246
Mountain	0.142	0.181	0.165	0.253	0.205	0.161
Pacific	0.159	0.189	0.191	0.256	0.189	0.176
Total US	**0.143**	**0.182**	**0.191**	**0.297**	**0.237**	**0.221**

Table 4: Life expectancy at age 50 before and after the removal of smoking deaths by Division: 1990 - 2004

	Females						Males					
	1990			2004			1990			2004		
	Without Smoking	With Smoking	Diff.	Without Smoking	With Smoking	Diff.	Without Smoking	With Smoking	Diff.	Without Smoking	With Smoking	Diff.
New England	29.87	27.87	2.0	32.32	29.25	3.1	26.37	22.94	3.4	28.72	25.90	2.8
Mid-Atlantic	29.42	27.38	2.0	31.90	29.07	2.8	26.04	22.54	3.5	28.20	25.54	2.7
East North Central	29.20	27.32	1.9	31.19	28.32	2.9	26.10	22.38	3.7	27.89	24.78	3.1
West North Central	30.01	28.33	1.7	31.50	28.90	2.6	26.36	22.89	3.5	28.37	25.37	3.0
South Atlantic	29.96	27.79	2.2	31.70	28.71	3.0	26.60	22.33	4.3	28.19	24.96	3.2
East South Central	28.83	26.93	1.9	29.74	26.88	2.9	25.87	21.06	4.8	27.09	23.13	4.0
West South Central	29.39	27.38	2.0	30.52	27.71	2.8	26.36	22.03	4.3	27.47	24.24	3.2
Mountain	29.47	27.53	1.9	32.01	29.52	2.5	25.75	22.62	3.1	28.28	26.18	2.1
Pacific	30.42	28.22	2.2	32.98	29.87	3.1	27.04	23.87	3.2	28.75	26.35	2.4
Total U.S.	29.71	27.71	2.0	31.70	28.80	2.9	26.42	22.63	3.8	28.23	25.25	3.0

Table 5: Mortality attributable to cigarette smoking in the U.S.: A comparison of estimates

	Females		Males	
	1990	2004	1990	2004
Current Model[1]	0.143	0.190	0.296	0.220
Preston, Glei, Wilmoth[2]	0.183	0.247	0.297	0.238
Peto-Lopez[3]	0.176	0.205[†]	0.277	0.243[†]
CDC Method[4]	0.140	0.153[*]	0.278	0.235[*]
Rogers[5]	—	0.126[†]	—	0.212[†]

[1]Coefficient estimates across 50 U.S. states, 1990 - 2004, ages 50+

[2]Using coefficient estimates across 20 developed countries, 1950 - 2003, ages 50+

[3]Ages 35+. Peto-Lopez estimates from (http://www.ctsu.ox.ac.uk/deathsfromsmoking)

[4]Estimates reported by the Centers for Disease Control and Prevention (2008), ages 35+

[5]Figures reported in Rogers et al. (2005) for the year 2000, ages 35+

[†] Estimates based on data for 2000

[*] Estimates based on data for the period 2000-2004

Table A1: Divisions and states

New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
Mid-Atlantic	New Jersey, New York, Pennsylvania
East North Central	Illinois, Indiana, Michigan, Ohio, Wisconsin
West North Central	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
South Atlantic	Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
East South Central	Alabama, Kentucky, Mississippi, Tennessee
West South Central	Arkansas, Louisiana, Oklahoma, Texas
Mountain	Arizona, Colorado, Idaho, Montana, New Mexico, Utah, Wyoming
Pacific	Alaska, California, Hawaii, Nevada[1], Oregon, Washington

[1] Included with Pacific division as opposed to Mountain division since it shares more in common culturally and socially with the Pacific than with the Mountain states. This change does not alter the conclusions substantially.

References

Adhikari, B., Kahende, J., Malarcher, A., Husten, C., & Asman, K. (2009). State-Specific Smoking-Attributable Mortality and Years of Potential Life Lost-United States, 2000-2004 (Reprinted from MMWR, vol 58, pg 29-33, 2009). *Jama-Journal of the American Medical Association, 301*(9), 928-929.

CDC. (2009). State-specific smoking-attributable mortality and years of potential life lost--United States, 2000-2004. *MMWR Morb Mortal Wkly Rep, 58*(2), 29-33.

Devesa, S. S., Grauman, D. J., Blot, W. J., & Fraumeni, J. F. (1999). Cancer surveillance series: Changing geographic patterns of lung cancer mortality in the United States, 1950 through 1994. *Journal of the National Cancer Institute, 91*(12), 1040-1050.

Doll, R., Peto, R., Boreham, J., & Sutherland, I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal, 328*(7455), 1519-1528.

Ezzati, M., & Lopez, A. D. (2003). Measuring the accumulated hazards of smoking: global and regional estimates for 2000. *Tobacco Control, 12*(1), 79-85.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J. Q., Smigal, C., et al. (2006). Cancer statistics, 2006. *Ca-a Cancer Journal for Clinicians, 56*(2), 106-130.

Jha, P., Peto, R., Zatonski, W., Boreham, J., Jarvis, M. J., & Lopez, A. D. (2006). Social inequalities in male mortality, and in male mortality from smoking: indirect estimation from national death rates in England and Wales, Poland, and North America. *Lancet, 368*(9533), 367-370.

Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2004). Actual causes of death in the United States, 2000. *Jama-Journal of the American Medical Association, 291*(10), 1238-1245.

Peto, R., Lopez, A. D., Boreham, J., Thun, M., & Heath, C. (1992). MORTALITY FROM TOBACCO IN DEVELOPED-COUNTRIES - INDIRECT ESTIMATION FROM NATIONAL VITAL-STATISTICS. *Lancet, 339*(8804), 1268-1278.

Peto, R., Lopez, A. D., Boreham, J., Thun, M., & Heath, C. (1994). *Mortality from Smoking in Developed Countries 1950-2000*. Oxford: Oxford University Press.

Peto, R., Lopez, A. D., Boreham, J., Thun, M., & Heath, C. (2006). *Mortality from smoking in developed countries 1950-2000 (2nd edition)*. Oxford: Oxford University Press.

Preston, S. H., Glei, D. A., & Wilmoth, J. R. (2009). *Contribution of Smoking to International Differences in Life Expectancy.* Paper presented at the Panel on Divergent Trends in Life Expectancy, National Research Council.

Preston, S. H., & Wang, H. D. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography, 43*(4), 631-646.

Rogers, R. G., Hummer, R. A., Krueger, P. M., & Pampel, F. C. (2005). Mortality attributable to cigarette smoking in the United States. *Population and Development Review, 31*(2), 259-+.

Rosenbaum, W. L., Sterling, T. D., & Weinkam, J. J. (1998). Use of multiple surveys to estimate mortality among never, current, and former smokers: Changes over a 20-year interval. *American Journal of Public Health, 88*(11), 1664-1668.

Shopland, D. R., Hartman, A. M., Gibson, J. T., Mueller, M. D., Kessler, L. G., & Lynn, W. R. (1996). Cigarette smoking among US adults by state and region: Estimates from the current population survey. *Journal of the National Cancer Institute, 88*(23), 1748-1758.

Shopland, D. R., Niemcryk, S. J., & Marconi, K. M. (1992). GEOGRAPHIC AND GENDER VARIATIONS IN TOTAL TOBACCO USE. *American Journal of Public Health, 82*(1), 103-106.

Thun, M., Day-Lally, C., Myers, D. G., Calle, E. E., Flanders, W. D., & Zhu, B. P. (1997). Trends in tobacco smoking and mortality from cigarette use in cancer prevention studies I (1959 through 1965) and II (1982 through 1988). In D. Burns, L. Garfinkel & J. Samet (Eds.), *Changes in cigarette-related disease risks and their implications for prevention and control, smoking and tobacco control. Smoking and tobacco control monograph no. 8, NIH publication no. 97-4213 (pp. 305-382)*. Bethesda, MD: Cancer Control and Population Sciences, National Cancer Institute, U.S. National Institutes of Health.

Thun, M. J., Hannan, L. M., Adams-Campbell, L. L., Boffetta, P., Buring, J. E., Feskanich, D., et al. (2008). Lung Cancer Occurrence in Never-Smokers: An Analysis of 13 Cohorts and 22 Cancer Registry Studies. *Plos Medicine, 5*(9), 1357-1371.