Smoking-attributable mortality by cause of death in the United States: An indirect approach

Joseph T. Lariscy

Department of Sociology, University of Memphis, 223 Clement Hall, Memphis, TN 38152, USA

ABSTRACT

More than 50 years after the U.S. Surgeon General’s first report on cigarette smoking and mortality, smoking remains the leading cause of preventable death in the United States. The first report established a causal association between smoking and lung cancer, and subsequent reports expanded the list of smoking-attributable causes of death to include other cancers, cardiovascular diseases, stroke, and respiratory diseases. For a second level of causes of death, the current evidence is suggestive but not sufficient to infer a causal relationship with smoking. This study draws on 1980–2004 U.S. vital statistics data and applies a cause-specific version of the Preston-Glei-Wilmoth indirect method, which uses the association between lung cancer death rates and death rates for other causes of death to estimate the fraction and number of deaths attributable to smoking overall and by cause. Nearly all of the established and additional causes of death are positively associated with lung cancer mortality, suggesting that the additional causes are in fact attributable to smoking. I find 420,284 annual smoking-attributable deaths at ages 50+ for years 2000–2004, 14% of which are due to the additional causes. Results corroborate recent estimates of cause-specific smoking-attributable mortality using prospective cohort data that directly measure smoking status. The U.S. Surgeon General should reevaluate the evidence for the additional causes and consider reclassifying them as causally attributable to smoking.

1. Introduction

More than 50 years after the U.S. Surgeon General’s first report on cigarette smoking and mortality, smoking remains the leading cause of preventable death in the United States (GBD 2015 Tobacco Collaborators, 2017; Rogers, Hummer, Krueger, & Pampel, 2005; U.S. Department of Health and Human Services [U.S. DHHS] 2014). The first Surgeon General’s report concluded that evidence was sufficient to infer that smoking is causally linked to lung cancer and chronic bronchitis among men and possibly linked to lung cancer among women (U.S. Department of Health, Education, and Welfare 1964). Subsequent reports expanded the list of smoking-attributable causes of death to include cancers in several organs, cardiovascular diseases, stroke, respiratory diseases, and diabetes (U.S. DHHS 2014). In addition to the first level of causes of death for which a causal association with smoking is established, a second level includes additional causes of death for which evidence is “suggestive but not sufficient” to infer a causal relationship with smoking (U.S. DHHS 2014:3). Research has found elevated mortality risk from these causes among smokers relative to nonsmokers and identified routes by which the carcinogens and other toxic substances in cigarette smoke could harm those parts of the body. However, the evidence for these additional causes of death does not yet meet the Surgeon General’s criteria for causality.

Two recent studies provide compelling evidence that these additional causes may in fact be attributable to smoking (Carter et al., 2015; Lariscy, Hummer, & Rogers, 2018). Both studies find elevated risk of death among current and former smokers compared with never smokers for established and additional causes. These studies (and many others assessing smoking-attributable mortality) use prospective cohort data, in which survey respondents report their smoking behavior, and then vital status is ascertained years later. This approach faces criticism because survey respondents generally report current cigarette use without providing lifelong smoking histories. A population’s smoking-attributable deaths are a result of smoking patterns from 20–30 years ago rather than current smoking prevalence (Lopez, Collishaw, & Piha, 1994; Pampel, 2005). Alternative approaches indirectly assess a population’s cumulative smoking burden using lung cancer mortality rates (Peto, Lopez, Boreham, Thun, & Heath, 1992, 1994; Preston, Glei, & Wilmoth, 2010a, 2010b). Additional techniques then determine the proportion of smoking-related deaths from causes of death other than lung cancer. Indirect approaches may be better able to measure cohorts’ full smoking histories and to account for the lag between a population’s adoption of smoking and subsequent smoking-attributable mortality.

This study applies a cause-specific version of the indirect method...
developed by Preston et al. (2010a, 2010b) to determine whether the established and additional causes of death are attributable to smoking among U.S. adults ages 50+. Specifically, I use the association between lung cancer death rates and death rates from each established and additional cause to assess whether causes are attributable to smoking, and if so, what fraction and number of deaths from each cause are attributable to smoking. Accurate estimation of smoking-attributable mortality by cause of death is important to determine the contribution of smoking to sociodemographic disparities in population health (Denney, Rogers, Hummer, & Pampel, 2010; Lariscy, Hummer, & Hayward, 2015; Pampel, 2005) and the extent to which smoking explains inequalities in life expectancy in international comparisons (Crimmins, Preston, & Cohen, 2011; Rostron & Wilmoth, 2011).

2. Background

2.1. Estimating smoking-attributable mortality: Direct method

Two methodological approaches are used to estimate smoking-attributable mortality. One approach (hereafter, “direct method”) uses survey data that directly measure smoking status and are prospectively linked to death records to compare mortality risk among current and former smokers with nonsmokers. The direct approach allows researchers to statistically adjust for individual characteristics that contributed to the smoking-mortality association such as sociodemographic factors and other health risk behaviors. Rogers et al. (2005) used an early version of the National Health Interview Survey Linked Mortality Files (NHIS-LMF) data to compare all-cause adult mortality risk of various smoker groups defined by different intensities and durations with the risk among never smokers. They showed that mortality risk is highest among current heavy smokers and that smoking cessation reduces mortality risk.

Studies that use the direct method to examine specific causes of death generally focus on broad cause categories or only the causes established by the U.S. Surgeon General as being causally linked to smoking (Doll, Peto, Boreham, & Sutherland, 2004; Hummer, Niam, & Rogers, 1998; Rostron, 2013). These established causes include cancers of the lung, lip and oral cavity, esophagus, stomach, colon and rectum, liver, pancreas, larynx, cervix, urinary bladder, kidney and renal pelvis, and bone marrow; ischemic heart disease; other heart diseases; stroke; atherosclerosis; aortic aneurysm; other arterial diseases; pneumonia, influenza, and tuberculosis; chronic obstructive pulmonary disease (COPD); and diabetes (Doll et al., 2004; International Agency for Research on Cancer, 2004; U.S. DHHS 2014).

Two recent studies go beyond the established causes of death to examine additional causes for which the current evidence is suggestive but not sufficient to infer a causal association with smoking. Carter et al. (2015) compiled a list of such additional causes that includes infections, breast cancer, prostate cancer, rare cancers, cancers of unknown site, hypertensive heart disease, essential hypertension and hypertensive renal disease, respiratory diseases (other than pneumonia and influenza), intestinal ischemia, liver cirrhosis, other digestive diseases, renal failure, additional rare causes combined, and deaths from unknown causes. They found that smokers exhibit higher risk of mortality from these causes than never smokers, net of sociodemographic and behavioral confounders. However, they analyzed five cohort samples that are disproportionately non-Hispanic white and college-educated. Lariscy et al. (2018) replicated and extended the study by Carter and colleagues by using nationally-representative data (i.e., 1990–2011 NHIS-LMF) and controlling for additional risk factors.

Despite the insights obtained with the direct approach, it encounters several issues. First, direct methods measure respondents’ current smoking status, whereas smoking-attributable mortality results from accumulated smoking burden. Most smokers initiate in adolescence or young adulthood, but smoking-attributable morbidity and mortality develop decades later (Pampel, 2005). Second, with cross-sectional survey data linked to death records, respondents report their smoking status at baseline, and changes in smoking status are not observed. Mortality risk ratios may be biased if respondents’ smoking status changes during the vital status follow-up period, whether nonsmokers begin smoking, current smokers quit, or former smokers relapse (Taylor, Hasselblad, Henley, Thun, & Sloan, 2002).

2.2. Estimating smoking-attributable mortality: Indirect method

The second approach for estimating smoking-attributable mortality (hereafter, “indirect method”) uses a population’s lung cancer mortality rate as a proxy for their cumulative tobacco burden. Lung cancer serves as an indicator of a population’s smoking burden because smoking is the main source of variation in lung cancer mortality, and 70–90% of lung cancer deaths occur among smokers (Ezzati & Lopez, 2003; Thun et al., 1997). In contrast, lung cancer deaths are rare among lifetime nonsmokers (Thun et al., 2008). Indirect methods may be better able to measure cohorts’ full smoking histories than direct methods.

One indirect approach developed by Peto et al. (1992, 1994; hereafter, “Peto-Lopez method”) uses the difference between a population’s lung cancer death rate and the lung cancer death rate among lifetime nonsmokers to estimate the proportion of a population exposed to smoking. The method imports smoker-nonsmoker death rate ratios for upper aerodigestive cancers, other cancers, COPD, and other medical causes from the Cancer Prevention Study II (CPS-II) and applies them to the population exposed to smoking. The cause-specific rate ratios are reduced by half to account for possible confounding. No deaths from liver cirrhosis, accidents, or violence are counted as attributable to smoking. Researchers have used the Peto-Lopez method to estimate the global impact of tobacco, examine mortality variation by country, and project future trends in smoking-attributable mortality (Peto, Lopez, Pan, Boreham, & Thun, 2015).

Another indirect method developed by Preston et al. (2010a, 2010b; hereafter, “PGW method”) relies on the statistical association between lung cancer mortality and mortality from all other causes of death combined to estimate smoking-attributable mortality. Specifically, the PGW method uses negative binomial regression to model death rates from all causes other than lung cancer combined as a function of lung cancer death rates and control variables (age, year, country, and interaction terms). They used data by sex and 5-year age group for 20 high-income countries at advanced stages of the tobacco epidemic during 1950–2006. Cause-specific death counts came from the World Health Organization (WHO) Mortality Database and all-cause death counts, population exposures, and death rates came from the Human Mortality Database. This macrolevel model-based approach avoids the need to import risk ratios from CPS-II. The resulting coefficients are translated into smoking-attributable fractions (SAFs) to indicate the percentage of deaths due to smoking. Fanelon and Preston (2012) re-estimated the PGW coefficients using U.S. data and controlling for state of residence. The U.S. coefficients are better suited for the mature stage of the tobacco epidemic occupied by the United States and correct for implausibly high smoking-attributable mortality among older women that stems from using the international data. Using U.S. coefficients, Fanelon and Preston (2012) found that 21% (17%) of male (female) deaths were attributable to smoking in 2004. To date, researchers have used the PGW method to make international comparisons of smoking-attributable mortality (Kelly & Preston, 2016; Preston et al., 2010a, 2010b; Rostron, 2010) and examine the contribution of smoking to mortality disparities by race/ethnicity (Blue & Fanelon, 2011; Ho & Elo, 2013), educational attainment (Ho & Fanelon, 2015), and geographic context (Fanelon, 2013; Fanelon & Preston, 2012).
2.3. Applying indirect estimation to smoking-attributable causes of death

Preston et al. (2010a, 2010b) developed their method to examine the association between lung cancer mortality and mortality from other causes of death combined, without the need to specify individual causes of death other than lung cancer. As a sensitivity analysis to validate their approach, Preston et al. (2010a) examined the association between lung cancer death rates and death rates for a few broad cause-of-death categories. They showed that lung cancer mortality is strongly correlated with mortality from respiratory diseases and smoking-related cancers (i.e., categories established as causally associated with smoking) and weakly or negatively correlated with mortality from other cancers and external causes (i.e., categories not associated with smoking). I extend their analysis by examining the correlation between lung cancer death rates and death rates for specific causes of death, including the established causes acknowledged by the U.S. Surgeon General and the additional causes Carter and colleagues (2015) compiled. In sum, my application of the PGW indirect method to specific causes of death offers a unique opportunity to assess whether the additional causes of death are attributable to smoking and to estimate the fraction and number of excess deaths due to smoking overall and by cause in the United States.

3. Material and methods

3.1. Data

This study drew on vital statistics data for U.S. adults aged 50+ in 1980–2004. Mortality data came from the National Bureau of Economic Research (NBER 2016), which harmonizes Multiple Cause-of-Death files available from the National Center for Health Statistics (NCHS 2004). Population denominators for death rates came from the Surveillance, Epidemiology, and End Results (SEER) Program, National Cancer Institute (NCI 2016), which harmonizes data from the U.S. Census Bureau. I calculated lung cancer death rates (per 1000) for sex-, age-, period, and U.S. division-specific groups. Analyses were based on about 48 million deaths and over 1.7 billion person-years.

3.2. Measures

Cause-specific models adjusted for age, period, and U.S. division of residence. I stratified analyses by sex throughout given sex differences in smoking prevalence and mortality risk (Preston & Wang, 2006; Rogers & Powell-Griner, 1991). I recoded age into five-year groups 50–54 through 85+. Adults ages 50+ were selected because lung cancer deaths are rare before age 50, and the PGW method relies on the association between lung cancer and other causes of death to estimate smoking-attributable mortality. Period was based on years 1980–2004 and recoded into five-year periods. Other applications of the PGW method have used individual years, but I combined years to produce stable mortality rates for rare causes of death. I grouped the 50 U.S. states (excluding the District of Columbia and U.S. territories) into nine divisions (U.S. Census Bureau, 2015) to account for geographic clustering of smoking-attributable mortality (Fenelon, 2013; Fenelon & Preston, 2012). Although divisional differences are not a focus of this study, I adjusted for division to approximate the U.S.-based PGW model by Fenelon and Preston (2012). Geographic information is not available in public-use U.S. mortality data after 2004.

I coded the established and additional smoking-attributable causes of death based on ICD-9 and ICD-10 cause-of-death codes (WHO, 2010). ICD-9 was implemented in 1979 and ICD-10 was implemented in 1999. I harmonized the two ICD versions with guidelines by Anderson, Miniño, Hoyert, and Rosenberg (2001). Although data are available for previous ICD versions, going back earlier would require harmonizing cause-of-death codes across periods with greater variability in cause-of-death coding standards (Anderson, 2011; Preston et al., 2010a). I used complete underlying cause-of-death codes, which include four-character codes with a decimal digit to indicate cause subcategories. Appendix Table A1 lists the ICD-9 and ICD-10 codes for the established and additional causes.

3.3. Analytic approach

I applied a cause-specific version of the PGW indirect method, which fits negative binomial models to predict the logged mortality rate of each cause of death other than lung cancer, using lung cancer mortality rate (per 1000), age, period, and U.S. division as independent variables. I modified the original PGW model in two ways to correct for the implausibly high mortality among older women produced by the original PGW model. First, I excluded adults aged 85+ from negative binomial models and then used the average of the coefficients for ages 70–74, 75–79, and 80–84 as the coefficient at ages 85+ to estimate SAFs (Preston et al., 2010b). Second, I added an age-by-time interaction term, which makes PGW results more consistent with published estimates (Rostron, 2010).

I estimated the association between death rates for each cause of death other than lung cancer and death rates for lung cancer with the following negative binomial model:

$$\ln(M_i) = \beta_0 M_i + \beta_a X_a + \beta_t X_t + \beta_{at}(X_a \times X_t) + \beta_{it}(t \times M_i) + \beta_d(X_a \times X_t),$$

(1)

where \(\ln(M_i)\) is the logged mortality rate for cause of death \(c\), \(M_i\) is the lung cancer mortality rate, \(X_a\) is a series of dummy variables for age group, \(X_t\) is a series of dummy variables for period, \(X_d\) is a series of dummy variables for division, and \(t\) is a linear period term. The model includes an age-by-lung cancer death rate interaction term, a time-by-lung cancer death rate interaction term, and an age-by-time interaction term. Beta coefficients accompany each first-order and interaction term. The negative binomial models include a statistical offset of the natural log of exposure to model the logged rates for each cause of death other than lung cancer.

I calculated the fraction of lung cancer deaths attributable to smoking (\(A_L\)) with the following formula:

$$A_L = \frac{M_L - \bar{M}_N}{M_L},$$

(2)

where \(\bar{M}_N\) is the expected lung cancer death rate among lifetime nonsmokers (Table 1), derived from CPS-II data (Thun et al., 1997). I assume that age- and sex-specific lung cancer death rates would match these rates in the absence of smoking.

The association between lung cancer mortality and mortality from cause \(c\) is represented by \(\beta_{lc}\), which equals the sum of the coefficients

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

Note: Lung cancer death rates among lifetime nonsmokers are imported from CPS-II data for years 1982–1988 (Thun et al., 1997).
from the lung cancer first-order term, age-by-lung cancer interaction term, and time-by-lung cancer interaction term in formula (1):

\[ \hat{\beta}_c = \hat{\beta}_c + \hat{\beta}_c + \hat{\beta}_c. \]

I estimated 34 different sets of age-specific \( \hat{\beta}_c \) values, a set for each smoking-attributable cause other than lung cancer for women and men. Using \( \hat{\beta}_c \) values, I estimated the fraction of deaths attributable to smoking from cause \( c \) other than lung cancer (\( A_c \)) with the formula:

\[ A_c = \frac{e^{\hat{\beta}_c(M_c)} - e^{\hat{\beta}_c(M_c)}}{e^{\hat{\beta}_c(M_c)}}, \quad (3) \]

This formula combines the effect of smoking on lung cancer mortality (indicated by the difference in the observed lung cancer mortality rate and the expected lung cancer mortality rate among nonsmokers) with the association between lung cancer mortality and mortality from cause \( c \) (represented by \( \hat{\beta}_c \)) to estimate the fraction of deaths from cause \( c \) due to smoking.

Although the coefficients and SAFs for specific causes of death are informative for determining whether the causes are associated with lung cancer mortality (i.e., evidence that they are attributable to smoking), other population health metrics more effectively convey the impact of smoking on adult mortality (Rogers et al., 2005). Translating the coefficients into the fraction of all deaths due to smoking and the impact of smoking on adult mortality (Rogers et al., 2005). Translating smoking), other population health metrics more effectively convey the additional causes are positive, suggesting that some portion of deaths from these causes are attributable to smoking. Coefficients vary considerably; among women, the largest coefficients occur for COPD, laryngeal cancer, and esophageal cancer (established causes) and additional rare causes combined.

### Table 2

Female coefficients representing the association between cause-specific death rates and lung cancer death rates. Source: 1980–2004 U.S. vital statistics data.

| Cause of death                   | Mean  | 50–54 | 55–59 | 60–64 | 65–69 | 70–74 | 75–79 | 80–84 | 85+ |
|----------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| Lip and oral cavity cancer       | 0.35  | 1.43  | 0.80  | 0.11  | 0.15  | 0.31  | 0.11  | -0.17 | 0.08|
| Esophageal cancer                | 0.53  | 2.74  | 0.67  | 0.54  | -0.02 | 0.24  | 0.05  | -0.03 | 0.09|
| Stomach cancer                   | -0.05 | -0.62 | 0.14  | -0.06 | 0.03  | -0.03 | 0.04  | 0.08 | 0.03|
| Colorectal cancer                | 0.19  | 0.46  | 0.42  | 0.19  | 0.10  | 0.12  | 0.07  | 0.06 | 0.09|
| Liver cancer                     | 0.01  | 0.13  | -0.03 | -0.18 | -0.18 | 0.00  | -0.05 | -0.08|
| Pancreatic cancer                | 0.18  | 0.50  | 0.35  | 0.22  | 0.16  | 0.13  | 0.05  | 0.01 | 0.06|
| Laryngeal cancer                 | 0.80  | 2.71  | 1.08  | 0.86  | 0.49  | 0.51  | 0.25  | 0.19 | 0.32|
| Cervical cancer                  | -0.41 | -0.74 | -0.90 | -0.56 | -0.23 | -0.23 | -0.22 | -0.20 | -0.21|
| Uterine bleeding cancer          | 0.25  | 0.91  | 0.09  | 0.32  | 0.26  | 0.20  | 0.16  | 0.15 | 0.15|
| Kidney and renal pelvis cancer   | 0.11  | 0.23  | 0.23  | 0.00  | 0.14  | 0.07  | 0.15  | -0.03 | 0.07|
| Acute myeloid leukemia            | 0.12  | 0.20  | 0.20  | 0.26  | 0.14  | 0.24  | -0.04 | 0.04 | 0.08|
| Diabetes                         | 0.05  | 0.25  | 0.22  | 0.11  | -0.02 | -0.02 | 0.01  | -0.14 | -0.05|
| Ischemic heart disease           | 0.38  | 1.18  | 0.82  | 0.32  | 0.12  | 0.11  | 0.14  | 0.22 | 0.15|
| Other heart disease              | 0.41  | 1.03  | 0.96  | 0.44  | 0.33  | 0.23  | 0.12  | 0.03 | 0.13|
| Stroke                           | 0.19  | 0.94  | 0.30  | 0.12  | 0.12  | 0.03  | 0.06  | 0.02 | 0.01|
| Atherosclerosis                  | 0.39  | 1.88  | 1.33  | 0.61  | 0.23  | -0.20 | -0.28 | -0.22 | -0.23|
| Aortic aneurysm                  | 0.32  | 0.48  | 0.41  | 0.61  | 0.43  | 0.22  | 0.19  | 0.06 | 0.15|
| Other arterial diseases          | 0.22  | 0.13  | 0.41  | 0.25  | 0.35  | 0.17  | 0.23  | 0.06 | 0.15|
| Pneumonia, inflammation, and tuberculosis | 0.49  | 0.67  | 0.77  | 0.38  | 0.30  | 0.36  | 0.29  | 0.16 | 0.27|
| Chronic obstructive pulmonary   | 1.07  | 2.96  | 1.77  | 0.87  | 0.60  | 0.60  | 0.58  | 0.62 | 0.60|

### 4. Results

#### 4.1. Coefficients for established and additional smoking-attributable causes of death

Tables 2 and 3 show age-specific and mean coefficients for cause-specific death rates regressed on lung cancer death rates among women and men, respectively. Nearly all coefficients for both established and additional causes are positive, suggesting that some portion of deaths from these causes are attributable to smoking. Coefficients vary considerably; among women, the largest coefficients occur for COPD, laryngeal cancer, and esophageal cancer (established causes) and additional causes.
intestinal ischemia and essential hypertension/hypertensive renal disease (additional causes). Among men, the largest coefficients occur for COPD and aortic aneurysm (established causes) and liver cirrhosis (an additional cause). Coefficients often differ by sex but are not consistently higher for either group; coefficients are higher among men for some causes but higher among women for other causes.

Age-specific and mean coefficients are negative for some causes of death and age groups. Although seemingly anomalous, negative coefficients generally occur for causes with multiple risk factors. Among women, these causes include cancer of the stomach, liver, or cervix; liver cirrhosis; and unknown causes. Among men, these causes include liver cancer, additional rare causes combined, and unknown causes. Liver cancer and cirrhosis are associated with heavy alcohol use, stomach and liver cancer are linked to obesity, and additional rare causes combined and unknown causes have numerous risk factors other than smoking. Coefficients tend to decrease with increasing age and, for some causes of death, fall below zero. Negative coefficients may occur among older adults for selected causes because they experience higher death rates overall.

4.2. Cause-specific smoking-attributable fractions

Table 3 presents mean SAFs for each cause of death in 1980–2004. SAFs calculated from negative coefficients are also negative, so I set negative mean SAFs to zero. The largest cause-specific SAFs are for lung cancer (83% among women and 87% among men), affirming that the majority of lung cancer deaths are attributable to smoking in the United States. Among established causes, SAFs are large for COPD among women (61%) and men (43%), laryngeal cancer among women (45%), and aortic aneurysm among men (45%). Among additional causes, SAFs are large for intestinal ischemia among women (44%) and liver cirrhosis among men (47%). SAFs for some additional causes are as large as or larger than SAFs for established causes.

Examining the SAF for all established and additional causes combined (A) showed that 15% of female deaths and 22% of male deaths were due to smoking among U.S. adults at ages 50+ in 1980–2004. In supplemental analyses, I estimated SAFs following the original PGW method (i.e., regressing death rates from all causes of death other than lung cancer combined on death rates from lung cancer). This approach found that 17% (24%) of female (male) deaths were attributable to smoking in 1980–2004. Thus, the cause-specific indirect method produces only slightly lower estimates than the original PGW method.

4.3. Annual number of excess deaths by cause attributable to smoking, 2000–2004

I estimated the annual number of excess deaths due to smoking for each established and additional cause in 2000–2004 by multiplying the cause-specific SAFs ($A_i$ and $A_{L_i}$) by the corresponding number of deaths from each cause ($D_i$ and $D_{L_i}$) in years 2000–2004 divided by five (Table 5). The indirect method finds that 420,284 annual deaths would have been averted in the absence of smoking. The number of excess deaths attributable to smoking is higher among men (253,041) than women (167,243). The most sizable numbers of preventable deaths are observed for the established causes lung cancer and ischemic heart disease, which affect each group for more excess deaths than all additional
causes combined. Although the additional causes account for fewer deaths individually compared with the established causes, when accumulated, the additional causes account for 60,418 deaths (18,006 among women and 42,412 among men), or about 14% of all smoking-attributable deaths.

I compared these indirect estimates to estimates using the direct method and NHIS-LMF data (Lariscy et al., 2018) to determine whether indirect and direct methods produce similar results. Lariscy and colleagues used data from the 1990–2011 NHIS-LMF and controlled for several sociodemographic and behavioral confounders. They estimated the annual number of smoking-attributable deaths at ages 35+ in 2010. I recalculated their annual number of excess deaths for ages 50+ in 2000–2004 in order to match the ages and period of my indirect method estimates. The direct method identified 486,476 annual smoking-attributable deaths at ages 50+ in 2000–2004. Among women and men, the indirect method estimate is lower than the direct method estimate.

5. Discussion

This study makes several important contributions to the study of smoking-attributable mortality by cause of death in the United States. First, it is the first application of the PGW indirect method to examine a comprehensive list of smoking-attributable causes of death. The evidence for many of these causes is currently suggestive but not sufficient to infer a causal association with smoking. Second, it draws on U.S. data that include the entire population and covers a 25-year period. In contrast, prospective cohort studies rely on sample data that exclude certain segments of the U.S. population. For example, the NHIS-LMF is nationally-representative but excludes the institutionalized population whereas the CPS-II is not nationally-representative and is disproportionately non-Hispanic white and college-educated (Lariscy et al., 2018; Malarcher et al., 2000). Finally, it produces new estimates of the annual number of excess deaths overall and by cause of death.
that would have been averted in the absence of smoking, which are lower than estimates from the direct method.

My results contribute to the mounting evidence supporting a causal association between smoking and the additional causes of death. Both the indirect and direct approaches suggest that several additional causes are attributable to smoking. Carter et al. (2015) and Lariscy et al. (2018) used the direct method with prospective cohort data to compare cause-specific mortality risk among smokers relative to nonsmokers and found that relative risks tend to be larger for established causes than additional causes, but still significantly elevated. Likewise, the current study uses the indirect method with vital statistics data to show that the associations with lung cancer mortality rates tend to be larger for established causes than additional causes, but are still positive for nearly all causes of death examined. This general agreement between the indirect and direct approaches confirms that the indirect method is a valid alternative for estimating smoking-attributable mortality among populations without survey data linked to death records but where cause-of-death reporting in vital statistics data is available.

Despite the similarity between results from the direct and indirect approaches, the comparison reveals that the direct method may be better able to detect smoking-attributable causes of death that are also linked to other risk factors than the indirect method. The PGW indirect model produces negative mean coefficients for cancers of the stomach, liver, and cervix among women and liver cancer among men, even though the U.S. Surgeon General recognizes these causes of death as causally linked to smoking. Using the direct method, Lariscy et al. (2018) found mortality risk of current smokers, compared with never smokers, to be 198% higher from liver cancer among men, 128% higher for cervical cancer among women, and 92% higher for stomach cancer among women.

The evidence that breast and prostate cancer (additional causes) are linked to smoking presents an important new contribution. The coefficients are positive for breast and prostate cancer for all age groups, and SAFs derived from those coefficients indicate that 12% (17%) of breast (prostate) cancer deaths are due to smoking. Among U.S. women, lung cancer is the leading cause of cancer death, and breast cancer is the second leading cause of cancer death. Among U.S. men, lung cancer is the leading cause of cancer death, and prostate cancer is the second leading cause of cancer death (American Cancer Society, 2018). Although other social and biological risk factors account for the majority of breast and prostate cancer deaths, my results add to the compelling evidence indicating that smoking is causally related to breast and prostate cancer.

Both the direct and indirect methods find that liver cirrhosis is linked to smoking among men. In sensitivity analyses, Lariscy et al. (2018) limited their sample to the 1997–2011 NHIS-LMF Sample Adult Files to adjust for alcohol use and other socioeconomic and behavioral confounders. Net of alcohol use, mortality risk for liver cirrhosis among men was significantly elevated among current smokers (risk ratio = 3.29) and former smokers (risk ratio = 1.83) relative to nonsmokers. Corresponding cirrhosis mortality risk ratios among women were not statistically significant when models adjusted for alcohol use. In the present study, cirrhosis mortality is positively associated with lung cancer mortality for all age groups among men, but the association is weakly positive or negative among women. This evidence that a portion of cirrhosis deaths are linked to smoking among men has important methodological implications for the Peto-Lopez indirect method. Recall that the Peto-Lopez method assumes that no cirrhosis deaths are due to smoking. Future applications of the Peto-Lopez method should import smoker-nonsmoker risk ratios for cirrhosis mortality among men when assessing the number and fraction of deaths due to smoking, with some reduction to the risk ratio to account for confounding from heavy alcohol use.

5.1. Limitations

Several limitations must be considered. First, the lag period between cigarette smoke exposure and smoking-attributable death may differ for lung cancer and other smoking-attributable causes of death. For instance, some heart diseases develop at younger ages than lung cancer. Different lag periods could affect the age-specific coefficients between lung cancer death rates and rates from other causes. This lag difference could potentially be the reason for negative coefficients among some causes of death. Fenelon and Preston (2012) address this possibility and conclude that differential lags are most problematic for populations at early stages of the tobacco epidemic but not for countries at an advanced stage like the United States. Second, the cause-specific PGW approach relies on accurate and consistent cause-of-death coding on death certificates. Although cause-of-death reporting quality is generally high in the United States (Mathers, Fat, Inoue, Rao, & Lopez, 2005), coding accuracy may vary by time, age, or region. This may particularly be an issue among older decedents with multiple chronic conditions. Finally, my findings are not conclusive evidence of causality between smoking and the additional causes, but they contribute to the mounting evidence that meets the Surgeon General’s criteria to infer a causal association with smoking (U.S. DHHS 2004).

5.2. Conclusion

Despite declines in smoking prevalence and smoking-attributable mortality since the Surgeon General’s first report, smoking continues to kill nearly half a million American adults every year. Especially troubling, declines in youth smoking prevalence have stalled so that smoking endures among U.S. adolescents and young adults. Even with reductions in smoking prevalence, considerable smoking-attributable mortality will continue well into the twenty-first century due to the existing smoking burden among adults now and the decades-long lag between smoke exposure and eventual morbidity and mortality from smoking-attributable diseases (Nam, Rogers, & Hummer, 1996; Pampel, 2005). At the same time, encouraging research suggests that smoking cessation greatly reduces and perhaps even erases the detrimental effects of smoking, especially if smokers quit at younger ages (Doll et al., 2004; Jha et al., 2013). Tobacco control policies implemented to discourage smoking initiation and promote cessation will save lives well into the future.

Acknowledgements

Earlier versions of this article were presented at the 2018 meeting of the Population Association of America, Denver, Colorado, and the 2017 meeting of the Southern Demographic Association, Morgantown, West Virginia. I thank Evelyn Patterson and SSM-PopulationHealth reviewers for helpful comments, Andrew Fenelon for assistance with applying the Preston-Glei-Wilmoth model to U.S. data, and Caleb Cooley and Samia Tasmim for assistance with data analysis. I thank NCHS and NBER for providing mortality data and the U.S. Census Bureau, SEER, and NCI for providing population data. Results of this article do not necessarily express the views of the U.S. Census Bureau, NBER, NCHS, or NCI.

Ethical statement

Declarations of interest

None.
Appendix A

(See Table A1).

Table A1
ICD-9 and ICD-10 codes for smoking-attributable causes.

| Causes of death                                    | ICD-9 codes | ICD-10 codes |
|---------------------------------------------------|-------------|--------------|
| Diseases established as caused by smoking         |             |              |
| Lip and oral cavity cancer                        | 140–149     | C00–C14      |
| Esophageal cancer                                 | 150         | C15          |
| Stomach cancer                                    | 151         | C16          |
| Colorectal cancer                                 | 153, 154    | C18–C20      |
| Liver cancer                                      | 155         | C22          |
| Pancreatic cancer                                 | 157         | C25          |
| Laryngeal cancer                                  | 161         | C32          |
| Lung cancer                                       | 162         | C33–C34      |
| Cervical cancer                                   | 180         | C53          |
| Urinary bladder cancer                            | 188         | C67          |
| Kidney and renal pelvis cancer                    | 189         | C64–C66      |
| Acute myeloid leukemia                            | 205         | C92.0        |
| Diabetes mellitus                                 | 250         | E10–E14      |
| Ischemic heart disease                            | 410–414, 429.2 | I20–I25   |
| Other heart disease                               | 390–398, 415–429.1, 429.3–429.9 | I00–I09, I26–I51 |
| Total stroke                                      | 430–434, 436–438 | I60–I69   |
| Atherosclerosis                                    | 440.1, 440.2 | I70          |
| Aortic aneurysm                                    | 441         | I71          |
| Other arterial diseases                           | 442–448     | I72–I78      |
| Pneumonia, influenza, tuberculosis                | 010–018, 480–487 | A16–A19, J10–J18 |
| Chronic obstructive pulmonary disease             | 490–492, 496 | J40–J44      |
| Additional diseases associated with smoking       |             |              |
| All other infections                              | 001–009, 020–139 | A00–A15, A20–B99 |
| Breast cancer                                     | 174         | C50          |
| Prostate cancer                                   | 185         | C61          |
| Rare cancers                                      | 152, 156, 158–160, 163–171, 175, 179, 181, 182, 186, 187, 190, 192–198, 200, 201, 209–238 | C17, C21, C23, C24, C26, C30, C31, C37–C41, C45–C49, C51, C52, C54, C55, C58, C60, C62, C63, C68–C70.1, C72–C79, C81.0–C81.3, C96, C97, D00–D48.7 |
| Cancers of unknown site                            | 199, 239    | C80, D48.9   |
| Hypertensive heart disease                        | 402         | I11          |
| Essential hypertension/hypertensive renal disease | 401, 405    | I10, I15     |
| All other respiratory diseases                    | 460–478, 500–516.2, 516.4–519 | J00–J09, J20–J39, J45–J84.0, J84.8–J99 |
| Intestinal ischemia                               | 557         | K55          |
| Liver cirrhosis                                   | 571.0–571.3, 571.5–571.9 | K70, K74 |
| All other digestive diseases                      | 520–556, 558–569, 574–579 | K00–K52.9, K56–K67, K80–K93 |
| Renal failure                                     | 584–586     | N17–N19      |
| Additional rare causes combined                   | 360–389, 680–799.8 | H00–H95, L00–L99, M00–M99, O00–R98 |
| Unknown causes                                    | 799.9       | R99          |

Note: ICD-9 and ICD-10 codes are harmonized based on guidelines by Anderson et al. (2001).

References

American Cancer Society (2018). Cancer facts and figures 2018. Atlanta, GA: American Cancer Society. Retrieved from <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018-cancer-facts-and-figures-2018.pdf>.

Anderson, R. N. (2011). Coding and classifying causes of death: Trends and international differences. In R. G. Rogers, & E. M. Crimmins (Eds.). International handbook of adult mortality (pp. 467–489). Dordrecht, The Netherlands: Springer.

Anderson, R. N., Miniño, A. M., Hoyert, D. L., & Rosenberg, H. M. (2001). Comparability of cause of death between ICD-9 and ICD-10: Preliminary estimates. National Vital Statistics Reports, 49(2), 1–32.

Blue, L., & Fenelon, A. (2011). Explaining low mortality among U.S. immigrants relative to native-born Americans: The role of smoking. International Journal of Epidemiology, 40(3), 786–793.

Carter, B. D., Abnet, C. C., Feskanich, D., Freedman, N. D., Hartge, P., Lewis, C. E., Ockene, J. K., Prentice, R. L., Speizer, F. E., Thun, M. J., & Jacobs, E. J. (2015). Smoking and mortality—Beyond established causes. New England Journal of Medicine, 372(7), 631–640.

Crimmins, E. M., Preston, S. H., & Cohren, B. (Eds.). (2011). Explaining divergent levels of longevity in high-income countries. Washington, DC: National Academies Press.

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), 1519–1527.

Ezzati, M., & Lopez, A. D. (2003). Estimates of global mortality attributable to smoking in 2000. Lancet, 362(9387), 847–852.

Fenelon, A. (2013). Geographic divergence in mortality in the United States. Population and Development Review, 39(4), 611–634.

Fenelon, A., & Preston, S. H. (2012). Estimating smoking-attributable mortality in the United States. Demography, 49(3), 797–818.

GBD 2015 Tobacco Collaborators (2017). Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: A systematic analysis from the Global Burden of Disease Study 2015. Lancet, 389(10082), 1865–1906.

Ho, J. Y., & Elo, I. T. (2013). The contribution of smoking to black-white differences in U.S. mortality. Demography, 50(2), 545–568.

Lyon, France: International Agency for Research on Cancer. Retrieved from <http://monographs.iarc.fr/ENG/Monographs/vol83/mono83.pdf>.

Jha, P., Ramasundarabhatte, C., Landsman, V., Rostron, B., Thun, M. J., Anderson, R. N.,...
