# A new method for estimating smoking-attributable mortality in high-income countries

**Samuel H Preston,[1] Dana A Glei[2]\* and John R Wilmoth[2]**

[1]Population Studies Center, University of Pennsylvania, Philadelphia, PA, USA and [2]Department of Demography, University of California, Berkeley, CA, USA

\*Corresponding author. 5985 San Aleso Ct., Santa Rosa, CA 95409-3912, USA. E-mail: danaglei@sonic.net

**Background**  Cigarette smoking is responsible for a massive loss of life in both developed and developing countries. This article develops an alternative to the Peto–Lopez method for estimating the number or fraction of smoking-attributable deaths in high-income countries.

**Methods**  We use lung cancer death rates as an indicator of the damage caused by smoking. Using administrative data for the population aged ⩾50 years from 20 high-income countries in the period from 1950 to 2006, we estimate a negative binomial regression model that predicts mortality from causes other than lung cancer as a function of lung cancer mortality and other variables. Using this regression model, we estimate smoking-attributable deaths based on the difference between observed death rates from lung cancer and expected rates among non-smokers.

**Results**  Combining the estimated number of excess deaths from lung cancer with those from other causes, we find that among males in 1955 the smoking-attributable fraction was highest in Finland (18%); among women, no country exceeded 1%. By 2003, Hungary had the highest fraction of smoking-attributable deaths among males (32%), whereas the USA held that position among women (24%). Our estimates are remarkably similar to those produced by the Peto–Lopez method, a result that supports the validity of each approach.

**Conclusions**  We provide a simple and straightforward method for estimating the proportion of deaths attributable to smoking in high-income countries. Our results demonstrate that smoking has played a central role in levels, trends and international differences in mortality over the past half century.

**Keywords**  Smoking, mortality, smoking/mortality, high-income populations, lung neoplasms, cause of death

## Introduction

Although the negative effects of smoking on mortality at the individual level are well established, measuring the mortality impact at the population level is more challenging because of the difficulty of obtaining accurate cohort histories of smoking behaviour. Cohort data are necessary because smoking generally begins relatively early in life, whereas the full impact

on mortality is not revealed until much older ages. The most persuasive evidence identifying the mortality risks associated with smoking has been drawn from prospective cohort studies that compare the death rates of current smokers and former smokers with the death rates of those who never smoked regularly. The Cancer Prevention Study II (CPS-II) in the USA is the largest such study, but is based on a sample of volunteers who are more likely to be White, middle class and college-educated than the US population as a whole.[1] Beyond the problem of generalizability, cohort studies are also subject to attenuation bias resulting from imprecise classification of smoking status. For example, occasional and short-term smokers are included among 'lifetime non-smokers' in CPS-II.[2] Moreover, smoking status is typically identified at baseline and assumed constant thereafter; Taylor et al.[3] showed that correction for this bias substantially raised the estimated risk of smoking.

Given the absence of cohort studies in many populations for which attributable risk estimates are sought, Peto and Lopez and colleagues developed an ingenious method for filling this gap.[4] The method 'borrows' the relative risks of cause-specific mortality for smokers versus non-smokers from CPS-II and applies them to the population of interest. Instead of applying them to the distribution of the population by smoking status, however, they used observed lung cancer death rates as an indicator of the population's cumulative smoking exposure. Specifically, the level of excess lung cancer mortality due to smoking—which is determined by comparing the observed lung cancer rates to the rates among non-smokers and smokers in CPS-II within the corresponding sex and age group—is presumed to reflect the maturity of the smoking epidemic in a population and is used in combination with the relative risks observed in CPS-II to estimate the fraction of cause-specific deaths that is attributable to smoking.[5] This 'indirect' indicator may in fact be a more reliable index of the damage from smoking than directly measured smoking behaviour based on self-report.

Their method allows one to estimate smoking-attributable deaths in a given population without requiring any information on smoking behaviour. However, the Peto–Lopez approach is heavily dependent upon the assumption that CPS-II estimates of lung cancer death rates for smokers and non-smokers and relative risks for other causes of death can be applied to other countries and across time.[6] Furthermore, because smokers are self-selected, some of the mortality differential between smokers and non-smokers may be attributable to confounding with other risk factors. Thus, to avoid overstating the impact of smoking, Peto et al. rather arbitrarily halved the CPS-II excess risks for causes other than lung cancer—If the relative risk is represented by RR, the excess risk equals RR–1; thus, halving the excess

risk yields a relative risk of $1 + \frac{1}{2}(RR - 1)$. Others have lowered the reduction to 30%[7] or adjusted directly for confounding factors.[8] The Peto–Lopez method has been widely used although not widely validated. A comparison of the attributable fraction based on the Peto–Lopez method with estimates based on 13 prospective cohort studies in five northern European countries found a mean absolute difference of 6.2%, or 30.0% of the mean attributable fraction;[9] this relatively large discrepancy may result from the fact that the cohort studies' indicator of smoking was smoking status as reported in survey data, whereas the Peto–Lopez method uses observed mortality from lung cancer, which is presumably a more comprehensive and more reliable measure of cumulative smoking exposure.

This article develops an alternative to the Peto–Lopez method for calculating deaths attributable to smoking. Although we also use lung cancer mortality as the indicator of smoking damage, we do not rely on the relative risks from CPS-II or any other study. Instead, we investigate the macro-level statistical association between lung cancer mortality and mortality from all other causes of death among 20 countries during 1950–2006. This approach assumes that lung cancer mortality is a reliable indicator of the damage from smoking and that such damage has left an identifiable imprint on other causes of death at the population level. A related approach has been applied to sub-national time-series data for various cancers.[2,10]

## Methods

### Modelling strategy

Our model for estimating the impact of smoking on mortality is formally developed in Supplementary data available at *IJE* online. This model is based on the assumption that lung cancer mortality is a good proxy for the impact of smoking on mortality from other causes. Specifically, we assume that, after adjusting for sex and age, smoking is the only source of variation in lung cancer death rates in the populations under consideration. The Peto–Lopez model uses this same assumption, which is justified by evidence suggesting that changes in lung cancer rates result primarily from the history of smoking behaviour.[11–13]

Using negative binomial regression,[14] we model mortality at ages 50–54, 55–59, ... 80–84, ≥85 years from causes other than lung cancer [$M_O$] as a function of lung cancer mortality [$M_L$] and other variables. Preliminary analyses indicated that variation in $M_O$ was greater than would be present in a Poisson process, thus justifying the choice of a negative binomial model.[15] A log-linear relationship is assumed between mortality and its predictors (thus, a unit increase in $M_L$ is associated with a constant proportional increase

**Table 1** Coefficients for lung cancer death rates in 2003 and assumed values of lung cancer death rates among non-smokers

| Age group (years) | Model coefficients for lung cancer death rate (per 1000) in 2003[a] | | Age group (years) | Assumed lung cancer death rates (per 1000) among non-smokers[b] | |
|---|---|---|---|---|---|
| | Males | Females | | Males | Females |
| 50–54 | 0.342 | 0.699 | 50–54 | 0.06 | 0.06 |
| 55–59 | 0.183 | 0.455 | 55–59 | 0.05 | 0.07 |
| 60–64 | 0.113 | 0.281 | 60–64 | 0.12 | 0.12 |
| 65–69 | 0.075 | 0.147 | 65–69 | 0.22 | 0.17 |
| 70–74 | 0.053 | 0.072 | 70–74 | 0.35 | 0.31 |
| 75–79 | 0.042 | 0.036 | 75–79 | 0.52 | 0.33 |
| $\geqslant$80 | 0.050 | 0.147 | 80–84 | 0.89 | 0.58 |
| | | | $\geqslant$85 | 0.87 | 0.61 |

[a]Based on a negative binomial regression model predicting mortality from causes other than lung cancer. The coefficients shown here correspond to values of $\beta'_L$ as defined in the description of equation (2). Thus, a 0.001 change in the lung cancer death rate implies that the death rate for other causes combined is higher by a factor of $e^{\beta'_L}$ for the specified age–sex group in 2003, taking into account interactions with both age and calendar year. Each sex-specific model also includes dummy variables for country, calendar year and age group as well as interactions between country and year (treated as linear).
[b]Based on observed lung cancer rates among persons in the 1982–88 CPS-II who never smoked regularly.[1]

in $M_O$). The outcome variable is the number of deaths from causes other than lung cancer for a given country–year–age group divided by the number of person-years of exposure. All analyses were conducted using Stata 10.1.

Because of potential sex differences in the effects of smoking and in age-patterns of mortality, we model the sexes separately. Control variables include country fixed effects; dummies for each calendar year and each age group and interactions between country and year (treated as linear) to allow for inter-country differences in the pace of mortality decline. We include an interaction between $M_L$ and year of observation (linear) because previous studies suggest that the relative risk associated with smoking has increased over time.[1,16] Finally, we interact the smoking indicator with a set of age dummies (50–54, 55–59, ... $\geqslant$80) to allow the association between $M_L$ and $M_O$ to vary across age.[1]

To assess the validity of the model, we estimated a parallel set of models, substituting in turn the death rate from each of four groups of causes as the dependent variable. We expect that the damage from smoking signalled by high lung cancer mortality will also be clearly visible in two of these categories—smoking-related cancers and respiratory diseases. Thus, we expect that mortality from these causes of death will be positively correlated with lung cancer mortality across time and space. We also examine a category in which mortality should be unrelated to lung cancer mortality: external causes (i.e. accidents, homicide and suicide). Finally, we examine mortality from 'other cancers' (i.e. neither lung cancer nor smoking-related cancers). A weak relationship with lung cancer mortality is anticipated for this category based on relative risks observed in CPS-I.[17]

We also investigated the robustness of the results to two alternative specifications: the use of a second-degree polynomial rather than a set of dummy variables to represent the interaction between age and lung cancer mortality; and deletion of the variable representing trends in the relation between lung cancer mortality and mortality from other causes. In addition, we examined the sensitivity of the results to the exclusion of Hungary and Japan when fitting the model. Hungary is the only Eastern European country in our dataset and exhibits excess mortality in middle-adulthood similar to that observed in post-Soviet countries. Japan is the sole Asian country in our dataset and has a very low level of mortality combined with a rapid increase in smoking prevalence.

## Estimating the attributable fraction

To estimate the fraction of deaths attributable to smoking, we assume that, in the absence of smoking, lung cancer death rates (by sex and 5-year age group) would match those observed among Americans in the CPS-II study (1982–88) who never smoked regularly.[1] These rates are presented in Table 1 and are generally similar to those observed in other samples of non-smokers.[18,19] No trend in lung cancer mortality among non-smokers in the USA was observed over a 20-year period.[20] In some populations where the prevalence of smoking is thought to have been very low (e.g. Spanish women in 1951–54), lung cancer rates were even lower than among non-smokers in CPS-II.

For each country–year–sex–age group, we calculate the fraction of lung cancer deaths attributable to smoking as:

$$A_L = \frac{M_L - \lambda_L^N}{M_L}, \tag{1}$$

where $M_L$ is the observed lung cancer death rate and $\lambda_L^N$ is the expected rate among non-smokers. In cases where $M_L - \lambda_L^N$ is negative, the value of $A_L$ is set at 0. For mortality from other causes, we compare the number of deaths predicted by the negative binomial regression model under two assumptions about the lung cancer death rate: that it equals the observed level for the population or that it equals the level assumed for non-smokers in the corresponding sex–age group. The difference between these two predicted numbers of deaths, divided by the prediction based on the observed level of lung cancer mortality, provides an estimate of the fraction attributable to smoking. This procedure is equivalent to implementing the following formula:

$$A_O = 1 - e^{-\beta'_L(M_L - \lambda_L^N)}, \qquad (2)$$

where $\beta'_L$ is the coefficient of lung cancer mortality in the regression model of $M_O$, including any interactions between $M_L$ and time (since 1950) or age. If both $\beta'_L$ and $M_L - \lambda_L^N$ are positive (as they are in the large majority of cases), then $A_O$ lies between 0 and 1. If either $\beta'_L$ or $(M_L - \lambda_L^N)$ is negative, we set the value to zero before computing $A_O$ (this occurs rarely, mostly in situations where $M_L$ is very low, suggesting that smoking-related mortality is negligible).

Finally, the overall attributable fraction for deaths from all causes is a weighted average:

$$A = \frac{A_L D_L + A_O D_O}{D}, \qquad (3)$$

where $D_L$, $D_O$ and $D$ represent observed numbers of deaths from lung cancer, other causes and all causes combined, respectively.

### Data

Death counts by cause of death are drawn from the WHO Mortality Database.[21] All-cause death counts, exposure estimates and death rates come from the Human Mortality Database (HMD).[22] We use annual data by sex and 5-year age groups (50–54, . . . 80–84, ⩾85 years) for 20 high-income countries since 1950 (see Table S1; Supplementary data are available at *IJE* online). The selected countries include all those for which data were available from both the HMD and WHO for the period since 1955. We use data from the HMD because it is the most easily accessible source of reliable all-cause mortality estimates for a wide range of countries with high-quality data. Because the HMD does not include data by cause of death, we obtain the cause-specific distribution of deaths from the WHO Mortality Database, which is the most comprehensive, publicly accessible collection of such data. The dataset used for this analysis contains 280.6 million deaths and 9.8 billion person-years of exposure. For each country–year–sex–age group, we apply the distribution of deaths by cause from WHO to the

death counts and rates from the HMD to derive cause-specific death counts and rates.

Lung cancer deaths are defined to include malignant neoplasms of trachea, bronchus and lung (ICD-7: 162, 163; ICD-8 and ICD-9: 162; ICD-10: C33, C34). The quality of lung cancer mortality data is generally good and may be less affected by variation in diagnostic practices than data on other causes such as vascular diseases.[11] Some of the effects of changes in coding practices may be reflected statistically in our period effects, but they are not expected to be large.

## Results

Table 1 presents the estimated age and sex-specific regression coefficients from the negative binomial model depicting the relationship between lung cancer death rates and mortality from other causes for 2003 (see also Supplementary Table S2; Supplementary data are available at *IJE* online). Each coefficient indicates the proportionate effect of a 0.001 change in the lung cancer death rate on mortality from other causes of death. The interactive variable between lung cancer mortality and time has a positive (though small) coefficient, suggesting that the relationship between lung cancer mortality and mortality from other causes of death has strengthened over time. This coefficient indicates a linear time trend (in a logarithmic scale) of 0.00038 [95% confidence interval (CI) 0.00026–0.00050, $P \sim 0.000$] for men and 0.00277 (95% CI = 0.00218–0.00337, $P \sim 0.000$) for women. Thus, *ceteris paribus*, the predicted value of $M_O$ corresponding to a particular value of $M_L$ is estimated to have increased by 0.8% for males and 5.7% for females over a 20-year period. Several studies have suggested that the relative risk from smoking has increased over time for all major cause of death categories.[20,23] In addition, the increase in the coefficient of lung cancer may reflect a changing mix of causes of death.

The age-specific coefficients shown in Table 1 indicate that, as age advances, a given increment in lung cancer mortality is associated with a smaller proportionate impact on mortality from other causes. This declining pattern is reversed in the oldest age category for both sexes. This reversal may reflect under-recording of lung cancer among the oldest decedents, perhaps because of the multiplicity of morbid conditions typically present among them. For example, if only 90% of lung cancer deaths were recorded accurately at the highest ages compared with 100% at younger ages, then the multiplier would increase by ~11% in the highest age category.

### Attributable risk estimates

Estimates shown in Table 2 indicate that the attributable risk from smoking is much greater for men than for women. However, the risk for women,

**Table 2** Estimated smoking-attributable fraction (A) and 95% CI[a] among deaths at ages ≥50 years in 1955, 1980, 2003, by sex and country

| | Males | | | | | | Females | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1955 | | 1980 | | 2003 | | 1955 | | 1980 | | 2003 | |
| Country | A | 95% CI | A | 95% CI | A | 95% CI | A | 95% CI | A | 95% CI | A | 95% CI |
| Australia | 0.07 | (0.063, 0.076) | 0.23 | (0.216, 0.236) | 0.18 | (0.172, 0.189) | 0.00 | (0.000, 0.001) | 0.03 | (0.027, 0.037) | 0.13 | (0.114, 0.142) |
| Austria | 0.15 | (0.141, 0.162) | 0.22 | (0.210, 0.234) | 0.18 | (0.174, 0.195) | 0.01 | (0.004, 0.009) | 0.02 | (0.011, 0.024) | 0.06 | (0.048, 0.076) |
| Belgium | 0.09 | (0.081, 0.094) | 0.32 | (0.305, 0.330) | 0.34[b] | (0.329, 0.354) | 0.00 | (0.001, 0.004) | 0.01 | (0.005, 0.015) | 0.05[b] | (0.039, 0.060) |
| Canada | 0.07 | (0.061, 0.072) | 0.23 | (0.226, 0.244) | 0.26 | (0.248, 0.266) | 0.00 | (0.001, 0.005) | 0.05 | (0.043, 0.056) | 0.23 | (0.209, 0.243) |
| Denmark | 0.07 | (0.062, 0.078) | 0.23 | (0.213, 0.239) | 0.24[b] | (0.222, 0.249) | 0.00 | (0.002, 0.008) | 0.05 | (0.040, 0.061) | 0.19[b] | (0.172, 0.216) |
| Finland | 0.18 | (0.168, 0.200) | 0.29 | (0.274, 0.304) | 0.19 | (0.172, 0.198) | 0.01 | (0.003, 0.010) | 0.02 | (0.009, 0.025) | 0.05 | (0.036, 0.070) |
| France | 0.05 | (0.047, 0.053) | 0.18 | (0.169, 0.182) | 0.20 | (0.194, 0.207) | 0.00 | (0.001, 0.002) | 0.00 | (0.002, 0.003) | 0.02 | (0.018, 0.022) |
| Hungary | 0.07 | (0.066, 0.078) | 0.23 | (0.218, 0.239) | 0.32 | (0.313, 0.337) | 0.01 | (0.007, 0.013) | 0.04 | (0.032, 0.047) | 0.15 | (0.131, 0.162) |
| Iceland | 0.03 | (0.000, 0.062) | 0.06 | (0.025, 0.098) | 0.17 | (0.113, 0.235) | 0.00 | (0.000, 0.001) | 0.08 | (0.023, 0.142) | 0.21 | (0.101, 0.314) |
| Italy | 0.04 | (0.039, 0.044) | 0.21 | (0.205, 0.219) | 0.25 | (0.244, 0.260) | 0.00 | (0.001, 0.003) | 0.01 | (0.009, 0.011) | 0.06 | (0.051, 0.063) |
| Japan | 0.01 | (0.011, 0.014) | 0.12 | (0.113, 0.122) | 0.22 | (0.211, 0.225) | 0.00 | (0.000, 0.001) | 0.02 | (0.021, 0.027) | 0.13 | (0.121, 0.140) |
| The Netherlands | 0.10 | (0.092, 0.107) | 0.34 | (0.327, 0.352) | 0.28 | (0.270, 0.293) | 0.00 | (0.000, 0.002) | 0.01 | (0.005, 0.014) | 0.10 | (0.086, 0.110) |
| New Zealand | 0.08 | (0.071, 0.097) | 0.22 | (0.203, 0.238) | 0.18 | (0.166, 0.199) | 0.00 | (0.000, 0.002) | 0.05 | (0.033, 0.060) | 0.14 | (0.115, 0.167) |
| Norway | 0.02 | (0.015, 0.025) | 0.09 | (0.083, 0.102) | 0.17 | (0.157, 0.185) | 0.00 | (0.000, 0.003) | 0.01 | (0.006, 0.012) | 0.08 | (0.059, 0.095) |
| Portugal | 0.02 | (0.014, 0.022) | 0.07 | (0.064, 0.076) | 0.13 | (0.121, 0.137) | 0.00 | (0.000, 0.001) | 0.00 | (0.000, 0.003) | 0.01 | (0.004, 0.008) |
| Spain | 0.04 | (0.034, 0.040) | 0.14 | (0.138, 0.149) | 0.23 | (0.226, 0.241) | 0.00 | (0.000, 0.001) | 0.00 | (0.001, 0.003) | 0.00 | (0.003, 0.005) |
| Sweden | 0.03 | (0.026, 0.034) | 0.11 | (0.102, 0.117) | 0.10 | (0.091, 0.107) | 0.00 | (0.000, 0.003) | 0.02 | (0.010, 0.023) | 0.07 | (0.061, 0.088) |
| Switzerland | 0.09 | (0.086, 0.104) | 0.20 | (0.190, 0.213) | 0.17 | (0.158, 0.180) | 0.01 | (0.000, 0.003) | 0.01 | (0.005, 0.010) | 0.05 | (0.034, 0.062) |
| UK | 0.17 | (0.157, 0.175) | 0.31 | (0.302, 0.322) | 0.22 | (0.216, 0.232) | 0.07 | (0.009, 0.012) | 0.07 | (0.064, 0.079) | 0.19 | (0.178, 0.208) |
| USA | 0.09 | (0.081, 0.090) | 0.24 | (0.230, 0.244) | 0.24 | (0.232, 0.246) | 0.00 | (0.003, 0.004) | 0.07 | (0.064, 0.075) | 0.24 | (0.220, 0.252) |

[a]See Supplementary data available at *IJE* online for further details regarding the construction of the confidence intervals.
[b]Estimates based on data from 1997 for Belgium and 2001 for Denmark.

which was negligible in 1955, has been growing rapidly in most countries. Spain, Portugal and France are exceptions, where the imprint of smoking remains small for women. For men, trends in the attributable fraction are more mixed: the risk declined between 1980 and 2003 in 8 countries, rose in 11 and was constant in the USA. In every country except Iceland, the attributable fraction for 2003 is greater for males than for females. In 2003, the largest estimated proportion of deaths >50 years of age that is attributable to smoking occurred in Hungary among men (0.32) and in the USA among women (0.24).

### Validity

A parallel set of models that use the death rate from a specific group of causes rather than $M_O$ as the dependent variable provides a check on the validity of our model (see Supplementary data available at *IJE* online for details). Here we summarize the results in the form of the mean of age-specific coefficients relating lung cancer mortality to a particular cause-of-death category. For respiratory diseases, the mean of these coefficients is 0.903 for females and 0.394 for males for 2003. Comparable figures for smoking-related cancers are 0.455 and 0.149, for other cancers 0.254 and 0.056 and for external causes, 0.016 and –0.023. Thus, we observe the expected relationships for both males and females: lung cancer mortality is powerfully related to mortality from respiratory diseases across populations, strongly related to smoking-related cancers, positively but more weakly related to other cancers and unrelated (or even slightly negatively related) to mortality from external causes. If the criterion were the mean smoking-attributable fraction of deaths from each cause rather than the mean coefficient (Table S7; Supplementary data are available at *IJE* online), we would observe the same ordering of causes among females but smoking-related cancers and other cancers would reverse position among males.

### Robustness

Across the four alternative specifications, we evaluate robustness based on the change in the attributable fraction; in particular, the mean across countries of the change (in absolute value) of the attributable fraction in 2000 (see Supplementary data available at *IJE* online for details). Results are not sensitive to the alternative specification for the age $\times M_L$ interaction. Across countries, the mean change in the attributable fraction is <0.001 for males and 0.012 for females. When the time trend in $M_L$ effect is dropped, the mean change in the attributable fraction is 0.009 for men and 0.010 for women. When the model is fit excluding data for Hungary, the mean change in the attributable fraction is 0.004 for men and 0.007 for women. When data for Japan are excluded, the comparable figures are 0.004 for men and 0.002 for women. Since women's attributable fractions are generally lower and women's absolute changes tend to be larger, their results are more sensitive than those of men to most of these changes in specification. This greater sensitivity to changes in specification is likely to reflect a lighter footprint of smoking on mortality levels for women than for men.

## Discussion

We have described a new method for estimating the fraction of deaths at ages $\geqslant 50$ years that is attributable to smoking. Although this new approach, like that of Peto and Lopez, uses lung cancer mortality as an indirect measure of smoking history, the two approaches diverge radically at that point. The Peto–Lopez approach assumes that relative risks from CPS-II can be exported (sometimes with a mechanical adjustment) to other countries. We make no such assumption but rather, within the confines of a macro-level statistical model, allow the data to determine the relationship between mortality from lung cancer and the aggregate of other causes of death.

Table 3 presents a comparison of the smoking-attributable fraction estimated by our model with the Peto–Lopez estimates for 2000, the latest year for which the Peto–Lopez method has been widely applied to data from developed countries.[24] In this application, a 50% reduction is assumed in smokers' excess risk of death in CPS-II from causes other than lung cancer. The Peto–Lopez results pertain to ages $\geqslant 35$ years, whereas ours apply to ages $\geqslant 50$ years. Because deaths between ages 35 and 50 years are few relative to deaths at ages $\geqslant 50$ years, the difference in age spans should have little effect on the comparison.

The two methods produce very similar results for both males and females in terms of the level of attributable risk as well as its international distribution. The correlation between the attributable fractions for the two methods is 0.96 for males and 0.94 for females. In view of the substantial differences in methodology, such similarity of results provides important validation for both approaches.

The largest disparity between the two sets of estimates occurs in Japan for both males and females. In this case, the present method gives estimates of the attributable fraction that are higher than those derived from the Peto–Lopez method. It is possible that the Peto–Lopez estimates are more valid because they reflect variation in the structure of 'other causes of death' in Japan, particularly its unusually low death rate from coronary heart disease. Katanoda *et al.*[25] pool data from three Japanese prospective studies to estimate the smoking-attributable fraction. Their results are more consistent with the present study for males (28 vs 22% for the present study and 18% for Peto–Lopez) and less consistent for females (7 vs 12 and 6%, respectively). However, the age range in their study is somewhat younger than

**Table 3** Comparison of smoking-attributable fraction in 2000, by sex and country

| Country | Males | | Females | |
|---|---|---|---|---|
| | Based on model[a] (ages ⩾50 years) | Peto–Lopez[b] (ages ⩾35 years) | Based on model[a] (ages ⩾50 years) | Peto–Lopez[b] (ages ⩾35 years) |
| Australia | 0.20 | 0.20 | 0.12 | 0.11 |
| Austria | 0.19 | 0.19 | 0.07 | 0.06 |
| Belgium | 0.34[c] | 0.31 | 0.05[c] | 0.05 |
| Canada | 0.25 | 0.25 | 0.20 | 0.18 |
| Denmark | 0.23 | 0.25 | 0.18 | 0.20 |
| Finland | 0.19 | 0.18 | 0.04 | 0.04 |
| France | 0.21 | 0.21 | 0.01 | 0.02 |
| Hungary | 0.32 | 0.31 | 0.12 | 0.12 |
| Iceland | 0.14 | – | 0.21 | – |
| Italy | 0.25 | 0.25 | 0.05 | 0.05 |
| Japan | 0.22 | 0.18 | 0.12 | 0.06 |
| The Netherlands | 0.30 | 0.28 | 0.07 | 0.10 |
| New Zealand | 0.18 | 0.20 | 0.15 | 0.15 |
| Norway | 0.15 | 0.17 | 0.06 | 0.10 |
| Portugal | 0.12 | 0.15 | 0.01 | 0.01 |
| Spain | 0.22 | 0.25 | 0.00 | 0.00 |
| Sweden | 0.10 | 0.10 | 0.05 | 0.07 |
| Switzerland | 0.18 | 0.19 | 0.05 | 0.06 |
| UK | 0.23 | 0.23 | 0.16 | 0.16 |
| USA | 0.25 | 0.24 | 0.22 | 0.20 |

'–': Data are not available.
[a]Estimates based on the model represent the fraction of all deaths at ages ⩾50 years.
[b]Estimates based on Peto *et al.*[24] (http://www.ctsu.ox.ac.uk/~tobacco/SMK_P5_6.pdf) represent the fraction of all deaths at ages ⩾35 years.
[c]Estimates are based on data for 1997, the latest year available in the WHO database.

that considered here, the experience is earlier in time and the data are not nationally representative. Two factors bias downwards the attributable risk estimate in their study as in most prospective studies: smoking status is measured only at baseline and not updated; and the mortality effects of second-hand smoke are not included, whereas they are reflected in lung cancer death rates that form the basis of the other two methods.

Differences for Japan notwithstanding, the correlation between results of the two methods is very high. In other words, the statistical association that we estimate between lung cancer mortality and mortality from other causes as observed over a half century in 20 countries is highly consistent with the procedures employed by Peto and Lopez. Neither method would be expected to work well in populations such as China where factors other than smoking have a major impact on lung cancer mortality.[7] In these situations, modification of the observed lung cancer death rate to account for these other factors would be required before either method could be suitably implemented. The Peto–Lopez method may work better than the present method in populations, such

as many in tropical Africa, where the cause-of-death structure is very different than the pattern typically observed in economically developed countries. On the other hand, the present method is less sensitive to errors in the recording of causes of death than Peto–Lopez, since it combines all causes other than lung cancer into one large group. The present method has the additional advantage of providing estimates of the uncertainty due to random variation in observed death rates and estimated regression coefficients, as shown here in Table 2 and described in the Supplementary data available at *IJE* online.

It will often be possible to use both methods to assess the fraction of deaths attributable to smoking in a particular setting. The difference between estimates will provide another indication of the degree of uncertainty in results. From the present article, it appears that such a comparison will be reassuring in the large majority of cases.

## Supplementary Data

Supplementary data are available at *IJE* online.

# Funding

---

### KEY MESSAGES

- We develop an alternative to the Peto–Lopez method for estimating the number or fraction of smoking-attributable deaths in high-income countries using lung cancer mortality as an indirect measure of smoking exposure.

- In 1955, among the 20 countries included in our analysis, the estimated smoking-attributable fraction among males was highest in Finland (18%); among women, no country exceeded 1%.

- By 2003, among the same 20 countries, Hungary had the highest smoking-attributable fraction among males (32%), whereas the USA held that position among women (24%).

- Our estimates are remarkably similar to those produced by the Peto–Lopez method, a result that supports the validity of both approaches.

- These results demonstrate that smoking has played a central role in levels, trends and international differences in mortality over the past half century.

---

# References

[1] Thun MJ, Day-Lally C, Myers DG, *et al*. Trends in tobacco smoking and mortality from cigarette use in Cancer Prevention Studies I (1959 through 1965) and II (1982 through 1988). In: Burns DM, Garfinkel L, Samet JM (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. Bethesda, MD: Cancer Control and Population Sciences, National Cancer Institute, U.S. National Institutes of Health, 1997, pp. 305–82.

[2] Leistikow BN, Kabir Z, Connolly GN, Clancy L, Alpert HR. Male tobacco smoke load and non-lung cancer mortality associations in Massachusetts. *BMC Cancer* 2008;**8:**341.

[3] Taylor DH Jr, Hasselblad V, Henley SJ, Thun MJ, Sloan FA. Benefits of smoking cessation for longevity. *Am J Public Health* 2002;**92:**990–96.

[4] Peto R, Lopez AD, Boreham J, Thun M, Heath C Jr. Mortality from tobacco in developed countries: indirect estimation from national vital statistics. *Lancet* 1992;**339:**1268–78.

[5] Lopez AD. On the assessment of smoking-attributable mortality. In: *Proceedings of the International Population Conference*. Vol. 1. Liege, International Union for the Scientific Study of Population, 1993, 449–58.

[6] Sterling TD, Rosenbaum WL, Weinkam JJ. Risk attribution and tobacco-related deaths. *Am J Epidemiol* 1993;**138:**128–39.

[7] Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet* 2003;**362:**847–52.

[8] Ezzati M, Henley SJ, Thun MJ, Lopez AD. Role of smoking in global and regional cardiovascular mortality. *Circulation* 2005;**112:**489–97.

[9] Valkonen T, van Poppel F. The contribution of smoking to sex differences in life expectancy: four Nordic countries and the Netherlands 1970-1989. *Eur J Pub Health* 1997;**7:**302–10.

[10] Leistikow BN, Tsodikov A. Cancer death epidemics in United States Black males: evaluating courses, causation, and cures. *Prev Med* 2005;**41:**380–85.

[11] Lopez AD. The lung cancer epidemic in developed countries. In: Lopez AD, Caselli G, Valkonen T (eds). *Adult Mortality in Developed Countries: From Description to Explanation*. Oxford: Clarendon Press, 1995, pp. 111–34.

[12] Preston SH, Wang H. Sex mortality differences in the United States: the role of cohort smoking patterns. *Demography* 2006;**43:**631–46.

[13] Brennan P, Bray I. Recent trends and future directions for lung cancer mortality in Europe. *Br J Cancer* 2002;**87:**43–48.

[14] Long JS, Freese J. Regression models for categorical dependent variables using Stata. College Station, TX: Stata Press, 2006.

[15] McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd edn. Boca Raton: CRC Press, 1989.

[16] Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* 2004;**328:**1519.

[17] Burns DM, Shanks T, Choi W, Thun M, Health C, Garfinkel L. The American Cancer Society Cancer Prevention Study I: 12-year followup on 1 million men and women. *In: Changes in Cigarette-related Disease Risks and their Implications for Prevention and Control, Smoking and Tobacco Control*. NIH publication no. 97-4213, Smoking and tobacco control monographs. Volume 8. Bethesda, MD: Cancer Control and Population Sciences, National Cancer Institute, U.S. National Institutes of Health, 1997, 113–304.

[18] Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994;**309:**901–11.

[19] Enstrom JE. Rising lung cancer mortality among nonsmokers. *J Natl Cancer Inst* 1979;**62:**755–60.

[20] Rosenbaum WL, Sterling TD, Weinkam JJ. Use of multiple surveys to estimate mortality among never, current,

and former smokers: changes over a 20-year interval. *Am J Public Health* 1998;**88:**1664–68.

[21] World Health Organization. *Detailed Data Files of the WHO Mortality Database.* www.who.int/whosis/mort/download/en/index.htm (17 November 2008, date last accessed).

[22] Human Mortality Database (HMD). *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).* www.mortality.org (30 December 2008, date last accessed).

[23] Thun MJ, Day-Lally CA, Calle EE, Flanders WD, Heath CW, Jr. Excess mortality among cigarette smokers: changes in a 20-year interval. *Am J Public Health* 1995;**85:** 1223–30.

[24] Peto R, Lopez AD, Boreham J, Thun M. *Mortality From Smoking in Developed Countries 1950–2000*, 2nd edn, updated June 2006. http://www.ctsu.ox.ac.uk/~tobacco/ (23 January 2009, date last accessed).

[25] Katanoda K, Marugame T, Saika K, *et al.* Population attributable fraction of mortality associated with tobacco smoking in Japan: a pooled analysis of three large-scale cohort studies. *J Epidemiol* 2008;**18:**251–64.